ASHG 2024 Poster Abstracts

As of November 20, 2024

Instructions

1. View the Table of Contents on the next page.
    a. The Table of Contents is sorted chronologically and alphabetically by Session Title and then alphabetically by Abstract Title.
2. Click on the Abstract Title of the abstract you would like to view, and you will be taken to the page the abstract is on.

Table of Contents

Wednesday, November 6, 2024, 2:30 pm - 4:30 pm 164

Session Title: Cancer Poster Wednesday Session 164

Board 8001W: A breast cancer risk prediction model integrating both genetic and non-genetic factors in women of Nigeria, Cameroon, and Uganda 164

Board 8002W: A comprehensive quantitative catalog of differentially expressed mice tissues whole transcriptomefor understanding early cancer-causing events 166

Board 8003W: Acute endothelial stresses identify microRNA let-7b and non-coding *SLC11A2 (NRAMP2/ DMT1)* exon as biomarkers that overlap with those detected in malignant and non-malignant diseases 167

Board 8004W: AlphaMissense pathogenicity scores predict response to immunotherapy and enhances the predictive capability of tumor mutational burden ⭐ 168

Board 8005W: An ex vivo screening for growth factor combinations reveals estrogen-responsive cellular signatures contributing to immunosuppressive subtypes of ovarian cancer 169

Board 8006W: An integrated platform for cancer precision medicine: Genetic testing coupled with patient-derived organoid (PDO) and patient-derived xenograft (PDX) models ⭐ 170

Board 8007W: An MDS patient with deletion 20q and a t(9;22)(q34;q11.2): A case report and review of the literature 171

Board 8008W: Assessing the impact of genetic variations on metformin's efficacy in treating colorectal cancer 171

Board 8009W: Association of*HOXB13* G84E with any and lethal prostate cancer among 648,421 men 172

Board 8010W: Building an Ecosystem for Genetic Testing of Hereditary Cancer Syndromes in Algerian Population: The Experience of our Research Laboratory (2008-2024) 174

Board 8011W: Challenges of germline predisposition testing in patients with hematologic malignancies 175

Board 8012W: Characterization of chromosome 22q deletion in papillary thyroid carcinoma 176

Board 8013W: Charting Cancer Evolution Landscape on the Coordinate of Somatic Mutations by Single-cell NanoSeq 177

Board 8014W: Circadian Rhythm Genes in Chronic Myeloid Leukemia: An Untold Tale 178

Board 8015W: Clinical characteristics and Lynch Syndrome Associated Variants in Taiwan study 179

Board 8016W: Coding mutations of genitourinary development genes carry high risk for prostate cancer. 180

Board 8018W: Comprehensive Clinico-genetic Profiling Reveals Prognostic Significance of Cytogenetic Aberrations in Multiple Myeloma: Implications for Risk Stratification and Treatment Outcomes in an Indian Patient Cohort 181

Board 8019W: Conquering SPP1-driven resistance of solid tumors to CAR T cell therapy 182

Board 8020W: Consideration about Japanese variants in *TP53* based on the ToMMo-54KJPN: enhanced prediction accuracy with AlphaMissense 183

Board 8021W: Data Science Approaches to Understanding Hereditary Cancer Syndromes in Colombia 184

Board 8022W: Deciphering the genetic mechanism of sex differences within multiple myeloma 185

Board 8023W: Differential gene expression at diagnosis in high hyperdiploid pediatric acute lymphoblastic leukemia with negative and positive MRD at the end of induction treatment phase 186

Board 8024W: DNA Methylation and Glucose Metabolism Gene Variants as Predictive Markers for Pancreatic Cancer Risk 187

Board 8025W: Effective batch effect correction in large-scale TruSeq and capture-based RNA sequencing 188

Board 8026W: Evaluating CRLF2 rearrangement status along with relevant diagnostic impact in Ph-Like acute lymphoblastic leukemia 189

Board 8027W: Expanding our understanding of the genetics of mole count 190

Board 8028W: Expanding the central nervous system tumor spectrum of *POT1* germline variants 191

Board 8029W: Exploring the yield of germline genetic testing on skin biopsy samples . 193

Board 8030W: Generative AI impact on protein stability prediction in breast cancer genes 194

Board 8031W: Genetic Ancestry and Risk of Thyroid-Related Adverse Events Among Lung Cancer Patients Receiving Immune Checkpoint Inhibitors. 195

Board 8033W: Genetic testing for solid tumors: A 59-gene panel versus A 1021-gene panel 196

Board 8034W: Genome-wide association analysis of tumor markers in WHALE cohort with East Asian ancestries..... 197

Board 8035W: Genome-Wide Mapping of Regulatory Elements and Clusters of Non-Coding Somatic Mutations in Liver Cancer Genomes ⭐ 198

Board 8036W: Genomic data analysis of large lung cancer cohort with novel Native Hawaiian data reveals racial disparity in the landscape of somatic mutations..... 199

Board 8037W: Germline and somatic human leukocyte antigen (HLA) variations across racial/ethnic and ancestral populations in 10,479 patients with cancer 200

Board 8038W: GWAS and Functional Genomics Reveal Candidate Susceptibility Genes, Including Known Drivers *MECOM* and *KRAS*, in Endometrial Cancer..... 201

Board 8039W: High-throughput assay for evaluation of unclassified sequence variants in the breast and pancreatic cancer susceptibility gene *ATM* 202

Board 8040W: Homologous repair deficiency characterization from RNA-seq using allelic imbalance..... 203

Board 8041W: Identification of breast cancer risk genes which modulate responses to immune-mediated destruction 204

Board 8042W: Identification of cross-cancer risk loci among groups of cancers defined by biological and etiological similarities 205

Board 8043W: Interim GWAS in bladder cancer: Towards a multi-omics approach 206

Board 8044W: Investigating the relationship between breast cancer risk factors and mammographic AI-generated texture feature 207

Board 8045W: Large-scale population-based study highlights increased cancer risk among individuals with elevated genetic predisposition to autoimmune diseases..... 208

Board 8046W: Leveraging Custom CRISPR/Cas9 Screens to Identify AraC-Daunorubicin-Etoposide (ADE), Clofarabine, Gemtuzumab Ozogamicin (GO), and Bortezomib Response Modulators Association with Clinical Outcomes in Pediatric AML 210

Board 8049W: Multimodal analysis of rare *BARD1* missense variant found in a patient with multiple primary tumors 211

Board 8050W: Multiomic 6-base data from cell-free DNA enhances the performance of liquid biopsy classifiers ⭐ 212

Board 8051W: Multiple newly identified Ewing sarcoma susceptibility loci expand etiologic knowledge and nominate mechanisms of risk 213

Board 8052W: Mutational analysis of *FANCA* gene deletion in Tunisian Fanconi anemia patients 214

Board 8053W: Mutational Penetrance of the *CYP17A1* Gene in the Evolution of Uterine Fibroids Coexisting with Pregnancy in Senegalese Women ⭐ 215

Board 8054W: One patient with acute promyelocytic leukemia (M3) bearing tetraploid plus double t(15;17) and PML/RARA rearrangements achieved a complete remission and remained heathy during the six-years follow-up period 216

Board 8056W: Preclinical Model Atlas to Enable Clinical Translatability for Precision Oncology 217

Board 8057W: Prevalence and cancer phenotype of biallelic pathogenic variants in *CHEK2*: A genome-first approach in the Geisinger and UK Biobank cohorts. 218

Board 8058W: Quantifying the impact of coverage and switch accuracy on detection of somatic copy number alterations in cancer using whole exome sequencing. 219

Board 8060W: Sex chromosome gene expression does not match reported sex in a large proportion of cancer cell lines 220

Board 8061W: Single-cell analysis of gastric premalignant lesions and gastric cancer reveals cell lineage diversity and micro-environmentheterogeneity 221

Board 8062W: Single-cell landscape deciphering cancer cell-of-origin and cellular heterogeneity in malignant transformation of 13 major tissues..... 222

Board 8063W: Smoking and Pathogenic Germline Variants in Patients with Lung Cancer 223

Board 8064W: Spatial guided contrast learning for predicting spatial transcriptomics from histology images to enhance breast cancer patient stratification and immunotherapy response 224

Board 8065W: Testing results and clinical applications from 73 lymphoma patients using a 115 gene panel for lymphoma 225

Board 8066W: The mutational spectrum of arsenic-related human tumors 226

Board 8067W: Unveiling Telomere Dynamics: Alterations in Telomeric Nuclear Architecture and Variability in Telomere Lengths in Alternative Lengthening of Telomere 227

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session 228

Board 5001W: A Combination of polygenic risk score and measured blood pressure for risk stratification in cardiovascular disease mortality. 228

Board 5002W: A genetic analysis of brain cortical curvature and major depressive disorder in Mexican Americans 229

Board 5003W: A genetic case-control study of three single nucleotide polymorphisms in *ABCB4*, *ABCG8* and *SLC10A2* to explore their contribution towards the onset of cholelithiasis in Pakistani population 230

Board 5004W: A loss-of-function *CCR2* variant is associated with lower cardiovascular risk 231

Board 5005W: A modular network-based approach improves post-omics and post-GWAS hypothesis generation about disease genes and mechanisms ⭐ 232

Board 5006W: A Multi-ancestry Transcriptome-Wide Association Study of Alzheimer's Disease 233

Board 5007W: A multi-pronged strategy for target identification in Polycystic Kidney Disease 234

Board 5009W: A novel stop gain *SORL1* mutation in a Peruvian family with Alzheimer's Disease of the PeADI Study 236

Board 5010W: A polygenic risk score and case finding for chronic obstructive pulmonary disease 237

Board 5011W: Aberrant recursive splicing as a cause of human disease ⭐ 238

Board 5012W: Advanced Genetic Characterization and Pathway Analysis of IgA Nephropathy: Insights from Rare Variant, Exome-Wide Studies, and Genetic Diagnostics 239

Board 5013W: Alteration of JAK-STAT signaling in a patient with TYK2 variants 240

Board 5014W: Alzheimer's Disease risk and somatic loss of the Y chromosome..... 241

Board 5015W: Alzheimer's disease microglia exhibit an enrichment of somatic cancer driver mutations, correlating with disease-associated inflammatory states ⭐ 242

Board 5016W: An iPSC-based neural model of sialidosis uncovers glycolytic impairment-causing presynaptic dysfunction and deregulation of Ca^{2+} dynamics..... 243

Board 5017W: Analysis of the coding regions of 33,763 individuals identifies known and novel gene associations for idiopathic nephrotic syndrome 244

Board 5018W: Ancestry-aware GWAS identifies novel common variants associated with susceptibility to disseminated coccidioidomycosis..... 246

Board 5020W: *APOE4* stratified Alzheimer's disease polygenic risk score association with circulating proteins but not with circulating metabolites..... 247

Board 5021W: Application of Evolutionary Theory and Machine Learning to Discover Novel Risk Genes in Coronary Artery Disease 248

Board 5022W: Application of polygenic risk scores to discriminate different types of diabetes in early onset diabetics and to predict the prediabetic stage in the general population 249

Board 5023W: Assessment of a novel syndromic T2D subset based on IGF1 Resistance with high unmet medical need..... 250

Board 5024W: Association of Single Nucleotide Polymorphism in OCT1 Gene in type II diabetes mellitus patients from North Indian population. 251

Board 5025W: Associations between a polygenic predisposition to later pubertal timing and diagnostic odysseys for patients with abnormal growth..... 252

Board 5026W: Autism diagnosis at birth via functional analysis of IP_3receptor Ca^{+2} signaling in fibroblasts 253

Board 5027W: Brain proteome of well-being: A proteome-association study unveils novel proteins related to increased well-being..... 254

Board 5028W: Characterization of the obese proteome and downstream sequalae in Hispanic Latino study participants..... 255

Board 5029W: Characterizing the genetic architecture of Autism from a multi-ancestry perspective ⭐ 256

Board 5030W: Circulating Mfge8 levels increase the risk of human T2DM and Cardiovascular disease..... 257

Board 5031W: Clinical Variability Among Three Siblings with Tubulin-Folding Cofactor D-Related Encephalopathy: Further Delineation of Clinical Presentation 258

Board 5032W: Comparisons of subcutaneous adipose and plasma metabolite associations with BMI and with genome-wide genetic variation..... 260

Board 5033W: Comprehensive analysis of the genetic variation in the *LPA* gene from short-read sequencing..... 261

Board 5034W: Comprehensive Genomic Analysis of Healthy Aging in Super Seniors .. 262

Board 5035W: Contribution of copy number variants to schizophrenia in East Asian populations 263

Board 5036W: Cross-population statistical fine-mapping and exome-wide association study identify causal condensate variants for type 2 diabetes 264

Board 5038W: Decoding the Genetic Basis of Autoimmune Gastritis and Pernicious Anemia. 265

Board 5039W: Decreasing polygenic prediction accuracy with age is driven by age-dependent genetic and non-genetic disease liability 266

Board 5040W: Delineating ancestry-specific genetic patterns in type 2 diabetes: Implications for treatment efficacy from real-world and trial data 267

Board 5041W: Detecting gene and isoform expression changes in the IBDs using long-read sequencing 268

Board 5042W: Detecting genetic effects on variability in body fat distribution in UK Biobank to uncover gene-environment interactions 269

Board 5043W: Development of a multi-ancestry type 1 diabetes genetic risk score and application to genomic changes in the pancreas 271

Board 5044W: Diagnostic potential of miR-548c-3p in CKD pathogenesis by integrated computational and experimental analyses 272

Board 5045W: Differential gene expression of opioid overdose death across human brain regions 272

Board 5046W: Disentangling allele-specific effects of the HLA class II (HLAII) haplotype DR15/DQ6 in the development of inhibitory anti-drug-antibodies to therapeutic factor VIII proteins (tFVIIIs) using a peptidomics approach 274

Board 5047W: Dissecting type 1 diabetes etiology with human genetics and population-scale multiome profiling of human islets under basal and pro-inflammatory conditions 275

Board 5048W: DNA damage-associated protein co-expression modules provide insight into genetic variation tolerance and cardiovascular disease risk 276

Board 5049W: Doubling the number of genetic risk loci for mosaic loss of chromosome Y in 544,112 diverse ancestry participants from the Million Veteran Program ⭐ 277

Board 5050W: Educational Attainment as a Cognitive Reserve Factor in Alzheimer Disease: Implications for *APOE e4* Carriers 279

Board 5051W: Elucidating Sex-Differential Genetic Associations on Plasma Metabolites and Coronary Artery Disease Risk: Insights from Sex-stratified GWAS Analyses in the UK Biobank. ⭐ 280

Board 5052W: *EMILIN2* dissociates obesity from abdominal fat distribution by enhancing adipocyte differentiation and function 281

Board 5053W: Engineering 3D corneal organoids from patient-derived iPSCs to model and investigate underlying genetic influences on keratoconus pathophysiology 282

Board 5054W: Enrichment of rare copy number variations in metabotropic glutamate receptor gene networks in ADHD patients with comorbid neurodevelopmental phenotypes 283

Board 5055W: Establishment and Characterization of Mouse Models for the Study of Human Major Histocompatibility Complex Class I Gene Effects 284

Board 5056W: Estimated selection coefficients of protein coding variants support the large fitness impact of pathogenic variants in neurodevelopmental disorders 285

Board 5057W: Evaluation of sex-stratified polygenic risk scores for type 2 diabetes and glycemic traits in the Framingham Heart Study 286

Board 5058W: Evidence for widespread gene-environment interactions in genome-wide data 287

Board 5059W: Examining the shared genetic architecture of keloid scars and uterine fibroids 288

Board 5060W: Exome sequencing of >1 million individuals identifies 232 genes associated with adult height,bridging the gap between common traits and rare diseases 289

Board 5061W: Expanding genetic insight into steatotic liver disease contributed by Finnish founder effect 290

Board 5062W: Exploring the genetic complexity of Alzheimer's disease in African ancestry using single-cell multi-omics integrative analysis ⭐ 291

Board 5063W: Exploring the impact of Life's Essential 8 and APOE4 allele on brain aging in the UK Biobank 292

Board 5065W: FGFR2-related syndromes and de novo variants in a sub-Saharan African cohort with craniosynostosis 293

Board 5066W: Functional dissection of genetic interactions contributing to phenotype variability of complex disorders 295

Board 5067W: Functionally-informed fine-mapping and gene prioritization in height-associated loci 296

Board 5068W: G6PD deficiency masks the diagnosis of abnormal glucose tolerance: insight from the Africans in America Study 297

Board 5069W: Gene clusters and patterns of neurodevelopmental phenotypes in cerebral palsy 298

Board 5070W: Gene-diet interaction analysis in UK Biobank identified genetic loci that modify the association between fish oil supplementation and the incidence of dementia 299

Board 5071W: Genetic analysis of chromosome X nominates a new risk locus in females with Lewy body dementia 300

Board 5072W: Genetic factors associated with depression and cognitive decline 301

Board 5073W: Genetic Landscape of Postpartum Psychosis from Common to Rare Variants 302

Board 5074W: Genetically Predicted Clonal Hematopoiesis of Indeterminate Potential and Telomere Length and Risk of Multiple Primary Cancers 303

Board 5075W: Genome wide association study investigating increased risk of postpartum depression among women affected by immune-mediated inflammatory conditions 304

Board 5076W: Genome-wide association study of dermatochalasis reveals novel susceptibility loci 305

Board 5077W: Genome-wide association study of multiple neuropathology endophenotypes identifies novel risk loci and provides insights into genetic risk of dementia 306

Board 5078W: Genome-wide association study of non-alcoholic fatty liver disease based on ultrasound diagnosis in Korean population 307

Board 5079W: Genome-wide association study of proteomic aging reveals a shared genetic architecture with longevity, reproductive timing, and age-related diseases 308

Board 5080W: Genome-wide association study of responsiveness by SARS-CoV-2 mRNA vaccination in Japanese population 309

Board 5081W: Genome-wide linkage analysis reveals a novel locus for increased plasma p-tau 181 levels on Chromosome 1 in the Amish 310

Board 5082W: Genomic Signatures of Aging for Frailty in the NHLBI TOPMed Program ⭐ 312

Board 5083W: Germline genetics, disease, and medication exposures influence longitudinal dynamics of clonal hematopoiesis of indeterminate potential 313

Board 5084W: Girl case report: Nevus Spilus Syndrome coexisting with Ehlers-Danlos Syndrome and psychiatric disorders 314

Board 5085W: Global microRNA Profiling in Human Urinary Extracellular Vesicles: Identifying Novel Biomarkers and Pathways for Diabetic Nephropathy 315

Board 5086W: GWAS of neuropsychiatric symptoms in MCI/ADRD 316

Board 5087W: Harnessing Polygenic Scores to Personalize Smoking Cessation 317

Board 5088W: Healthcare provider responses to high polygenic risk disease reports: Return of results experience at the University of Alabama at Birmingham 318

Board 5089W: Hippocampal single nucleus transcriptomics of environmentally impoverished mice to reveal the addictome 319

Board 5090W: HLA-focused genetic risk prediction in populations of diverse ancestry with type 1 diabetes 320

Board 5091W: Hypothesis-free Mendelian Randomisation Identifies New Metabolites and Athletes Metabolites Linked to Risk of ALS 321

Board 5092W: Identification of circular RNA in Alzheimer's Disease 322

Board 5093W: Identification of common genetic risk loci associated with ischemic stroke in Taiwan 323

Board 5094W: Identification of Genes Associated with Alzheimer’s Disease Using Novel Region Specific Microglia Cell Models 324

Board 5095W: Identification of multiple new loci associated with late-onset Alzheimer’s disease in Japanese 325

Board 5096W: Identifying Genetic Determinants for Multi-Ancestry Portability of LDL Polygenic Risk Scores 326

Board 5097W: Identifying rare variants for cognitive preservation on chromosome 2 in the Midwestern Amish 327

Board 5098W: Impact of *HLA* diversity in classifying diabetes subtypes in Indians using type 1 diabetes genetic risk score 328

Board 5099W: Impact of lifestyle modifications on individuals with high polygenic scores for obesity: insights from the TMM CommCohort Study 329

Board 5100W: Inferred course of hospitalized COVID-19 progression from longitudinal multi-omics data associates with disease severity, patient survival, dynamic molecular changes, and genetic variation. 330

Board 5101W: Integrating Brain Cell-Type Specific Omics Data to Identify Causal Genes and Cell Types for Parkinson's Disease Associated Loci 332

Board 5102W: Integrating Multi-Omic Data to Unveil Molecular and Cellular Pathways Mediated by Common Polygenic Variation Using Deep Neural Networks 333

Board 5103W: Integrative Analysis of Allele Specific Expression and miRNA Profiling Reveals Underlying Disease Mechanism in Age-Related Macular Degeneration ⭐ 334

Board 5104W: Interaction effects of genetic variants and white matter hyperintensities on Alzheimer's disease biomarkers 335

Board 5105W: International meta-analysis of genome-wide association study on adolescent idiopathic scoliosis 336

Board 5106W: Investigating CYP26B1 as novel obesity locus in indigenous Americans living in Arizona 338

Board 5108W: Joint Fluid Single Cell Transcriptomics in Pediatric Septic Arthritis Discovers Signatures of Autoimmunity in Culture Negative Patients 339

Board 5109W: Joint modeling across tissues via tensor decomposition implicates GCM2 and ZNF73 regulons in genetic risk of Alzheimer's disease 340

Board 5110W: Large-scale analysis of sequencing data uncovers association between ultra-rare protein-truncating variants in *ITSN1* and Parkinson's disease ⭐ 341

Board 5112W: Large-scale Tourette Syndrome whole-exome sequencing analysis reveals a significant contribution of de novo mutations different from Autism Spectrum Disorder. 342

Board 5113W: Leveraging trisomy 21's unique characteristics to unveil the genetic basis of common complex diseases 343

Board 5115W: Mapping AD Progression: Multi-region spatial transcriptomics to study susceptibility to amyloid beta (Aβ) plaque in Alzheimer's Disease (AD) 344

Board 5116W: Mechanistic insights from liver fat GWAS loci 345

Board 5117W: Medical phenome-scale evaluation of the Polygenic Score Catalog in the BioMe Biobank in New York City 346

Board 5119W: Mendelian Randomization reveals causal links between circulating metabolites and Type 1 Diabetes. ⭐ 347

Board 5120W: Meta-analysis of European and East Asian GWAS Identifies Novel Susceptibility Loci in Females with Adolescent Idiopathic Scoliosis 348

Board 5121W: Metabolic pathways in aortic stenosis: A Mendelian randomization approach. 349

Board 5122W: Monogenic diabetes in Tunisia: Genetic testing reveals novel variants and an oligogenic signature enabling a new classification of diabetic patients and improved health care management ⭐ 350

Board 5123W: Mosaic Loss of Chromosome X in Peripheral Blood Cells from Whole-Genome Sequencing and Alzheimer's Disease in the Amish 352

Board 5124W: Multi-Ancestry Genome Wide Association Study (GWAS) in the *All of Us* Cohort for Primary Open Angle Glaucoma 352

Board 5125W: Multi-ancestry Genome Wide Association Study Meta-analysis of Non-syndromic Orofacial Clefts 353

Board 5126W: Multi-ancestry genome-wide association study for neuroticism 355

Board 5127W: Multi-ancestry meta-analysis investigating the genetic susceptibility of intracranial aneurysms in high-risk populations 356

Board 5128W: Multimodal Deep Learning for Disease Forecasting using UK Biobank ⭐ 357

Board 5129W: Multi-Trait Analysis of the FinnGen Population 358

Board 5130W: Multi-trait GWAS leverages power to identify childhood adiposity signals and reveals life-stage effects of leptin-melanocortin pathway genes 359

Board 5131W: Nanopore long-read sequencing of Lewy body dementia cases from the DementiaSeq1 cohort for structural variation and integrative genetics study..... 360

Board 5132W: Novel Genetic Variants Linked to Liver Cirrhosis and Cancer: Evidence of Gene-Environment Mediation and Interaction in the Million Veteran Program ⭐ 361

Board 5133W: Optimization and Validation of Multi-Ancestry Polygenic Risk Scores... 363

Board 5134W: PBMC profiling in single ventricle/hypoplastic left heart patients by single-cell analysis 363

Board 5135W: Phenomics Network Analysis for Depression Reveals Novel Comorbidity Architecture, Genomic Relationships, and Pleiotropic Variants 365

Board 5136W: Pima/O'odham Informed Consent in Genomic Studies: a Scoping Review 366

Board 5137W: Polygenic prioritization of monogenic IBD genes 366

Board 5138W: Polygenic risk score model based on polymorphisms related to metabolic syndrome offers valuable genetic insights 367

Board 5139W: Polygenic risk score predicts incident cardiovascular disease in people with diabetes at borderline or intermediate 10-year ASCVD risk ⭐ 368

Board 5140W: Polygenic Risk Score Predicts Lung Function Response to PM2.5 Exposure in Children 369

Board 5141W: Polygenic score-wide association study of comorbidities to incident dementia 370

Board 5142W: Polygenic trait-wide analysis identifies complex traits associated with immune-mediated inflammatory diseases and their clinical phenotypes 371

Board 5143W: Potential Causal Link between Hyperuricemia and Incident Chronic Kidney Disease in Eastern Asians: Insights from Large-scale Mendelian Randomization Analyses 373

Board 5144W: Potential genetic regulators of a novel plasma-based biomarker for Alzheimer's disease 374

Board 5145W: Prediction of coronary heart disease in Hispanics/Latinos using state-of-the-art polygenic risk scores: Findings from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) 375

Board 5146W: Prenatal autoimmune endocrine disorders and risk of autism 377

Board 5147W: Rare Deleterious ATXN2 Variants in Extreme Adult Obesity 378

Board 5148W: Rare variants identify novel autism genes and different genetic architecture in simplex and multiplex autism families ⭐ 378

Board 5149W: Revolutionizing precision health in Type 2 Diabetes: AI integration of multimodal medical imaging and genome-wide genetic data in a large-scale biobank 380

Board 5150W: Ribosomal DNA copy number modifies the genetic risk for Type 2 diabetes via dysregulated ribosome biogenesis in pancreatic islets 381

Board 5151W: Sex-specific causal inference assessment of immune cells and protein levels in neurodegenerative diseases 382

Board 5152W: Sex-Stratified Single-Cell Analysis Uncovers Gene Expression Differences in Schizophrenia, Bipolar Disorder, and Autism 383

Board 5153W: Sexually Dimorphic Association of CPS1 with Atherosclerosis in Humans and Mice 384

Board 5154W: SGLT2 Inhibitors Attenuate Risk of *APOL1*-mediated Chronic Kidney Disease 386

Board 5155W: Shaping the Genetic Landscape of ATXN2: Insights from a Novel Autosomal Recessive CAG Repeat Expansion at the Spinocerebellar Ataxia Type 2 Locus in a Homozygous Acadian Family 387

Board 5156W: Shared genetics of plasma protein level network guides composite risk scores for coronary artery disease and type 1 diabetes 388

Board 5157W: Single-cell multiomics study in hiPSC-derived neurons from large cohorts reveals novel gene regulatory mechanisms of schizophrenia ⭐ 389

Board 5158W: Single-nucleus multimodal dissection of Alzheimer's Disease identifies brain region-specific dysregulation of the regulome. ⭐ 390

Board 5159W: SLC30A8 rare variant modify contribution of common genetic and lifestyle factors toward type 2 diabetes 391

Board 5160W: *SOD1* Related Amyotrophic lateral sclerosis Progression A Case Report 392

Board 5161W: Spatial Multi-omic architecture of Alzheimer's disease : Lessons from Human and Mouse model systems 393

Board 5162W: Spectrum of rare genetic variation in etiology of non-syndromic autism spectrum disorder 394

Board 5163W: Targeting Dementia Subtypes using Comparative Genomics and Drug Repositioning 395

Board 5164W: The DNA of distress: Uncovering genomic associations of allostatic load 396

Board 5165W: The genetic determinants and genomic consequences of non-leukemogenic somatic point mutations ⭐ 397

Board 5166W: The impact of lifestyle intervention on cardiovascular and kidney outcomes in genetically high-risk obese individuals ⭐ 399

Board 5168W: Genome-Wide Association Study Identifies Variants Associated with Neural Pathway Morphology 400

Board 5169W: Tractor Analysis of African American Inflammatory Bowel Disease Identifies Ancestry Specific Associations 401

Board 5170W: Transcriptomic and proteomic analysis of isogenic *ARC* gene deficiency SH-SY5Y cells 402

Board 5171W: Translating genomic risk models to clinical trials in a learning health system: the Prostate CAncer integrated Risk Evaluation (P-CARE) model ⭐ 403

Board 5172W: Type 2 Diabetes Is Causally and Independently Associated with Pulmonary Tuberculosis: A Mendelian Randomization Study 404

Board 5173W: Uncoupling the genetics driving obesity and cardiometabolic disease . 405

Board 5174W: Understanding the common genetic architecture of spontaneous coronary artery dissection using genetically related phenotypes 406

Board 5175W: Understanding the relationship between giant cell arteritis and polymyalgia rheumatica using population biobank data 408

Board 5176W: Understanding the role of APOE genotype in the genetic regulation of gene expression and splicing in human microglia 409

Board 5177W: Unraveling the Genetic Architecture of Inflammatory Bowel Disease Through Large-Scale Exome Sequencing ⭐ 410

Board 5178W: Whole genome sequencing analysis identifies rare non-coding variants with large effects and regulatory regions associated with circulating protein levels ⭐ 411

Board 5179W: Whole genome sequencing analysis of severe, early-onset chronic obstructive pulmonary disease 412

Board 5180W: Whole genome sequencing identifies deleterious novel variants in craniofacial enhancers of Africans with orofacial clefts 413

Board 5181W: Whole-genome sequencing identifies *SIKE1, CD36,* and *IFFO1* as candidate susceptible genes for systemic lupus erythematosus 415

Session Title: Epigenetics Poster Wednesday Session 416

Board 1002W: 3D-genome impacts of TAD-boundary-disrupting ultra-rare structural variants identified in schizophrenia ⭐ 416

Board 1004W: An atlas of genetic effects on the monocyte methylome across European and African populations 417

Board 1005W: An improved DNA methylation signature for Kabuki Syndrome for classification of KMT2D variants 418

Board 1006W: Atlas of single-cell ATAC plus gene expression reveals differential epigenetic regulation across many tissue types and species 419

Board 1007W: Characterizing Epigenomic Differences in Pulmonary Hypertension Subtypes 420

Board 1008W: Childhood adversity increases young adult Elevated Alcohol Consumption and contributes to increased young adult cardiac risk among Black youth 421

Board 1009W: Comprehensive profiling of genetic and epigenetic features for FSHD diagnosis using targeted nanopore sequencing 422

Board 1010W: Controlled Fusion of Heterogenous ATAC-seq Samples for Robust Quantification of Chromatin Accessibility 423

Board 1011W: Development of bias correction methods for cord blood-based epigenome-wide association studies 424

Board 1012W: Diagnosis of Mendelian diseases using genome-wide DNA methylation signatures: Preliminary practice from a Chinese patient cohort..... 425

Board 1013W: DNA methylation patterns and differences in two long-read sequencing platforms in a clinical pediatric cohort 426

Board 1014W: DNA methylation signatures in the prefrontal cortex of alcohol-naïve nonhuman primates predict risk for future ethanol drinking. 427

Board 1015W: Downregulation of nuclear receptor-binding SET domain protein induces neurotoxic cytokines via p38 mitogen-activated protein kinase and Jun kinase pathways in a glioma cell line, U87MG 428

Board 1016W: Endogenous retroviruses are a source of regulatory variants in colorectal cancer 429

Board 1017W: Endothelial cell to cell adhesion is altered as consequence of loss of function of neurofibromin and afadin6 in neurofibromatosis type1 430

Board 1018W: Epigenetic clocks as novel biomarkers of checkpoint inhibitor treatment response..... 431

Board 1019W: Epigenetic reprogramming successfully reactivates gamma globin expression, paving way for β-hemoglobinopathies treatment ⭐ 432

Board 1020W: Epigenetic variation linked to opioid overdose converges on key target genes in the nucleus accumbens..... 433

Board 1021W: Epigenome-wide association study of placental DNA methylation and cumulative maternal glycemic levels throughout pregnancy 434

Board 1022W: Evaluating CRISPR-Cas9 scaRNA1 deletion on biochemical modification and function of the spliceosome..... 435

Board 1023W: Functional redundancy and pathogenic mechanism in *RING1* variants: Insights from H2AK119ub1 dosage in human neural development 436

Board 1024W: Generation of hypothalamus with normalized gene expression of the responsible imprinting genes derived from Prader-Willi syndrome patient cells using epigenome editing technology. 437

Board 1025W: Genetic regulation of methylation across East Asian and European populations 438

Board 1026W: Identification of circulating epigenetic biomarkers for hemorrhage in brain arteriovenous malformation by whole-genome methylome profiling 439

Board 1027W: Identification of microRNA signatures carried by neuronal released extracellular vesicles in Alzheimer's Disease and their putative effect in altering target cell gene expression with population specificity 440

Board 1028W: Identifying myogenic enhancers and 3D genome regulation across human development 441

Board 1029W: Investigating epigenetic alterations to the chromatin landscape in the progression of gastric-esophageal adenocarcinomas to peritoneal metastasis. 442

Board 1030W: Multi-modal inference of gene regulatory networks in systemic lupus erythematosus suggests ethnicity-specific differences in epigenetic regulation ⭐ ... 443

Board 1031W: Peakverse:Robust suite of tools for prioritizing and visualizing single cell ATAC sequencing peaks 445

Board 1033W: Single cell analysis of neurogenesis in adult human olfactory epithelium ⭐ 446

Board 1034W: Spatiotemporal Twist1 deficiency shifts forelimb to hindlimb expression pattern resulting in multidigit phenotype 447

Board 1035W: Targeted cell-free DNA methylation haplotype signatures in prostate cancer progression ⭐ 448

Board 1036W: Tip60 and ventrolateral neurons (LNv): insights into alcohol-induced neuroadaptations in Drosophila 449

Board 1037W: Transcription factor activation induces site-specific copy gain of the interferon receptor cluster ⭐ 450

Board 1038W: Using DNA methylation to investigate phenotypic heterogeneity 451

Board 1039W: Whole Genome Methylation Profiling to Enhance Diagnostic Yield in Neurodevelopmental Disorders 452

Board 1040W: Whole genome methylation sequencing in blood from persons with mild cognitive impairment and dementia due to Alzheimer's disease identified extensive differences that are sex-specific and predictive of cognitive status 453

Session Title: Evolutionary and Population Genetics Poster Wednesday Session 454

Board 3002W: A multidisciplinary approach to understand the prehistorical societies of Qatar 454

Board 3003W: A population-specific pangenome reference improves variant discovery in Pacific Islanders 455

Board 3004W: Ancestral origins and post-admixture adaptive evolution of highland Tajiks 456

Board 3005W: Assessment of a microhaplotype panel for human identification, population structure and ancestry inference in Brazil 457

Board 3006W: Balancing selection and the malaria parasite genome: Insights into human-parasite co-evolution through genetic variation 458

Board 3007W: Characterization of the coronary artery disease-associated 9p21 locus across ancestries, and its role in genetic susceptibility across populations 459

Board 3008W: Comparative genomic analysis of differential gene regulation using multimodal data in dozens of human and chimpanzee cell types 460

Board 3009W: Correlated trait effects of nearby variants reflect phenotypic selection and mutational architecture 461

Board 3010W: Digital image processing to detect adaptive evolution 463

Board 3011W: Enhancing rare variant discoveries with population genetics insights ... 463

Board 3013W: Expanding the utility of ancient DNA using large-scale genome sequencing 464

Board 3014W: Exploring Genetic History and the Spread of Lactase Persistence in Tunisian Populations 465

Board 3015W: Genetic-similarity-based approaches for matching to populations in reference datasets 467

Board 3016W: Geography and study design in the biobank era: a population genetics-informed approach to understanding rare, deleterious variation in human genetic samples 468

Board 3017W: Global diversity, recurrent evolution, and recent selection on amylase structural haplotypes in humans 469

Board 3018W: High Throughput phasing of HLA in a diverse ancestry set of individuals with high accuracy 470

Board 3019W: Human Germline Mutational Signatures in 525K UK Biobank and Genomics England 100K Whole Genomes 470

Board 3021W: Identifying and interpreting lineage-specific fixed differences between modern and archaic humans 472

Board 3022W: Investigating ghost introgression in Denisovans via ancestral recombination graphs 473

Board 3023W: Investigation of homozygosity effects using variant annotation in a sample with individuals from founder populations 474

Board 3024W: Landscape of clonal hematopoiesis among East Asian individuals aged 80 years and older 475

Board 3025W: Leveraging Identity-by-Descent based methods in biobanks to understand how haplotypic variation contributes to Sickle Cell Disease severity 476

Board 3026W: Phenome-wide association studies of Mucins in *All of Us* data reveal tandem repeat-associated phenotypic traits 477

Board 3027W: Polygenic adaptation leads to a higher reproductive fitness of native Tibetans at high altitude 478

Board 3028W: Population health monitoring leveraging fine-scale population structure in the biobank at the Colorado Center for Personalized Medicine. ⭐ 479

Board 3029W: Recapitulating the African Diaspora Genome through ONT Long Read Sequencing 480

Board 3030W: Recruiting and Engaging Native Hawaiian and Pacific Islander Elders: Insights from "Brain Train" and Community Partnerships 481

Board 3031W: Refining constraint estimates of stop-loss variants 482

Board 3032W: Study of natural selection in post-LGM East Asian populations using ancient DNA 483

Board 3033W: The allelic architecture of functionally deleterious variants in the Saudi Arabian population 484

Board 3034W: The *APOL1* p.N264K variant is co-inherited with the G2 kidney disease risk variant through a proximity recombination event 485

Board 3037W: Tracing the evolution of E-Y14891, a major Ashkenazi Jewish Y-chromosomal lineage 486

Board 3038W: Unraveling the genetic potentiators of modern human migration 487

Board 3039W: Unusual female-bias in the *de novo* mutation rate of a lemuriform primate, the aye-aye (*Daubentonia madagascarensis*) 488

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session 489

Board 7052W: A Case Study: Modifying NIH NExTRAC Public Deliberation Prompts for Community Consultation to Advance ELSI Study of Sociogenomics Research ⭐ 489

Board 7053W: A Framework for Equity in Genomic Medicine and Research..... 490

Board 7054W: Advancing Breast Cancer Care: The Role of Implementing a Genomic Counseling Model in Colombia 491

Board 7055W: Advancing genetic testing for neurological disorders in Tanzania: importance, challenges, and strategies for implementation ⭐ 492

Board 7057W: Assessing the Role of Genetic Diversity and Environmental Knowledge on Healthcare Among Qinghai Minority Students..... 493

Board 7059W: Comparing telehealth and in-person genetic counseling visit times across specialties 494

Board 7060W: Decoding Ambiguity: Evaluating the Psychological Harms and Utility of Returning Cancer Variants of Uncertain Significance ⭐ 495

Board 7061W: Defining social and behavioral traits in polygenic scores and understanding measurement limitations. 496

Board 7062W: Diversifying the genetic counselor workforce via community colleges: A look at higher education experience before graduate school, grit, and resilience 497

Board 7063W: Doing no Harm: Exploring the psychosocial impacts of receiving secondary findings from whole exome sequencing 498

Board 7064W: Educating primary care providers about newborn screening and genetic conditions..... 500

Board 7065W: Efficient Inquiry Management in Expanding Biobanks: Balancing Growth and Support 501

Board 7066W: Emerging ethical considerations of neurogenetics data for Indigenous Peoples 502

Board 7067W: Empowerment of Diversity in Genomic Careers: Promotion of Workforce, Employment and Readiness in Genomics (PoWER-G) and the Path to Success 503

Board 7068W: Engaging high school students in exploring big data via the *All of Us* Data Browser 504

Board 7069W: Enhancing Breast Cancer Awareness Among College Students: A Cross-Sectional Study in Nigeria 505

Board 7070W: Evaluating community perceptions and ethical considerations in genetics research in small scale northern Kenyan populations 506

Board 7071W: Evaluation of Somatic Genetic Testing as a Screening Tool to Detect Hereditary Cancer Predisposition in Patients Diagnosed with Myeloid Malignancies ... 507

Board 7072W: Filling the pipeline: The Greenwood Genetic Center's approach to developing the future genetic workforce. 508

Board 7074W: From Oversight to Opportunity: Redefining Rural Inclusivity in Genetic Research 509

Board 7075W: Genetics as a growing dimension of group identity: A study using Wikipedia data 510

Board 7076W: Hands-on educational approaches for teaching and learning pharmacogenomics 511

Board 7077W: Hereditary cancer syndrome registry at an urban safety net institution. 512

Board 7078W: Identifying Precision Medicine Program strengths and areas of improvement by curriculum mapping to the core genetics and genomics competencies by the Association of Professors of Human and Medical Genetics (APHMG) 513

Board 7079W: Impact of polygenic and monogenic risk results on patients' perceived risk of hypercholesterolemia in the eMERGE IV study 514

Board 7080W: Personal utility of genetic testing depends on patient population: Evidence from four clinical cohorts 516

Board 7081W: Population Descriptors in Genomic Legacy Data: Reflections and Recommendations for Future Research Directions from a Trans-NIH Workshop ⭐ ... 517

Board 7082W: Scaling genetic counselling services across medical specialties in Singapore 519

Board 7083W: Teaching the Genome Generation: Cultivating High School Genomics Through Teacher Education 520

Board 7084W: The challenge of genetic counselling in Xp22.31 related CNVs 521

Board 7085W: The development and preliminary evaluation of 'The Genetics Navigator': a digital solution for adult and pediatric genetics 522

Board 7086W: Utility of whole genome sequencing in pediatric neurological disorders in Karachi 524

Board 7087W: Validity of self-reported family history for hereditary cancer risk assessment in the primary care setting 525

Session Title: Genetic Therapies Poster Wednesday Session 526

Board 8081W: Building a human genetics-guided side effect score for drug therapeutics 526

Board 8082W: Deciphering the Therapeutic Effect of Dynamin-2 Reduction in SPEG-Related Myopathy: Perspectives from Multi-Omics Analysis 527

Board 8083W: Engineering poison exons for allele-specific silencing of the Huntington's disease gene *HTT* 528

Board 8084W: Identifying and therapeutically targeting novel splice defects in Glycogen Storage Disease IX 529

Board 8085W: Individualized ASO Therapy Development for a Patient with *BIN1*-Associated Autosomal Dominant Centronuclear Myopathy 530

Board 8086W: Integrating Genetic and Proteomic Evidence to Predict Drug Approval Success Rates: Insights from the UK Biobank 531

Board 8087W: L-Citrulline Treatment of Nitric Oxide Deficiency in MELAS: A Phase I Dose-Finding and Safety Study 532

Board 8088W: Long-read motor neuron RNAseq of deep intron 8 *IGHMBP2* variants reveals pathogenicity and informs a precision therapeutic strategy 533

Board 8089W: Modulating 5'UTR splicing to treat haploinsufficient neurodevelopmental disorders 534

Board 8090W: Pathohistological study of intracranial Gaucheroma causing deafness in a patient with Gaucher Disease Type 3: Effects of substrate reduction therapy 536

Board 8091W: Penetrance of Alpha-1 Antitrypsin Deficiency among PiZZ Individuals in UK Biobank 537

Board 8092W: Plant Exosome-like Nanovesicles for Delivery of Gene Editing Tool (Antisense Oligonucleotide) to Treat Cell Model of Amyotrophic Lateral Sclerosis 538

Board 8093W: Preclinical development of a gene therapy for Ogden syndrome, using new mouse models of the disease ⭐ 539

Board 8094W: Purity and stability assessment of AAT protein to support Beam 302 clinical bioanalysis 540

Board 8096W: Response to asfotase alfa with no mutation in *ALPL* 541

Board 8097W: Stimulating And Simulating DNA For Auto-Correction To Enhance The Regeneration Of Lost Structures Or Functions In The Human Body 542

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session 543

Board 1042W: A custom array for genotyping rare predicted loss-of-function variants in the Michigan Genomics Initiative biobank 543

Board 1043W: A draft pangenome reference for people with Pacific ancestry 544

Board 1044W: A Framework for Ensuring High Quality Genetics and Omics Training Data for Artificial Intelligence and Machine Learning 545

Board 1045W: A novel data model for population descriptors in genomic research of diverse populations developed in the Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium 546

Board 1046W: A rare Chromosome 6p25.3p25.1 deletion and duplication at 2q34q37.3: A case report with dental and other clinical features. 548

Board 1048W: A unified computing environment for genomics data storage, management, and analysis: NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space (AnVIL) 548

Board 1049W: Assessing large language model performance related to aging in genetic conditions 549

Board 1050W: Bioinformatics-based meta-analysis of gut microbiome in epilepsy 550

Board 1051W: Characterizing discrepant allele frequencies across the gnomAD v4 exomes and genomes 551

Board 1052W: Comprehensive analysis of variant concordance reveals optimized average read-depth of human whole-genome sequencing data 552

Board 1053W: Comprehensive Genomic Profiling of Stroke Cohorts in Taiwan: Integrative Analysis of ROH, Clonal Hematopoiesis, Pharmacogenetics, Mitochondrial Mutations, UPD, Short Variants, GWAS and HLA Using Next-Generation Sequencing. 553

Board 1054W: Comprehensive Global Genome Diversity Resource : Bridging Gaps in Genomic Data Accessibility for Inclusive Health Research ⭐ 555

Board 1055W: Comprehensive variant catalog and genotype-phenotype correlations for Joubert syndrome associated genes 556

Board 1056W: Convolutions Accelerate mRNA Ramp Sequence Research in ExtRamp 2.0 557

Board 1057W: Deciphering fine-scale population structure in the U.S. to uncover population-specific disease risks for health equity 558

Board 1058W: Development of a novel gene-disease relationship classification framework and comparison to existing ClinGen standard 559

Board 1059W: Enabling cloud-based sharing, harmonization, and analysis of secondary use data in the multi-institutional Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium 560

Board 1060W: Enhanced Genotype Imputation in the Chinese Population Utilizing the GSRD-100K^{WCH} Haplotype Reference Panel from 22,374 Deep Whole Genome Sequenced Subjects ⭐ 561

Board 1061W: Evaluation of new built-in AI features from the database management system to enable natural language conversation with local data 562

Board 1062W: FaceBase: Data Resources for Craniofacial and Dental Development, Molecular Genetics, and Genomics 563

Board 1063W: From query to visualization: programmatic access and interactive workspaces for Alzheimer's disease genetics and >100,000 omics datasets 564

Board 1064W: AI cloud-based end-to-end technology for accurate, fast and affordable newborn screening of genetic disorders / rare diseases 565

Board 1065W: Genome structure and function in individuals with sex chromosome aneuploidies: A systematic review. 566

Board 1066W: Genome-wide allele-specific expression analysis aids interpretation of disease associated common and rare variants ⭐ 568

Board 1067W: Genome-wide de novo tandem repeat variation in a four-generation family extensively sequenced with multiple long- and short-read technologies. 569

Board 1068W: Genomic Return of Results for The United States' *All of Us* Research Program 570

Board 1069W: Harmonizing 60,000 whole genome sequence data to advance Alzheimer's Disease genetics research - Joint effort by the Alzheimer's Disease Sequencing Project and the Genome Center for Alzheimer's Disease ⭐ 571

Board 1070W: High confidence small Open Reading Frames (smORFs) show evidence of translation outside annotated transcript boundaries 573

Board 1072W: Landscape of Colorado Center for Personalized Medicine research freeze 4 574

Board 1074W: Leveraging semantic information in biomedical literature to annotate publicly-available samples and studies 575

Board 1075W: Long-chain fatty acid oxidation disorders: A comprehensive novel database of LC-FAOD gene variants, genotypes and phenotypes 576

Board 1076W: Matched Annotation from NCBI and EMBL-EBI (MANE): Current Status, Stability and Future Plans for expansion of the MANE dataset 577

Board 1077W: Measuring the accuracy of electronic health record-based phenotyping in the All of Us Research Program using pathogenic *BRCA1* and *BRCA2* carrier frequencies in putative ovarian cancer cases 579

Board 1078W: MedGen at NCBI: Empowering Biomedical Discovery and Clinical Care by Integrating Multiple Ontologies 580

Board 1079W: Mucopolysaccharidosis Type VII (MPS VII): A novel, online *GUSB* gene variant database 581

Board 1080W: Multi-organ imaging-derived polygenic indexes for brain and body health 582

Board 1082W: NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS): 2024 Update 583

Board 1083W: Optimization of Insert Sizes for Illumina WGS with a Combination of Covaris AFA based Shearing and Bead-Based Size Selection 584

Board 1084W: Pathogenic variants of pulmonary hypertension associated genes in COPDGene 585

Board 1085W: PhenoGemini: Enhancing Molecular Diagnoses of Mendelian Disorders through Identifying Twin Patients with Large Language Models ⭐ 586

Board 1086W: Phenome-wide association study of *MTHFR* variant in UK Biobank and *All of Us* 587

Board 1087W: Prevalence of errors in lab-made plasmids across the globe 588

Board 1088W: Rapid Point-of-Care Genotyping for Sickle Cell Anemia Using CRISPR Technology 589

Board 1089W: Taiwan View: A multiomic visualization tool for Taiwan Biobank data. .. 590

Board 1090W: The *All of Us* Researcher Academy Internship Program: An Experiential Learning Opportunity Designed to Support Research Skills Development of Diverse Early-Stage Researchers 591

Board 1091W: The Alzheimer's Disease Sequencing Project Follow Up Study ADSP-FUS: *APOE* genotype status and demographic characteristics across African and African American datasets..... 592

Board 1092W: The BIG initiative's contributions to inclusive genomics 593

Board 1093W: The Ensembl Variant Effect Predictor: enhanced annotation for human variant interpretation 594

Board 1094W: The impact of a research resource: A 5-year review of peer-reviewed publications utilizing samples from the NIGMS Human Genetic Cell Repository 596

Board 1095W: The impact on clinical success from the 23andMe cohort 597

Board 1096W: Unraveling the genetic predisposition to non-syndromic Biliary atresia through genomic investigations 598

Board 1097W: Updates from the frontline of international genomic data sharing 598

Board 1098W: Validation of the variant classification rules for ornithine transcarbamylase, N-acetylglutamate synthase, and carbamylphosphate synthetase 1 599

Board 1099W: VarChat: the genAI-based genomic assistant available as an open-access platform..... 601

Session Title: Mendelian Phenotypes Poster Wednesday Session 602

Board 2002W: A Case Series of Pancreatitis in Patients with Poorly Controlled Classic Homocystinuria 602

Board 2003W: A founder variant in tubulin binding cofactor B is associated with autism spectrum and hereditary spastic paraparesis 603

Board 2004W: A homozygous frameshift mutation in *TRDN* causes a Cardiac Arrhythmia Syndrome in a Malian family 604

Board 2005W: A mechanism underlying anophthalmia by MAB21L1^{R51W} de novo mutation 605

Board 2006W: A Novel *De Novo KDM3B* Mutation in a Male Patient with Evolving Diets-Jongmans Syndrome Features 606

Board 2007W: A novel in-frame deletion of HNF1B in a family presenting with renal cysts and maturity onset diabetes of the young 607

Board 2008W: A novel *PPARG* germline variant segregating with familial lymphedema: challenges in disease gene discovery at the single-family level 608

Board 2009W: A novel*CUL7*variant in a 3-M syndrome Turkish case 609

Board 2010W: A null variant in the voltage-gated calcium channel beta-1 subunit (*CACNB1*) causes autosomal recessive muscular dystrophy 610

Board 2012W: A rare case of Knobloch Syndrome in a consanguineous Malian family 611

Board 2013W: A scan for genomic disorders of infertility in 1500 cases of spermatogenic failure 612

Board 2014W: Accuracy of rare diseases inference in generative AI (ChatGPT3.5): comparison with in-house AI diagnostic supporting system 613

Board 2015W: An innovative approach for identifying low-penetrance structural birth defect genes 614

Board 2016W: An update on microduplications in 17q21.31 region potential cause of *KANSL1* gene disruption 615

Board 2017W: Beta-thalassemia in India: A decade of molecular diagnostics in a tertiary care centre..... 616

Board 2018W: Biallelic variants in *LLGL1*, a hippo-yap pathway regulator, are a candidate cause for a rare neurodevelopmental syndrome 617

Board 2019W: *CAMK2D* and isolated dilated cardiomyopathy: genotype-phenotype expansion and functional analyses support an association with disease 618

Board 2021W: Complete chromosome 21 centromere sequences from a Down syndrome family reveal size asymmetry and differences in kinetochore attachment ⭐ 619

Board 2023W: *De novo* variants in *CDKL1* and *CDKL2* are associated with childhood-onset epilepsy, developmental delay, and poor motor coordination 620

Board 2024W: Deficiency of lncRNA *CHASERR* upregulates chromatin remodeler CHD2 and causes global transcriptomic changes in a novel neurodevelopmental disorder ⭐ 622

Board 2025W: Defining the molecular mechanism of *NARS1*-mediated dominant neurological disease..... 623

Board 2026W: Development of patient-friendly diagnostic method for pediatric mitochondrial cases in India..... 624

Board 2027W: Diagnostic yield of whole genome sequencing for patients with neurodevelopmental delays 624

Board 2028W: DNA methylation episignature analysis is useful for reclassification of a maternal inherited missense variant with uncertain significance in the *KDM5C* gene .. 626

Board 2029W: *Drosophila* models of *ABCD1*-related X-linked Adrenoleukodystrophy show possible implications in peroxisomal biogenesis 627

Board 2030W: *Drosophila* models of HNRNPH2-related neurodevelopmental disorder variants demonstrate developmental defects 628

Board 2031W: Elucidating the role of *de novo*structural variations in a Mendelian disease cohort consisting of 169 families 629

Board 2032W: Establishing a comprehensive registry and natural history study for Pol II-associated developmental disorders 630

Board 2033W: Estimating the prevalence of rare Mendelian diseases in a founder population, using whole genome sequences from the CARTaGENE Biobank in Quebec 631

Board 2034W: Examination of a novel biomarker for pyridoxine-dependent epilepsy, 6-oxo-pipecolate, and its potential implications for screening and treatment 632

Board 2035W: Exploring craniofacial variation in a background mouse population to better understand the genetic architecture of Mendelian diseases, using achondroplasia as a case study 633

Board 2036W: Exploring molecular mechanisms of GABA dysregulation in *SLC6A1*-related neurodevelopmental disorders 634

Board 2037W: Functional Validation of *HMCN1* as a Candidate Gene for Childhood-Onset Essential Hypertension 635

Board 2038W: Gene variants identified in pediatric patients with Hypermobile Ehlers-Danlos Syndrome (hEDS) 636

Board 2039W: Genetic Architecture of 22,125 Deafness Subjects across China 637

Board 2041W: Genetic models and *ASTN1* variant contributions to neurodevelopmental disorders via deficits in astrotactin-mediated neuronal migration 638

Board 2042W: Genetic variants in *DDX53* contribute to Autism Spectrum Disorder associated with the Xp22.11 locus. 639

Board 2043W: Genotype/phenotype correlation of MESP2 gene: A 20 years Cohort 640

Board 2044W: Genotypic and Phenotypic Spectrum of maple syrup urine disease: A single center Study 641

Board 2045W: Homozygous*CAPZA2*start-loss variant associated with developmental delay, epilepsy, and infantile hypertrophic cardiomyopathy 642

Board 2046W: *HUWE1* gene, a prominent player in intellectual disability: a case-series report 643

Board 2047W: Hyperinsulinism associated with non-coding *HK1* variants 644

Board 2049W: Long-read RNA sequencing reveals Parkinson's disease-associated transcription in oligodendrocytes ⭐ 645

Board 2050W: Maternal Gonosomal Mosaicism Causing Nail-Patella Syndrome: A Case Report and Implications for Genetic Counseling 646

Board 2051W: Molecular insights into the pathogenesis of Chediak-Higashi syndrome and the biology of Lysosomal Trafficking Regulator 647

Board 2052W: Multiple approach modeling of Hypermobile Ehlers-Danlos Syndrome to elucidate a genetic etiology 648

Board 2053W: New candidate genes associated with the development of Pulmonary Arterial Hypertension 649

Board 2054W: New HSD17B4 gene mutation associated with D-bifunctional protein deficiency in a Puerto Rican patient 650

Board 2055W: Noncanonical endosomal trafficking mediates GLUT10 ER to mitochondrial targeting to regulate vitamin C homeostasis ⭐ 651

Board 2059W: Novel *TAF8* variants in three patients with neurodevelopmental disorder with severe motor impairment, absent language, cerebral hypomyelination and brain atrophy expand the genotypes and phenotype associated with this disorder 652

Board 2060W: Phelan-McDermid syndrome caused by a single nucleotide deletion: A case study 653

Board 2061W: Population-level biobank sequencing reveals insights into *C9orf72*-associated repeat expansion disorder epidemiology, penetrance and clinical presentation ⭐ 654

Board 2062W: Prevalence of heart disease among adults with recurrent pathogenic copy-number variants 655

Board 2063W: Rare maternal inheritance of *ARFGEF1* variant associated developmental disability: A case report 656

Board 2066W: Retrospective Analysis of Diagnostic Yield of Genetic Testing for Hereditary Connective Tissue Disorders 657

Board 2067W: RNF2-dependent ubiquitination is critical for transcriptional plasticity and neural development 658

Board 2068W: The analysis of the *MCPH1* gene across curated databases suggest phenotypic expansion 659

Board 2069W: The impact of copy number variants in diagnosis and severity of autosomal dominant polycystic kidney disease 660

Board 2070W: The p.T1163M variant in *TTBK2* causes ciliary dysfunction 661

Board 2071W: *TNXB* mutations in Hypermobile Ehlers Danlos Syndrome and Hypermobility Spectrum Disorder 662

Board 2073W: Understanding neurobehavior in TANGO2 Deficiency Disorder and the role of B vitamins ⭐ 663

Board 2074W: Understanding the genetic contributions to Cerebral Palsy: A comprehensive analysis of the complement system in neuronal development and plasticity pathway 665

Board 2075W: Unlocking Diagnoses: Integrating Whole Genome Sequencing, RNA-Seq and collaborative multi-site research in the Indiana University Undiagnosed Rare Disease Clinic (URDC) cohort 665

Board 2076W: Unraveling the brain-imaging predictors for cerebral palsy: insights from mendelian randomization analysis of brain imaging-derived phenotypes 667

Board 2077W: Unraveling the pathogenesis of SAVI-associated pulmonary fibrosis: transcriptional investigation of iECs and the role of EndMT 668

Board 2078W: Untargeted metabolomics as a potential screening tool for 3-methylglutaconic aciduria syndromes 669

Board 2080W: Von-Hipple Lindau Syndrome - A case report from India 670

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session 671

Board 3041W: A *de novo* deep intronic variant in *PKD1* adjacent to a common polymorphism creates a novel splice site in an individual with polycystic kidney disease 671

Board 3042W: A deep dive comparing the diagnostic yield and cost of clinical actionable findings from exome sequencing and multigene panel testing for epilepsy 672

Board 3043W: Classes of variants contributing to the increased diagnostic yield of whole genome and RNA sequencing 673

Board 3044W: Classification of meiotic/mitotic origin of chromosome 21 nondisjunction error in Down syndrome using proband-only sequencing: Methodology and application in a cohort with acute lymphoblastic leukemia 674

Board 3046W: Deciphering the genetic mystery of Congenital Nephrotic Syndrome ... 675

Board 3047W: Deciphering the genetic terrain of dilated cardiomyopathy: A study of 100 cases from the Indian sub-continent ⭐ 676

Board 3048W: Delving into the abyss: Integrating multi-omics and targeted functional studies to inform variants of uncertain significance 678

Board 3049W: Development and assessment of bioinformatics approach for whole genome sequencing-based newborns screening in the Generation Study ⭐ 679

Board 3050W: Development and Scaling of a Sensitive NGS-based Assay for Clinical Diagnosis of Acute Meningitis and Encephalitis 680

Board 3051W: Diagnostic yield of rare genetic disorders applied genome sequencing as a first-line test among 2,188 rare disease patients in Korean population 681

Board 3052W: Evaluation of Illumina Complete Long Reads for Genome-Wide Variant Discovery 682

Board 3053W: Genome sequencing versus exome sequencing in undiagnosed pediatric genetic diseases: An updated systematic literature review and meta-analysis of diagnostic and clinical utility 683

Board 3054W: How to gain a sex chromosome: Monosomy X to the rescue 684

Board 3055W: Identification of Highly Pathogenic Avian Influenza Clade 2.3.4.4b 685

Board 3056W: Impact of Genome boards on the rate of molecular diagnosis for patients with neurodevelopmental and neurological disorders: An update and further developments. 686

Board 3057W: Increasing epilepsy-related diagnostic yield from short-read genome sequencing 687

Board 3058W: Integrating emerging data into genomic testing: Outcomes from an evidence-based reanalysis initiative ⭐ 688

Board 3059W: International nomenclature for genomics 689

Board 3060W: Long-read genomic sequencing reveals expanded GAA-GGA chimeric alleles in Friedreich ataxia 690

Board 3062W: Molecular diagnosis of patients with optic atrophy<:> a multimodal approach increased the diagnostic rate. 691

Board 3063W: Molecular testing for patients with congenital limb anomalies: yield analysis and new candidate genes 692

Board 3064W: mRNA-based analysis of non-coding variants increases the diagnostic yield of clinical whole genome/exome sequencing for rare diseases 693

Board 3065W: On the (ir)reproducibility of EpiSign, a methylation-based clinical assay 694

Board 3066W: Project FIND-OUT: Protocol and pilot study for whole genome sequencing in symptomatic infants 695

Board 3067W: Resolving complex genotypes in residual clinical samples with long range-PCR and nanopore sequencing assay 697

Board 3068W: Revisiting the genetic basis of DFNB9 through cryptic variant exploration and haplotypic phasing 698

Board 3069W: Specification of frequency criteria for secondary findings genes to improve classification concordance 699

Board 3071W: The Face of *RNU4-2* Neurodevelopmental Disorder, a New, Common and Recognizable Syndrome 700

Board 3072W: The impact of a machine learning framework in reducing uncertainty and increasing equity in clinical genetic testing 701

Board 3073W: The parental origin of genomic imbalances found in patients with mental retardation and developmental delay 702

Board 3074W: The Pitfalls of Whole Exome Sequencing Analysis in Clinical : A Single Medical Center Experience 703

Board 3075W: Uncovering hearing impairment-related structural variants via integrating long-read and short-read whole-genome sequencing 704

Board 3076W: VariantMatcher: phenotypic and genomic data sharing to facilitate variant classification and disease gene discovery 705

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session 706

Board 6002W: Genetic association of Glucose-6-phosphate dehydrogenase variants and Pre-eclampsia (PET) in Ghana ⭐ 706

Board 6003W: A data-driven map of human metabolism across the allele frequency spectrum ⭐ 707

Board 6005W: A summary statistics-based toolset for identifying *trans*—regulation of proteomic pathways by cancer-related variants in the UK Biobank 708

Board 6006W: Accurate HLA genotype inference in single cell RNA sequencing data by integrating allele transcript abundance with low-pass genotyping 710

Board 6007W: Ageing and age-related diseases are accompanied by longitudinal changes in alternative splicing dysregulation ⭐ 711

Board 6008W: Allele specific transcription factor binding across human brain regions gives mechanistic insight into eQTLs ⭐ 712

Board 6009W: An atlas of the genetic basis of alternative tandem transcriptional initiation for human diseases ⭐ 713

Board 6010W: Annotation database updates and their impact on missense variant effect predictions..... 714

Board 6011W: Applying high throughput deep mutational scanning to study transcription factorsimplicated in the biology of Substance Use Disorders 715

Board 6013W: BMI Interacts with the Genome to Impact Gene Expression Globally, with Emphasis in the Brain and Gut 716

Board 6014W: Can biobanks be leveraged to help interpret variants of uncertain significance? 717

Board 6015W: CC2D1A causes ciliopathy, intellectual disability, autism, heterotaxy, renal dysplasia and abnormal CSF flow 718

Board 6016W: cciQTLs: Linking Human Genetic Variation to Intercellular Communication in the Brain 719

Board 6018W: Comprehensive Mosaic Variant Characterization Across Human Tissues 720

Board 6019W: Copy Number Variation and Haplotype Analysis of 17q21.31 Reveals Increased Risk Associated with Progressive Supranuclear Palsy and Gene Expression Changes 721

Board 6020W: Correlation of APOE polymorphism and its expression level with cardiac comorbidity in HIV-1 lipodystrophy patients 722

Board 6021W: Critical assessment of missense variant effect predictors on disease-relevant variant data 723

Board 6022W: Critical role of the T-box Glu86 residue in TBX4 in human lung development. 724

Board 6023W: Decreased expression of Presenilin-1 gene in homozygous G206A carrier brain as a potential mechanism for rapid or accelerated cognitive decline 725

Board 6024W: Discovering a delayed adipogenesis signature in visceral fat linked to metabolic dysfunction-associated steatohepatitis 726

Board 6025W: Efficient count-based models improve power and robustness for large-scale single-cell eQTL mapping 728

Board 6026W: Elucidating the ancestry-specific architecture of expression quantitative trait loci in Alzheimer's Disease 729

Board 6027W: Enhanced Classification of Protein-Truncating Variants Reveals the Need for Improved Predictive Tools 730

Board 6028W: Enhancer variant screens at major QT interval regulating genes 731

Board 6029W: Evaluating the Association of ROCK1P1 Polymorphism in Glaucoma Subtypes in a Nigerian Cohort 733

Board 6030W: Examining the Role of *SORL1* in AD-Specific Microglia and Neuron Endolysosomal Dysregulation 734

Board 6032W: Functional analyses for the late-onset Alzheimer's disease-risk variants of *SHARPIN* using CRISPR/Cas9-mediated knock-in cells 735

Board 6033W: Functional characterisation of genetic variants affecting adipocyte maturation and insulin activity in SGBS cell model 736

Board 6034W: Functional characterization of Hartnup disease missense variants in *SLC6A19* shows loss of function by disparate mechanisms 737

Board 6035W: Functional characterization of *RECQL4* missense variants: A protein level-based approach 738

Board 6036W: Future-proofing variant classification in *KCNH2*: The industry experience 739

Board 6037W: Genetic mapping in sexual planarians to identify loci involved in reproductive aging rate 740

Board 6038W: Genetic regulation of the targeted plasma proteome in children and adolescents reveals obesity-, puberty-, and age-specific differences 741

Board 6039W: Genetic Variation in *KIF1A* Gene Associated to Neuronal Disorders 742

Board 6040W: Genetic variation in the immune response to COVID-19 vaccination reveals vaccine-specific associations for extremes of antibody response in the UK population involving HLA and non-HLA loci 743

Board 6041W: Genetic variation influencing brain oxygen stress responses contributes to disease 744

Board 6042W: GWAS for Alzheimer Disease Using Genetic Similarity Clusters 745

Board 6043W: HHIP protein interactions in lung cells provide insight into COPD pathogenesis 747

Board 6044W: High-throughput identification of the impact of genetic variants on splicing in the context of the endogenous genome 748

Board 6045W: Human MASLD associated *GCKR P446L* promotes an increase in fibroblast growth factor 21 via up-regulation of transcription factor ChREBP in Huh-7 cell line 749

Board 6046W: IBD-focused eQTL analysis identifies novel GWAS target genes using diseased colon tissue 750

Board 6047W: Identification & validation of intronic splice mutations in a cohort of inherited retinal diseases 751

Board 6048W: Identification of sex-specific Alzheimer's disease causal genes via proteome-wide association studies ⭐ 752

Board 6049W: Identification of *STRC* structural variations in patients with hereditary hearing impairment using long-read sequencing 753

Board 6050W: Identifying diabetes-associated regulatory variants in HepG2 cells using massively parallel reporter assays ⭐ 754

Board 6051W: Identifying genetic variants associated with cell state abundance in T cells ⭐ 756

Board 6052W: Identifying suppressors of stress-induced neurodegeneration in a *C. elegans* knock-in *sod-1 G85R* ALS model 757

Board 6053W: Immunogenetic pleiotropy and context-specific gene regulation in the response to *Yersinia pestis* 758

Board 6054W: Insights into immunogenetic variability in the IgG heavy chain in COPD 759

Board 6055W: Investigating the bi-directional regulatory effects of rs174557 and rs968567 on *FADS1* and *FADS2* expression 760

Board 6056W: Investigating the effect of APOE variants in iPSC-derived astrocytes model of Niemann-Pick disease, type C1 761

Board 6057W: Investigation of shared and divergent contextual effects on overlapping type 1 and type 2 diabetes variants in steady state, hyperglycemic, and proinflammatory human pancreatic beta cells 762

Board 6058W: *KANSL1* contributes to the gene regulatory activity of the 17q21.31 locus in the context of Parkinson's disease 763

Board 6059W: Large-scale protein variant characterization of the STIM1 calcium binding cEF-hand reveals residues critical for calcium influx 764

Board 6061W: Loss-of-function variants in *CUL3* cause a syndromic neurodevelopmental disorder ⭐ 765

Board 6062W: MAGPIE accurate pathogenic prediction for multiple variant types using machine learning approach 766

Board 6063W: Metabolomic profiling reveals distinct and common biological pathways in COPD and IPF 767

Board 6064W: Mis-ascertainment of human gene regulatory architecture is caused by evolutionary history, but can be improved by comparison with farm species 768

Board 6065W: Mitochondrial DNA heteroplasmy improves risk prediction for myeloid neoplasms 769

Board 6066W: Molecular Characterisation of Filipino patients with Tyrosinemia Type 1 detected via Expanded Newborn Screening 770

Board 6067W: Multi-ancestry frontal cortex transcriptomic imputation models show variable predictive performance in single ancestry populations 771

Board 6068W: Multi-ancestry GWAS and TWAS implicate cell adhesion in anxiety related phenotypes 772

Board 6069W: Multi-omics integration deciphers genetic influences of osteoporosis and the role of *AZIN1* in bone homeostasis ⭐ 773

Board 6070W: Multi-omics integration in Caribbean Hispanics identifies neuronal and cerebrovascular pathways in Alzheimer's disease 774

Board 6071W: Optimizing a primary bronchial epithelial cell culture model to map air pollution exposure response eQTL 776

Board 6072W: OSBPL11 is an important regulator of cholesterogenic genes in murine hepatocyte cell line 777

Board 6073W: P selectin: A possible biomarker for vaso-occlusive crisis in Sickle Cell Disease 778

Board 6074W: Population-scale non-coding variant effect predictions using an in-silico reporter assay 779

Board 6075W: Premarital screening and genomic counseling in the UAE enabled by the Emirati Genome Program 780

Board 6076W: Prequalification of genomic newborn screening for severe childhood genetic diseases through federated training based on purifying hyperselection 781

Board 6077W: Protein-protein interaction explains *trans* regulation of the proteome and complex trait associations ⭐ 782

Board 6078W: SARs-CoV-2 Vaccination Antibody Kinetics: Genetic basis for differences in humoral immune response half-life in the UK-basedCOVIG study cohort 783

Board 6079W: Sex differences in the genetic architecture of circulating metabolites and disease risks 784

Board 6080W: Single-cell analyses with host genetics reveal the cell state-dependent genetic regulation of transcriptional profiles and T and B cell receptors repertoire in immune cells from 234 Japanese including COVID-19 patients and healthy subjects ⭐ 785

Board 6081W: Single-Cell Genomic Dissection of Human Genetic Control of Alternative Splicing in Response to Various Influenza Strains 786

Board 6082W: Single-cell RNA-seq analysis uncovers transcriptional signatures in pancreatic islet-like cells derived from iPSCs isogenic for the *ABCC8* R1420H loss-of-function variant 787

Board 6083W: Single-nuclei profiling of mRNA alternative polyadenylation in Alzheimer's disease 788

Board 6084W: Tissue-specific protein quantitative trait loci: implications for drug discovery in cardiometabolic disease 789

Board 6085W: TLR4 gene Polymorphism and biochemical markers as a tool to identify risk of Osteoporosis in women from Karachi 790

Board 6086W: Transcriptome-wide characterization of genetic perturbation atlases across cell types and dosage levels 791

Board 6087W: Transgenic denAsCas12a-CRISPRa/i mouse models for combinatorial in vivo perturbations 792

Board 6088W: UBE2A mutations linked to X-linked intellectual disability type Nascimento produce variable effects on protein stability and stress resistance 793

Board 6089W: Uncovering Genetic Factors Influencing Gut Microbiome Composition in the Korean Population: A Genome-Wide Association Study 794

Board 6090W: Using bulk RNA-sequencing analysis to study the effects of rs174557 and rs968567 for gene expression 795

Board 6091W: Using co-expression networks of transcriptome, translatome, and proteome to dissect the risk of schizophrenia 796

Board 6092W: Validation and development of functional assays for *RYR1* variants 797

Board 6093W: Variant-aware training of machine learning models for predicting the effects of genetic variants on chromatin accessibility..... 799

Board 6094W: Whole exome sequencing of siblings identifies a novel variant p.A399fs in *ANOS1* gene associated with Kallman syndrome 800

Session Title: Omics Technologies Poster Wednesday Session 800

Board 1100W: A genome-wide association study of mass spectrometry proteomics using the Seer Proteograph platform 800

Board 1101W: A statistical framework to model and infer gene networks from genome-scale genetic perturbation data 801

Board 1102W: Accelerating interpretation of rapid whole genomes and diagnosis in the intensive care unit 802

Board 1103W: An automated variant verification system to improve reporting efficiency 804

Board 1104W: An Integrated RNA-Seq Pipeline for Concomitant Human Genetic and Microbial Analyses in Persistent Cutaneous Infections 805

Board 1105W: An interactive web application for exploring human plasma and fibroblast metabolomics data from patients with inborn errors of metabolism 806

Board 1106W: Assessing the readiness of Oxford Nanopore sequencing for clinical genomics applications 807

Board 1107W: Association of sitting time and physical activity with gut microbiome... 808

Board 1108W: B cell immune repertoire sequencing in smoking, vaping, and chronic obstructive pulmonary disease in the COPDGene cohort 809

Board 1109W: Best practices Terra workspace for processing PacBio long reads sequenced on the Revio system: Single sample, trio, and cohort modes..... 810

Board 1110W: Brain RNA degradation is highly regulated in Alzheimer's disease 811

Board 1111W: Building large-scale population databases of structural variant calls from low- and high-coverage long-read sequencing data 812

Board 1112W: Clinical and protein impact of human loss-of function mutations in British Bangladeshi and Pakistani participants of the Genes and Health Study 813

Board 1113W: Clinical Validation of a Haplotyped Whole Genome Sequencing Assay Using PacBio HiFi Sequencing 814

Board 1114W: ClioMD: an artificial intelligence model for ciliopathies 815

Board 1115W: Cohort-level differential distributional analysis for studying microglia in Alzheimer's disease via single-cell RNA-sequencing 816

Board 1116W: Community engagement efforts to decrease barriers to access to the undiagnosed diseases network for populations facing health disparities 817

Board 1117W: Comparative analysis of cell type annotation methods for single-cell RNA-seq 818

Board 1118W: Comparison of Five different library prep kits for mRNA sequencing optimization from low input and degraded FFPE samples 819

Board 1119W: Contamination detection in NGS libraries with mitochondrial DNA: Limits of detection and effects on library performance 820

Board 1121W: Cross-modal machine learning enables researchers to easily leverage public transcriptomics samples for data-driven discovery of biomedical contexts associated with their genes of interest 821

Board 1122W: Cross-validation of next-generation sequencing for genetic stability assessment in cell therapy products 822

Board 1123W: Curated loci prime editing (cliPE) for accessible multiplexed assays of variant effect 823

Board 1124W: Delineating intercellular and intracellular gene regulation from spatial genomics data with SLEGRN 824

Board 1125W: Delineating the spatial molecular landscape of human heart transplant rejection using longitudinal cardiac biopsies 825

Board 1126W: Differential depth targeted gene panel sequencing to simultaneously detect somatic and germline variants in saliva-derived DNA 825

Board 1127W: Dynamics of age-dependent alterations in the human cardiac transcriptome at single-nucleus resolution 827

Board 1128W: Empowering Biological Research: An Integrated Ecosystem of Tools in the Reactome Pathway Knowledgebase 828

Board 1129W: Enabling spatial transcriptomics to reveal gene expression landscapes with near single-cell resolution in large-sized tissues with iStar-XL 829

Board 1131W: Expression, splicing or protein QTLs have substantial contributions to SNP-based heritability and polygenic prediction in complex traits ⭐ 830

Board 1133W: Gene expression profiling of peripheral blood mononuclear cells from women with cervical lesions reveals new markers of cancer. 831

Board 1134W: Genetic Architecture of plasma metabolites in the Southeast Asian population ⭐ 832

Board 1135W: Genetically-predicted protein levels in UK Biobank and NIH *All of Us* identify thousands of associations of proteins and clinical phenotypes 833

Board 1136W: Haplotype reconstruction using low-coverage genotyping-by-sequencing in genetically diverse mice 834

Board 1137W: Harmonized framework for RNA-seq-based rare disease diagnostics in a pan-continental consortium - Solve-RD 835

Board 1138W: High throughput ancient DNA sequencing using the Ultima Genomics platform ⭐ 836

Board 1139W: High throughput single-cell TCR sequencing method on aging mouse T cells and antigen stimulation in human T cell powered by split pool combinatorial barcoding technology 838

Board 1140W: Identification of non-coding RNAs as early prenatal screening biomarkers for congenital heart diseases by integrated analysis of Omics and clinical data..... 839

Board 1141W: Improvements in Combinatorial scRNA Sequencing that Enable Scaling to Hundreds of Samples and Millions of Cells in a Single Experiment 839

Board 1142W: Inferring phenotypes from tabular genotype data with pre-trained Large Language Model ⭐ 840

Board 1143W: InFuse: a comprehensive framework for novel isoform and gene fusion detection from long-read RNA sequencing ⭐ 841

Board 1144W: Integrated Analytical Biobanking as a foundational component of NIH research in Neuroscience 842

Board 1145W: Integrating genomics, transcriptomics, and epigenomics for rare disease diagnosis 843

Board 1146W: Integration of proteomics and CpG methylation sites identifies novel biomarkers and biology of early-stage heart failure 844

Board 1147W: Integrative Machine Learning Approaches for ASD Prediction using Gut Microbiome Analysis 845

Board 1148W: Isolation of single nuclei from cryosections of human cornea 846

Board 1149W: Learning a sequence motif syntax from Transformer model attributions with SEMANTIC 847

Board 1150W: Leverage generative GNN model to explore cellular heterogeneity in Perturb-seq data 848

Board 1151W: Long Reads and Short Reads and Methylation Oh My: Options for low input/FFPE samples 849

Board 1152W: Long-read RNA sequencing reveals isoform diversity in human immune cells and implications for diseases 850

Board 1153W: Massively parallel single-molecule protein quantitation using protein identification by short-epitope mapping (PrISM) 851

Board 1154W: Measuring enhancer-enhancer cooperativity using single-cell CRISPR perturbations and single-cell multiome sequencing data 852

Board 1155W: Medical Action Ontology (MAxO) enables prioritization of diagnostic procedures to support rare-disease differential diagnosis 853

Board 1156W: Meta-analysis of RNAseq data from two large atrial tissue repositories identifies more than 8000 transcripts associated with atrial fibrillation and suggests a prominent role for mitochondrial ADP/ATP processing: the CATCH ME and RACE V Consortia 854

Board 1157W: Modification of an early barcoding bulk RNA-sequencing method to capture long non-polyadenylated transcripts 856

Board 1158W: More accurate and faster expression and splicing methods improve DROP - the Detection of RNA Outliers Pipeline 857

Board 1160W: New Genome In A Bottle HG002 T2TQ100 diploid assembly-based small and structural variant benchmark sets. 858

Board 1161W: New quantitative and qualitative analytical framework of scRNAseq data reveals the pathophysiology of systemic lupus erythematosus ⭐ 859

Board 1162W: Omics Wide Association Analysis Platform (OWAS) to identify personalized genetic risk in CVD 860

Board 1163W: Paired single-cell protein and RNA comparison in developing human brain reveals biological gap in current single-cell research 861

Board 1164W: Plasma metabolites are associated with bone density, microarchitecture, and bone strength; The Framingham Heart Study 862

Board 1165W: Plasma protein signature as easily accessible biomarkers to predict future MCI to dementia conversion 863

Board 1166W: ProGenAC: Capturing deep phylogeny through the hyperbolic embedding space of a Variational AutoEncoder 864

Board 1167W: Protein quantitative trait locus analysis in African American and non-Hispanic White individuals 865

Board 1168W: Proteoform quantitation to understand isoforms and protein heterogeneity using a single-molecule proteome analysis platform. 866

Board 1169W: Proteogenomic approaches to advance rare disease diagnostics: a pilot study in the 100,000 Genomes rare disease cohort 867

Board 1170W: Rare variants in non-*MECP2* genes modulating clinical outcome in Rett syndrome. 868

Board 1171W: Reconstructing refined spatial tissue architecture in single-cell RNA-seq with REPAIR. 869

Board 1172W: ResSAT: Enhancing Spatial Transcriptomics Prediction from H&E-Stained Histological Images with Self-Attention Transformer Networks 870

Board 1173W: Ribbit: Accurate identification and representation of complex tandem repeats in genomes 871

Board 1174W: ScRNAseq reveals altered stemness and reduced genesis of GnRH neurons in mutated *SMC3* patient-derived iPSCs 872

Board 1175W: Sensitivity and dynamic range of massively parallel single-molecule protein quantitation illustrate the capabilities of protein identification by short-epitope mapping (PrISM). 873

Board 1176W: Short tandem repeat genotyping in an immune disorder cohort 874

Board 1177W: Simultaneous profiling of 1M single-cell BCRs and whole transcriptomes from diseased and healthy B cells in a single experiment using split-pool combinatorial barcoding ⭐ 875

Board 1178W: Single-nucleus multiomics in three brain regions reveals the regulatory programs for sporadic early-onset Alzheimer's Disease 876

Board 1179W: Streamlining NGS Library Normalization: A Novel Method for Efficient and Flexible Equimolar Pooling 877

Board 1180W: SVTopo: enhance complex structural variation analysis from highly accurate long reads with human-readable images 878

Board 1181W: Telo-seq: Native, multiplexed telomere sequencing of individual chromosomes using 3' end adapter ligation ⭐ 879

Board 1182W: The blended genome and exome sequencing method captures genetic variation in an unbiased, high-quality- and cost-effective manner ⭐ 880

Board 1183W: The PRIDE Affinity-proteomics Archive (PRIDE-AP): Making affinity proteomics datasets FAIR. 882

Board 1184W: Tracking dynamic regulatory changes in rare diseases using single sample networks 883

Board 1185W: Uncovering New Disease Associated Cellular States Within and Across Cell Types Using residPCA in Single-Cell RNA and ATAC Sequencing Data 884

Board 1186W: Whole Genome Scanning of Deep Variants Existing in Human Tissues by Duplex-seq with Comprehensive Genome Coverage 885

Session Title: Pharmacogenomics Poster Wednesday Session 886

Board 7001W: A call from the stars in Pharmacogenetics: Accurate genotyping of *CYP2D6* and PGx relevant genes from next-generation sequencing data 886

Board 7002W: A methodology to assess concordance of last prescribed treatment with CPIC guideline recommendations: A case study application in a diverse All of Us research cohort 887

Board 7003W: Association of Chromosome 19 Variants with Nicotine Metabolite Ratio in American Indians: Insights from the Strong Heart Study 888

Board 7004W: Berberine signature and cardiometabolic diseases: a study leveraging randomized controlled trial, cohort study, and Mendelian randomization 889

Board 7005W: Characterizing drug-induced stuttering in biobank-linked electronic health records enables potential precision medicine approaches 890

Board 7006W: Clinical integration of NGS-based pharmacogenomics: Accurate genotyping and practical applications 891

Board 7007W: Design and development of a novel dPCR assay for determination of *CYP2D6* star (*) allele copy number variations. 892

Board 7008W: Distribution of *CYP2C8*2 and *3* in the South (Velingara Health District) of Senegal: Alleles involved in the metabolism of Artesunate Amodiaquine used in the treatment of malaria ⭐ 893

Board 7009W: Effect of a novel missense variant in TNF on TNF-alpha binding with receptor proteins 894

Board 7010W: Genetic variants in insulin pathway genes are associated with statin-induced diabetes 895

Board 7011W: Genome-wide association study identifies candidate genetic variants associated with postoperative opioid-related side effects (nausea/vomiting and sedation) in patients undergoing laparoscopic-assisted colectomy (LAC) 896

Board 7012W: Integration of transcriptome-wide association study of Alzheimer's disease with large-scale perturbation data for drug repositioning and new therapeutic approaches 897

Board 7013W: Large language models enable accurate and fully automated curation of discrete molecular pathway data 898

Board 7014W: Mitogen-activated protein kinase kinase 6 (MAP2K6) gene polymorphism is associated with TNF-α inhibitor treatment response in rheumatoid arthritis patients 899

Board 7015W: N-Acetylcysteine ameliorates loss of the electroretinogram b-wave in a Bardet-Biedl Syndrome Type 10 mouse model 900

Board 7016W: Patterns of New Pharmacogenomic Drug Prescriptions and Socioeconomic Disparities in Glasgow: Insights from Longitudinal Data Analysis 901

Board 7017W: Retrospective study shows pharmacogenomic testing could reduce adverse events associated with clopidogrel use by 38% ⭐ 902

Board 7018W: Searching for functional regulatory SNPs that regulate *CYP2C9* expression to improve *CYP2C9* biomarker panels 903

Board 7019W: Strong Heart Study: Serum nicotine metabolite ratio (NMR) biomarker informs nicotine metabolism variation in American Indians who smoke commercial cigarettes 904

Board 7020W: Structural Equation Modelling of Polygenic Risk and Environment in Late-Life Depression 906

Board 7021W: TAS-PGx: Targeted Adaptive Sampling-Long Read Sequencing for Enhanced Pharmacogenomics Profiling and Genome-Wide Variant Analysis 907

Board 7022W: Full coverage of Tier 1 Variants for Pharmacogenomics Studies 908

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session 909

Board 7025W: Amplicon-based next generation sequencing reveals low-level germline mosaicism for CNV deletion -discordance between somatic and germline mosaicism909

Board 7026W: Autosomal dominant mutations in the GEMINI male infertility cohort: Detection rates and genetic burden 910

Board 7027W: Cilia mutation and gene by environment interactions leading to increased phenotypic variance. 911

Board 7028W: Concomitant Prenatal Diagnoses of Jeune, Malan, and Sotos Syndromes in a Singleton Pregnancy 912

Board 7029W: Epigenetic regulator *Tet2* modulates ovarian dysfunction associated with Fragile X premutation 913

Board 7030W: Equitable testing and screening for cystic fibrosis in patients of diverse genetic ancestry 914

Board 7031W: Exploring the genetic architecture of Unexplained Pregnancy Loss and other Adverse Pregnancy Outcomes 915

Board 7032W: First cases of Hemogobin J in sub Saharan Africa 916

Board 7033W: Gene by Prenatal Alcohol Exposure Interaction Effects on Growth and Cognition in Mother-Child Dyads in a South African Birth Cohort..... 917

Board 7034W: Gene-environment interactions with maternal smoking and vitamin intake identified potential contributors for CL/P..... 918

Board 7035W: Genetic Control of Stage-Dependent Alternative Polyadenylation Events in the Developing Human Brain ⭐ 920

Board 7036W: Genomic Indicators of Female Infertility Using Transcriptome-Wide Association Studies in a European Population 921

Board 7037W: Identification of seminal plasma derived predictive biomarker microRNA in idiopathic non-obstructive azoospermia patients 922

Board 7038W: Identifying the causal role of human milk microbiota in childhood asthma using one- and two-sample Mendelian randomization..... 923

Board 7039W: Lifespan analysis of repeat expression in the neurotypical human brain reveals age-dependent upregulation of HERV-K 924

Board 7040W: Long-read whole genome sequencing for comprehensive genomic profiling of men with infertility ⭐ 925

Board 7041W: Multi-ancestral maternal GWAS meta-analysis of longitudinal trajectory of fetal growth ⭐ 926

Board 7042W: Placental gene expression signatures based on maternal parity 927

Board 7043W: Single nucleus molecular mechanisms associated with Caudal Regression Syndrome in mice 929

Board 7044W: Spatial transcriptomic analysis of the molecular drivers behind bronchopulmonary dysplasia 930

Board 7045W: The Australian Genomic Autopsy Study: Summary of outcomes from 414 families, and new diagnoses from systematic reanalysis and reclassification. 931

Board 7046W: Unraveling mechanisms of protamine-based packaging in the sperm genome 932

Board 7047W: Validating the addition of 39 ACMG-recommended genes to a 235-gene carrier screening panel 933

Board 7048W: Validation of fetal and maternal recessive disease genotyping with FirstGene: a combined, non-invasive prenatal cfDNA assay for fetal aneuploidy, recessive diseases, and serological screening 934

Board 7049W: Whole-genome sequencing of nonobstructive azoospermia patients identify rare predisposing regulatory variants ⭐ 935

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session 936

Board 4002W: A biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic epistasis 936

Board 4003W: A Multi-phenotype colocalization framework in*LocusFocus* 937

Board 4004W: A novel, scalable, and efficient blockLASSO PGS method in All of Us and the UK Biobank 938

Board 4005W: A structural mean modelling Mendelian randomization approach to investigate the lifecourse effect of adiposity: applied and methodological considerations 939

Board 4006W: Abundant parent-of-origin effect eQTL: the Framingham Heart Study ⭐ 940

Board 4007W: Accounting for confounding of different types and in large quantities in causal network inference with Mendelian randomization 941

Board 4009W: An Empirical Simulator of Gene Expression Profiles Capturing Regulatory Effects of Rare Genetic Variants 942

Board 4010W: An integration of local outlier factor analysis and ensemble machine learning algorithms predicts *MEFV* gene variants: Novel gene-specific voting classifier 943

Board 4011W: An interpretable matrix factorization approach for GWAS summary statistics robust to sample overlap 944

Board 4012W: Analysis of 3.6 million individuals yields minimal evidence of meaningful pairwise genetic interactions for height 945

Board 4013W: Analysis of ancestral impact on disease across Mendelian diseases ... 946

Board 4014W: Ancient selective pressures on metabolic traits contribute to obesity and weight-related diseases in modern populations 947

Board 4015W: Applications and Evaluations of Polygenic Risk Scores in Cancer Outcomes: Context Matters 948

Board 4016W: ASDecoder: a semi-supervised deep learning model for decoding autism risk genes 949

Board 4017W: Assessing the impact of social exposomes on genetic liability for bipolar disorder 949

Board 4018W: Association analysis of affected cancers with subtype of *TP53* gene pathogenic variants in Li-Fraumeni syndrome 951

Board 4019W: Association analysis of rare genetic variants in circulatory and respiratory phenotypes among participants from the *All of Us* Research Program ⭐ 952

Board 4020W: Associations of combined genetic and lifestyle risks with incident type 2 diabetes in the UK Biobank 953

Board 4021W: Bayesian network imputation methods applied to multi-omics data identify putative causal relationships in a type 2 diabetes dataset containing incomplete data: An IMI DIRECT Study 954

Board 4022W: Benchmarking gAIRR-Call in identifying *T Cell Receptor* alleles from whole genome sequencing data 955

Board 4023W: Benchmarking of missense variant effect predictors using known gene-trait associations 956

Board 4024W: Beyond the Liability Threshold: A Theoretical Framework for Genetic Disease Risk Based Upon Multiple Pathways ⭐ 957

Board 4025W: BinMap: A novel Method to Finemap Genetic Associations with Binary Traits with an Application to T2D 958

Board 4026W: CADET: Enhanced transcriptome-wide association analyses in admixed samples using eQTL summary data 959

Board 4027W: Causal relevance of adiposity for infectious diseases in diverse populations 960

Board 4028W: Comparison of Bayesian methods for rare variant association using whole exome sequencing and medical record data in the UK Biobank 961

Board 4029W: Coupling foundation models with chromatin QTL analysis reveals unique genetic-epigenetic associations that emerge in opioid overdose 962

Board 4030W: Deciphering the genetic structure underlying childhood asthma heterogeneity 963

Board 4031W: Deep learning estimation of lumbar vertebral and inter-vertebral morphology from the UK Biobank dual energy X-ray dataset 964

Board 4032W: Deep learning prediction of RNA binding proteins reveal variant-effect scores for Alzheimer's disease association studies 966

Board 4033W: Detecting and correcting discrepancies in summary statistics for genetic association studies 967

Board 4034W: Detecting interpretable, genetically-driven patterns of brain structure using a data-driven deep learning approach 968

Board 4035W: Dissecting the biological mechanisms behind type 2 diabetes causal effects on non-cardiometabolic comorbidities. 969

Board 4036W: ENPACT predicts personalized DNA methylation patterns from DNA sequences using a combined linear and deep-learning approach 970

Board 4037W: Estimating the effect of opioid prescribing on addiction with Mendelian randomization 971

Board 4038W: Estimating the length distribution of non-crossovers from gene conversions 972

Board 4039W: Evaluating the role of imputing gene expression in the brain for psychiatric phenotypes. 972

Board 4040W: Exploring the age-related cognitive impairment with data from Taiwan Biobank : A genome-wide association study 973

Board 4041W: Exploring the impacts of multi-generational environment sharing on phenotypic covariance 974

Board 4042W: Exploring the potential of protein domain-specific burden tests to unveil insights into protein function and association with disease risk 975

Board 4043W: Extending REGENIE's whole genome regression framework for increased power in large-scale association analysis 976

Board 4044W: Flexibly encoded genome-wide association study identifies novel nonadditive genetic risk variants for cardiometabolic traits 977

Board 4045W: From cortex to cortex: the shared genetic components of kidney function and Alzheimer's Disease 978

Board 4047W: Genetic and non-genetic factors impacting MRI-estimated liver iron in the iron overload genetic disease *HFE* hemochromatosis 979

Board 4048W: Genetic association meta-analysis is susceptible to confounding by between-study cryptic relatedness 980

Board 4050W: Genetic Basis of Partner Choice 981

Board 4052W: Genetic studies of liver fat and volume using whole-body MRI and extensive phenotyping provide metabolic insights 982

Board 4053W: Genetic Variants Associated with Lactase Persistence in the United Arab Emirates: Insights from the Emirati Genome Project 984

Board 4054W: Genetics of cardiovascular outcomes in individuals with chronic kidney disease: the Chronic Renal Insufficiency Cohort (CRIC) study 985

Board 4055W: Genome-wide machine learning analysis on Alzheimer's disease 986

Board 4056W: Genome-wide Search of Exonic Variants Affecting Translational Efficiency 987

Board 4057W: Haplotype analysis reveals pleiotropic disease associations in the HLA region 988

Board 4058W: Haplotype association using LOCATER improves trait association power at loci with allelic heterogeneity 989

Board 4059W: High polygenic risk assignment instability and mitigation 990

Board 4060W: iCLOUD: Utilizing family history for sex-difference analysis in genetic association study of Alzheimer's disease 991

Board 4061W: Identification of *trans* regulatory networks from multi-condition multi-omics quantitative trait loci with applications in Alzheimer's disease 992

Board 4062W: Identifying candidate risk and protective rare variant effects on Alzheimer's Disease using multi-omic variant prioritization 993

Board 4063W: Identifying genetic association with secondary phenotypes in a case-control study using tilted knockoffs..... 994

Board 4064W: Identifying loci with pleiotropic effects on orofacial clefts and facial morphology using conditional false discovery rate analyses 996

Board 4065W: Identifying Rare Variants Associated with hypermobile Ehlers-Danlos Syndrome Using a Case-Only Cohort and Biobank Controls: Overcoming Ancestry and Technical Differences in Separately Sequenced Samples 997

Board 4066W: IGEM: Novel software provides up-to-date biological knowledge for analyzing exposure-exposure interactions 998

Board 4067W: Impact of sex, ancestry, and local ancestry on Apolipoprotein E (APOE) risk for Alzheimer's disease (AD): a pooled case-control analysis of Alzheimer's Disease Genetics Consortium (ADGC)..... 999

Board 4068W: Inferring the evolutionary architecture of complex traits through the WF process..... 1000

Board 4069W: Inferringcell-type-specific methylation quantitative trait lociwith a hierarchical Bayesian interaction model: a simulation-based benchmark study..... 1001

Board 4070W: Integrating clinical and genetic data into graph attention networks for disease risk prediction using Biobank data 1002

Board 4071W: Integrating Perturb-seq and GWAS summary statistics identifies core genes for human complex disease ⭐ 1003

Board 4072W: Integration of Phenotype Similarity Embedding Model in Identifying Disease Genes Using Whole Exome Sequencing Data from Biobanks..... 1004

Board 4073W: Integrative genomic and transcriptomic investigation of human heart failure mechanisms - Trans-Omics for Precision Medicine in Congestive Heart Failure (TOPCHeF) ⭐ 1005

Board 4074W: Investigating mode of inheritance in rare disease using aggregated rare variant association on phenotypes derived from phenome-wide deep learning 1006

Board 4076W: Kinship estimation bias carries over to heritability estimation bias using variance components 1007

Board 4077W: Leveraging Mendelian randomization for target discovery from machine-learning predicted phenotypes 1008

Board 4078W: Limited overlap between genetic effects on disease susceptibility and disease survival: evidences from seven biobanks and nine common diseases 1009

Board 4079W: Local ancestry adjusted *cis*-eQTL mapping of TB-T2D comorbidity elucidates the involvement of African ancestry in TB susceptibility. 1010

Board 4080W: Mendelian Randomization Analysis of Latent Exposures Co-regulating Multiple Correlated Traits 1011

Board 4081W: Mendelian Randomization does not implicate MS risk factors as having a role in NMOSD susceptibility. 1012

Board 4082W: Meta Analysis of Rare Disease Cohorts Discovers Novel Disease-Causing Genes and Processes Governing Disease Prevalence 1013

Board 4083W: Metabolic and lipidomic profiles reveal crucial mediators and pathways influencing bone mineral density 1014

Board 4084W: Metabolite mediation of pathways from genotype to disease in the Canadian Longitudinal Study on Aging 1015

Board 4085W: MetaGLIMPSE: Meta Imputation of Low Coverage Sequencing Data for Modern and Ancient Genomes 1016

Board 4086W: MetaSCAMPI: A scalable variance- and covariance-based meta-analysis framework for genome-wide interaction testing utilizing multiple traits 1017

Board 4087W: *ML-MAGES*: a computationally efficient tool for multivariate genetic association analyses with effect size shrinkage via machine learning ⭐ 1018

Board 4088W: Model introspection techniques enable scalable variant effect prediction from genomic sequence models and improve mechanistically-driven functional fine-mapping 1019

Board 4089W: MRScan enables heritable confounder selection from public GWAS databases for bias reduction in multivariable Mendelian randomization 1020

Board 4090W: Multi-Ancestry Analysis Identifies Susceptibility Variants and Improves Polygenic Risk Scores for Breast Cancer Subtypes 1021

Board 4091W: Multi-ancestry gene expression prediction identifies novel genes regulating complex traits. 1023

Board 4092W: Multi-modal Polygenic Risk Scores and Functional Annotations Improve Glaucoma Prediction Across Ancestries 1024

Board 4093W: Multipopulation Genome Wide Association Study of Age at Menarche reveals potentially novel variants linked to TGF-β and ER stress 1025

Board 4094W: Multi-Trait GWAS Identifies Shared and Novel SNPs in Fibroproliferative Diseases Among Non-Hispanic Black (NHB) and White (NHW) Populations..... 1026

Board 4095W: Multi-trait polygenic risk score and all-cause and cardiovascular disease mortality -A cohort study of 14,086 Japanese individuals-..... 1027

Board 4096W: Multivariate SEM-PGS Model: Using Polygenic Scores to Investigate Cross-Trait Indirect Genetic Effects 1028

Board 4097W: Network biology algorithms reveal biological pathways underlying cigarette smoking behaviors..... 1029

Board 4098W: New-Onset Refractory Status Epilepticus (NORSE) in Individuals with POLG Gene Variants: A Systematic Review 1030

Board 4099W: No evidence of synthetic association across multiple complex traits . 1032

Board 4100W: Novel genome-exposome integration tool identifies gene-environment interactions for autism spectrum disorder and attention deficit hyperactivity disorder 1032

Board 4101W: Novel pipeline to genotype trisomic loci for the study of congenital heart defects in Down syndrome 1033

Board 4103W: Outlier rare-variants implicate CTCF-bound promotor like sequences upstream of core genes in children with high risk for obesity 1035

Board 4104W: PanTrio: Bespoke pangenome reference enables highly accurate genotyping of proband-parent trio sequencing samples 1036

Board 4105W: Pathway-level network-informed transcriptome-wide association study leveraging prior gene regulatory network information 1037

Board 4106W: PhenoLink: Using a fine-tuned large language model to extract HPO terms from natural language texts 1038

Board 4107W: Polygenic risk and progression prediction for longitudinal and familial data in large biobanks 1039

Board 4108W: Polygenic risk for poor lung function and its association with COVID-19 non-recovery outcomes 1040

Board 4109W: Pooled calling on rare noncoding variants in samples with low pass whole genome sequencing 1041

Board 4110W: Predicting Polygenic Score Accuracy with Neighbors in the Genetic Ancestry Continuum 1042

Board 4112W: Prioritization of mouse strains for modeling genetic variant effects in rare and common disease 1043

Board 4113W: pyGEMMA: A fast, scalable, and user-friendly Python/R implementation of linear mixed models for genome-wide association studies 1044

Board 4114W: Real-time dynamic polygenic prediction for streaming data 1045

Board 4115W: Reconstruct clonal relationship with mosaic variants at bulk and single-cell level 1046

Board 4116W: Reproductive factors and sex hormones levels on differentiated thyroid cancer risk A Mendelian Randomization study 1047

Board 4117W: Robust tests for association between polygenic scores and ancestry 1048

Board 4118W: Scoring familial risks in 7million Finns using a pedigree-based BLUP method provides a unique angle to perceive and predict human phenotypes ⭐ 1049

Board 4119W: Sex-specific genetic resource for identifying genes influencing neurological traits and disorders through expression captured by neuroimaging 1051

Board 4120W: Shared and unique 3D genomic features of substance use disorders across multiple cell types reveal a role for the pancreas in pathogenesis ⭐ 1052

Board 4121W: Single-subject omics approaches unravel pathway-specific heterogeneity in Alzheimer's disease 1053

Board 4122W: Sleep Deprivation and Stress are additively associated with shortened methylation based telomere length 1054

Board 4123W: Somatic CNVs in oncogenic and vascular malformation genes enriched in Cerebral Cavernous Malformations 1055

Board 4124W: Sparse Bayesian variable selection using global-local shrinkage priors for the analysis of cancer datasets 1056

Board 4125W: Statistical and population genetic approaches to enhance the discovery of trait-associated loci in understudied populations with small sample size ⭐ 1056

Board 4126W: Systematically quantifying gain- and loss-of-function biochemical consequences of missense variants using UK Biobank exomes ⭐ 1058

Board 4127W: The comparison of data-driven and knowledge-driven approaches to clustering a Mendelian Randomization instrument, using multiple methods 1059

Board 4128W: The Covariate Adjusted PERmutation (CAPER) tool for genetic association analysis 1060

Board 4129W: The cumulative effect of protein-truncating variants on human life expectancy and overall disease burden. 1061

Board 4130W: The impact of Indigenous American ancestry on the risk of acute lymphoblastic leukemia and the implications for future study designs 1062

Board 4131W: The Norwegian Childhood Diabetes Registry: clustering of pediatric diabetes cases 1063

Board 4132W: The PhenX Toolkit: New Protocols for Sickle Cell Disease Pregnancy, Medical Cannabis, Polysubstance Use and Recovery, and Updated Protocols for Tobacco Regulatory Research 1064

Board 4133W: The role of sleep in the human brain and body: insights from multi-organ imaging genetics 1065

Board 4134W: Trajectory analysis of heart failure-related phenotypes from biobank data uncovers novel genetic associations ⭐ 1066

Board 4135W: Transcriptome-wide association study using summary-level brain eQTL reference data to identify risk genes associated with adolescent neurocognitive function 1067

Board 4136W: Transfer-learning of PRS with pseudo-validation 1068

Board 4137W: Two decades of admixture mapping studies on human health: ancestry/sociodemographic diversity and insights 1069

Board 4139W: Two-sample Mendelian randomization with common and annotation-weighted rare variants 1070

Board 4140W: Unlocking the Power of Biobanks using GPUs - Accelerating GWAS with SAIGE on High Performance Computing and Clouds 1071

Board 4141W: Unsupervised phenotyping and genetic discovery in 62,495 cardiac MRIs using autoencoders ⭐ 1072

Board 4142W: Unveiling the Influence of Reference Sample Size on Ancestry Estimation of Population Structure. 1074

Board 4143W: Using an external reference panel and single-variant summary statistics for rare-variant aggregation tests for association 1074

Board 4144W: Using partial polygenic scores to detect gene-by-ancestry interactions in recently admixed populations ⭐ 1075

Board 4145W: VACANT: Variant Annotation Clustering AssociatioN Test 1076

Board 4147W: Validating the effects of genetic variants on protein level:Link between whole genome sequencing and proteomic data in UK Biobank 1077

Board 4148W: VINTAGE: A unified framework integrating gene expression mapping studies with genome-wide association studies for detecting and deciphering gene-trait associations 1079

Board 4149W: XPRS: A Tool for Interpretable and Explainable Polygenic Risk Score .. 1080

Thursday, November 7, 2024, 2:30 pm - 4:30 pm 1081

Session Title: Cancer Poster Thursday Session 1081

Board 8001T: A comparison of the diagnostic yield of comprehensive multigene panel testing with RNA analysis in hereditary cancer to smaller guidelines-based panels .. 1081

Board 8002T: A highly accurate risk factor-based multiethnic XGBoost model for identifying patients with skin cancer 1082

Board 8003T: A multi-omic approach to genetic variants in breast cancer: Insights from the Norwegian Breast Cancer Study 1083

Board 8004T: Analytical Evaluation of Whole Genome Sequencing for Acute Myeloid Leukemia 1084

Board 8005T: Beyond *RB1* : analysis of somatic copy number alterations in retinoblastoma 1085

Board 8006T: Biologically-targeted discovery-replication scan identifies G×G interaction in relation to risk of Barrett's esophagus and esophageal adenocarcinoma 1086

Board 8007T: Case report: a deep-intronic and splice-altering variant in the *ATM* gene 1087

Board 8008T: Characterization of 397T/C, 351A/G polymorphisms of ERα gene and 331G/A polymorphism of PR gene as genetic markers of breast cancer predisposition in Cameroon 1088

Board 8009T: Clinicogenomics data guide targeted cancer therapies for improved clinical outcomes 1089

Board 8010T: Cmai: predicting antigen-antibody interactions from massive sequencing data 1090

Board 8011T: Comprehensive analysis of rare genetic variants in hepatocellular carcinoma across diverse population groups 1091

Board 8012T: Cosegregation of *NF1* mutation and *PMS2* Structural Variant with Hodgkin Lymphoma- A Rare Event 1092

Board 8013T: Deciphering the genetic architecture of lung cancer survival: GWAS meta-analysis with 10K patients and integrative multi-omics study ⭐ 1093

Board 8014T: Deep learning variant annotation methods increase pathogenic variant detection in women with breast cancer and nominate candidate noncoding variants. 1094

Board 8015T: DeepSomatic: A SNV and Small-Indel Somatic Caller Using Deep Neural Networks 1095

Board 8016T: Development and application of *MUTYH*-specific ACMG/AMP variant classification criteria 1096

Board 8018T: Development of Cross-ancestry Polygenic Risk Scores for Breast Cancer in Women with African Ancestry ⭐ 1098

Board 8019T: Effect of Inactivation of MAPT Through Gene Knockdown on Tumor Invasion and Microtubule Stability in Prostate Cancer Cells 1099

Board 8020T: Functional characterization of the 1p36.33 pancreatic cancer GWAS locus ⭐ 1099

Board 8021T: Functional genetic variant rs59133000 of gastric cancer regulating *PTGER4* in response to *Helicobacter pylori* infection and affecting NK cell functions ⭐ 1101

Board 8023T: Genetic predisposition to pancreatic adenocarcinoma in Czech patients - *BRCA2, ATM* and *BRCA1* are the most frequent mutant genes 1102

Board 8024T: Genetic testing of healthy family members of Hereditary Breast and Ovarian Cancer patients in routine clinical practice 1103

Board 8025T: Genetically adjusting prostate-specific antigen strengthens associations with risk of upgrading and prostate tumor features in prostate cancer patients on active surveillance 1104

Board 8026T: Genetics of Familial Uveal Melanomas 1105

Board 8027T: Genomic and Clinical Insights from a Large-Scale Study of 172 Retinoblastomas from Saudi Arabia 1106

Board 8028T: Global multi-ancestry meta-analysis uncovers pleiotropic genetic associations with thyroid cancer and other thyroid diseases 1107

Board 8029T: Haplotype-resolved assemblies for somatic mutation benchmark development for a GIAB tumor-normal pair reference sample 1108

Board 8030T: Head and Neck Paraganglioma Study: Whole Genome & Transcriptome Analysis 1109

Board 8031T: HiChIP-based epigenomic footprinting identifies a promoter variant of *UXS*1 that confers genetic susceptibility to gastroesophageal cancer. 1110

Board 8032T: High Depth Allele-Specific Expression Across Healthy Tissues and Ovarian Cancer 1111

Board 8033T: Highly accurate classification of prostate cancer using a random forest classifier built on 3' cfDNA fragment ends 1112

Board 8034T: Identification of Gene Fusions Using RNA-Seq in a Large Cohort of CLL Tumors. ⭐ 1113

Board 8035T: Identification of rare non-coding variants in familial non-medullary thyroid carcinoma 1114

Board 8036T: Identifying patterns in testing and prevalence of gBRCA mutations among women with breast cancer who underwent hereditary cancer germline testing. A cross-sectional observational study 1115

Board 8037T: Identifying variance in pathway activity across grade and stage in Human Hepatocellular Carcinoma ⭐ 1116

Board 8038T: Impact of Exposure to Fire Smoke on miRNA Expression in Firefighters' Blood and Skin Samples 1118

Board 8039T: Insights into the Development of Metastasizing Uveal Melanoma by BAP1 knockdown of a Primordial Eye Organoid Model ⭐ 1119

Board 8040T: Intrinsic subtype distributions across inherited breast cancer genes: an opportunity to refine treatment? 1120

Board 8041T: Leveraging machine learning to resolve biological confounding and expedite lymphoma diagnosis 1121

Board 8042T: Long non-coding RNA expression profiling of breast cancer in Moroccan women ⭐ 1122

Board 8043T: Machine learning-inferred family history of cancers correlated with smoking behavior and tumor characteristics in lung cancer 1123

Board 8044T: MC4R rs79783591 Founder Mexican Variant Related to Aggressive Phenotype in Testicular Germ Cell Tumors 1124

Board 8045T: Modulation of Oncogenic Regulatory Networks by Exosomal miRNAs in Patients with Non-Small Cell Lung Cancers. ⭐ 1125

Board 8046T: Molecular analysis of patients with Richter's syndrome (RS) and progressive chronic lymphocytic leukemia (CLL)..... 1126

Board 8047T: mSigDetect: Homologous recombination deficiency detection using mutational signatures in targeted long read sequenced tumours 1127

Board 8048T: Multi-omic integrative analysis uncovers key transcription factors and target proteins associated with renal cell carcinoma and its subtypes 1128

Board 8049T: Multi-omic profiling of tumor-derived extracellular vesicles enables early cancer detection 1130

Board 8050T: NEK2-driven pathways: A mutli-omic analysis to uncover molecular insights into drug-resistant myeloma ⭐ 1130

Board 8051T: Novel non-coding promoter mutations in *BCL2L12* and *IRF3* impacting 5% of melanoma patients ⭐ 1132

Board 8052T: Optimizing NGS-based MSI Analysis: A Novel Microsatellite Loci Selection Model for Both Tumor-normal Pair and Tumor-only Applications 1133

Board 8053T: Optimizing sequencing depth for reliable ctDNA analysis in plasma samples using down-sampling strategies 1134

Board 8054T: Performance of various cell free DNA extraction kits in healthy plasma samples..... 1135

Board 8055T: Phenotypic characterization of individuals with germline *POT1* variants: A retrospective analysis 1136

Board 8056T: Populations and methods used to define prostate cancer polygenic risk score categories impact the interpretation of risk..... 1137

Board 8058T: Reclassification of VUS shows the value added of using ClinGen´s ENIGMA Expert Panel Specifications for BRCA1 and BRCA2 over the standard ACMG classification system 1138

Board 8059T: Shifting the lens: analysis of genomics cohorts beyond discrete ancestry categories reveals overlooked somatic-germline interactions across cancer types... 1139

Board 8061T: Synergistic effect of microRNA enhances target specificity with reduced cellular toxicity: A potential tool in cancer therapeutics ⭐ 1140

Board 8062T: Synergistic interactions of genetic alterations and immune microenvironmental features predict patient survival and aggressive cancer subtypes 1141

Board 8063T: The impact of cigarette smoke on the transcriptome across human tissues 1142

Board 8064T: Unraveling high-grade gliomas: Enhancing cytogenomic aberration assessment through subtelomeric MLPA analysis 1143

Board 8065T: Uptake of Risk-Reduction, Surveillance and Therapeutic Interventions among Breast Cancer Patients with Pathogenic Germline Variants 1144

Board 8066T: WGS as the new standard for cancer sequencing using the Ultima UG100 1145

Board 8067T: Integrating Germline and Somatic Data to Resolve Variants of Uncertain Significance (VUS) in Colorectal Cancer (CRC) Patientswho Self-Identified as Hispanic/Latino ⭐ 1147

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session 1148

Board 5001T: A CRISPR/Cas9 screen identifies several genes and non-coding regulatory elements involved in the regulation of red blood cell density 1148

Board 5002T: A hot spot loss of function variant in*PLAU*confers risk for thrombotic events in severe COVID19 patients 1149

Board 5003T: A multi-ancestral genetic association study of ischemic heart failure .. 1150

Board 5004T: A multiancestry genome-wide association study of 108,477 individuals with gallstone disease identifies 37 novel genetic loci 1151

Board 5005T: A multi-ancestry, Genome-Wide Association Study (GWAS) of sepsis risk identifies a risk locus in TRIM22 1152

Board 5006T: A Novel Hypervariable Number Tandem Repeat in the Dopamine Transporter Gene is Associated with Multiple Dopamine-Related Phenotypes 1153

Board 5007T: A VNTR containing multiple MEF2 binding sites is associated with the African origin haplotype protective for AD in *APOEε4* carriers 1154

Board 5008T: *ABCA7* deletion and *APOE4* allele lower the age-of-onset of Alzheimer Disease in African-Ancestry individuals 1156

Board 5009T: Absence of SCA27B and unique allelic distribution in Peruvian patients with undiagnosed cerebellar ataxia 1157

Board 5010T: Accelerating novel gene discovery in nonsyndromic cleft lip/palate using whole genome sequencing 1158

Board 5011T: Admixture mapping revels novel population-specific loci associated with kidney disease in African American individuals 1159

Board 5012T: Alcohol use disorder polygenic scores with effect sizes comparable to the family history and *ADH1B* in European ancestry populations ⭐ 1160

Board 5013T: ALS Identified: Enhancing ALS Genetic Diagnosis Through a Sponsored Genetic Testing Program 1161

Board 5014T: Altered bile acid metabolism in Down syndrome ⭐ 1162

Board 5015T: Alzheimer's disease pathway-specific polygenic risk score training and performance evaluation in the UK Biobank cohort 1163

Board 5016T: Analysis of 36,361 whole genome sequenced samples from the Alzheimer's Disease Sequencing Project (ADSP) ⭐ 1164

Board 5017T: Analyzing specific candidate genes implicated in the development of Class III skeletal malocclusion 1165

Board 5018T: *APOBR* is downregulated in EBV+ tonsils of children with obstructive sleep-disordered breathing 1166

Board 5020T: Are germline and somatic mutation rates correlated in fertile and subfertile men? 1167

Board 5021T: Assessing the Performance of a Multi-Ancestry Polygenic Risk Score for LDL with Gene-Environment Interactions of Smoking and PM2.5 Exposure in Racially and Ethnically Diverse Groups: Results from the PAGE Study 1168

Board 5022T: Assessment of polygenic risk score performance in East Asian populations for ten common diseases: A Korean cohort study 1170

Board 5023T: Association analysis of HLA-DQB1 with narcolepsy type 2 and idiopathic hypersomnia 1171

Board 5025T: Associations Between Gut Microbiota and Alzheimer's Disease Biomarkers in a Korean Population 1172

Board 5026T: Associations between HLA polymorphism and COVID-19 severity, susceptibility, and mortality outcomes 1173

Board 5027T: Bi-ancestral analysis of the utility of type 2 diabetes polygenic score for predicting incident diabetes 1174

Board 5028T: Can the polygenic component of idiopathic Parkinson's disease (PD) explain the penetrance of monogenic PD? Findings from polygenic scores. 1175

Board 5029T: Cardiac arrhythmia risk scales with cumulative coding and noncoding genetic liability 1176

Board 5030T: Cell type-specific gene-regulatory effects of structural variants on neuropsychiatric diseases..... 1177

Board 5031T: Characterization of copy number variation in people with Pacific ancestry using PacBio HiFi Sequencing..... 1178

Board 5032T: Characterization of the Phenotypic Consequences of the Duffy-Null Genotype 1179

Board 5033T: Characterizing Early Life Insulin Level-defined Asthma Endotypes with Population-level Single-Cell Multi-omics Profiling 1180

Board 5034T: Claudin genetic variants in kidney stone pathogenesis 1181

Board 5035T: Clonal hematopoiesis is associated with a seven-year delay in the onset of Alzheimer's disease in *APOE* ε4/ε4 cases 1182

Board 5036T: Cognitive Polygenic Index and occupational complexity 1183

Board 5037T: Common and rare variant analyses reveal novel genetic factors underlying Idiopathic Pulmonary Fibrosis and shared aetiology with COVID-19 1184

Board 5038T: Comparison of gestational and type 2 diabetes mellitus transcriptome-wide association studies in South Asian and multi-ancestries populations 1185

Board 5039T: Complex trait analysis of cardiovascular imaging phenotypes via manifold learning 1186

Board 5040T: Comprehensive analysis of Human Leukocyte Antigen (HLA) diversity and association with infectious and immune-mediated diseases in the BioMe biobank .. 1187

Board 5041T: Context Matters: Alzheimer Disease Heritability Differs by Population . 1188

Board 5042T: Context-specific eQTLs reveal causal genes underlying shared genetic architecture of critically ill COVID-19 and idiopathic pulmonary fibrosis 1189

Board 5043T: Convergence of evidence between two methods for cell-type specific association studies robustly identifies differentially expressed genes in schizophrenia 1190

Board 5044T: Copy number variants in familial bipolar disorder ascertained from Anabaptist founder populations 1191

Board 5045T: Cumulative incidence estimates across polygenic score strata in global populations 1192

Board 5046T: Deciphering the Genetic Connection in Juvenile Idiopathic Arthritis and Adult Rheumatic Diseases 1194

Board 5047T: DementiaSeq2, an extended public resource of whole genome sequencing data for study of Lewy Body Dementia and Frontotemporal Dementia / Amyotrophic Lateral Sclerosis genetics 1195

Board 5048T: Design of a pragmatic clinical trial to assess outcomes after incorporating genetic susceptibility into coronary heart disease risk estimates 1196

Board 5049T: Development of a scoring system to guide action based on genetics, biomarkers, and lifestyle data: Health Composition 1198

Board 5050T: Developmentally driven gene program in fat depot-specific adipogenesis. 1199

Board 5051T: Differential Gene Expression Related to Opioid Use Behaviors in a Hybrid Rat Diversity Panel 1200

Board 5052T: Differentially expressed transcripts in HIV-TB co-infected children from East and Southern Africa are enriched for genes impacting neutrophil function ⭐ .. 1201

Board 5053T: Discovery of Genetic Factors for Atypical Antipsychotic-Induced BMI Change: Obesity Related to Antipsychotic Liability & Exposure (ORAcLE) Consortium 1202

Board 5054T: Dissecting the genetic basis of clinical phenotype heterogeneity within IBD in more than 55,000 IBD patients ⭐ 1203

Board 5055T: Diverse genomic alterations in single neurons after chronic brain trauma 1204

Board 5056T: Downregulated transcriptome in dopaminergic neurons from the ventral midbrain is enriched for schizophrenia risk loci. 1205

Board 5057T: Dysregulated plasma miRNAs are associated with cerebrospinal fluid and neuroimaging biomarkers in Alzheimer's disease continuum 1206

Board 5058T: Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes 1207

Board 5059T: Elevated risk of schizophrenia in 22q11.2 Deletion Syndrome with genome-wide tandem repeat expansions 1208

Board 5060T: Enhancing cardiovascular disease risk prediction using polygenic risk scores and traditional risk algorithms in the UK Biobank 1209

Board 5061T: Enhancing Early Alzheimer's Disease Detection Using Machine Learning and SHAP-Enhanced Non-Invasive Biomarkers 1210

Board 5062T: Epistasis and Gene-Gene Interaction Network to Identify Core Genes of Complex Traits 1211

Board 5063T: ESCALATOR: An efficient bioinformatics pipeline to harmonize and apply external weights for PRS calculation in large-scale datasets ⭐ 1212

Board 5064T: Evaluating Approaches for Multi-Ancestry Polygenic Risk Scores Development in Alzheimer's Among Hispanic Populations 1212

Board 5065T: Evaluating reported ethnicity, genetic ancestry and three-way local-ancestry admixture with preeclampsia in GenPE, a multi-center case-control study of Colombian women. 1214

Board 5066T: Evidence for Y chromosome dosage effects on autism risk 1215

Board 5067T: Examining the link between 179 lipid species and 8 disease endpoints using genetic predictors ⭐ 1216

Board 5068T: Exploration of bulk TCR repertoire in Graves' disease patients to evaluate T cell receptor diversities and signature sequences associated with HLA alleles and disease phases 1217

Board 5069T: Explore genetic modifiers of disease progression in amyotrophic lateral sclerosis 1218

Board 5070T: Exploring Genetic Determinants of Type 2 Diabetes: The Causal Impact of *SREBF1* Revealed by Two-Sample Mendelian Randomization 1219

Board 5071T: Exploring the impact of rare variant burden on COVID-19 severity: A whole genome sequencing analysis 1220

Board 5072T: Exploring the potential causal association of gut microbiota on panic & conduct disorder using a two-sample Mendelian randomization approach 1222

Board 5073T: Fifty-one novel, replicated loci identified in genome-wide association study of polyunsaturated and monounsaturated fatty acids in 124,024 European individuals 1222

Board 5074T: Functional screening of AD genes via CRISPR activation in microglia... 1223

Board 5075T: Gene-set based analysis of whole genome sequencing data of families with schizophrenia and bipolar disorder from Portuguese Islands Collection 1225

Board 5076T: Genetic Analysis in Over 51,000 Mother-Child Pairs Suggests Higher Maternal Testosterone Protects Daughters from Anxiety 1226

Board 5077T: Genetic analysis of X-chromosome in myasthenia gravis 1227

Board 5078T: Genetic architecture and comprehensive phenotypic assessment of alcohol use disorder and related outcomes in All of Us. 1228

Board 5079T: Genetic architecture of the limbic white matter in aging 1229

Board 5080T: Genetic correlates of treatment-resistant depression: Insights from polygenic scores across cognitive, temperamental, and sleep traits in the all of us cohort 1231

Board 5081T: Genetic correlations inform machine learning based identification of heavy alcohol consumption using objective biomarkers 1232

Board 5082T: Genetic covariance analysis identifies *TMEM106B* and *ACE* as genetic loci shared between Alzheimer's disease and primary psychiatric disorders 1233

Board 5083T: Genetic differences between male & female pattern hair loss in a Korean population 1234

Board 5084T: Genetic mechanisms define novel drug targets for IgA vasculitis ⭐ ... 1235

Board 5085T: Genetic risk estimation of pregnancy complications through polygenic risk scores from low-coverage NIPS data 1236

Board 5086T: Genetic Risk, Adherence to Brisk Walking Pace, and Venous Thromboembolism 1237

Board 5087T: Genetic variants affect diurnal glucose levels across the day 1238

Board 5088T: Genetic variants associated with TST and LPS in insomnia patients: whole genome sequencing study 1239

Board 5089T: Genetics of postpartum depression in African American women 1240

Board 5090T: Genome sequencing identifies novel risk loci for multiple system atrophy 1241

Board 5091T: Genome-based Analysis of Patients with IEI Increases Diagnostic Yield by 53% 1243

Board 5092T: Genome-wide association studies and integrative analysis with multi-omics data reveal novel loci and their molecular mechanisms for polyunsaturated, monounsaturated, and saturated fatty acids 1244

Board 5093T: Genome-wide Association Study of Long COVID by Predictive Phenotyping of EHR Data using Deep Learning: Transferring Trained Model from National COVID Cohort Collaborative to All of Us Research Platform Enables Genetic Discovery 1245

Board 5094T: Genome-wide meta-analysis of Alzheimer's disease reveals genes causally involved in disease risk by linking to gene expression loci in whole blood and brain. . 1246

Board 5095T: Genomic analysis identifies risk factors in restless legs syndrome 1247

Board 5096T: Genomic Insights into Autism Spectrum Disorder: Comprehensive Analysis of Genetic Variants in Indian Trios 1248

Board 5098T: Genotype specific genetic architecture of alcohol consumption identified by a genotype-stratified GWAS in Japanese people 1249

Board 5099T: Gestational diabetes mellitus shares genetic risk factors with type 1 and type 2 diabetes and is predicted by diabetes polygenic scores 1250

Board 5100T: GWAS in Pacific Islanders identified established and novel risk loci for four serum lipid measures 1251

Board 5101T: GWAS uncovers the neuronal contribution to multiple sclerosis susceptibility ⭐ 1252

Board 5102T: GWAS-informed data integration and non-coding CRISPRi screen illuminate genetic etiology of bone mineral density 1253

Board 5103T: Heterogeneity in the Genetic Architecture of AD across Ancestry Groups 1254

Board 5104T: Higher gene expression of dynein heavy chains in the dorsolateral prefrontal cortex predict lower neuropathology and better cognitive outcomes in individuals with Alzheimer's disease 1255

Board 5105T: Identification of a rare missense variant in *VSX2* as a novel risk locus for retinal detachment ⭐ 1256

Board 5106T: Identification of genes influencing rhinovirus and respiratory syncytial virus IgG responses in children and adults 1257

Board 5108T: Identifying common, rare, and structural variants associated with cardiac structure and function by whole genome sequencing..... 1258

Board 5109T: Identifying of 10 Genes of Interest in 17 Families with dyslexia: The Influence of Simultaneously Present Variants on Individual Phonemic Decoding Ability 1260

Board 5110T: Identifying rare and less common SNPs associated with type 2 diabetes mellitus in multiplex families from the multiethnic GENNID study 1261

Board 5111T: Identifying splicing quantitative trait loci (sQTL) in colon tissue of IBD patients 1262

Board 5112T: Identifying ultra-rare genetic variants in Alzheimer's disease 1263

Board 5114T: Impulsivity facets and substance use involvement: Insights from genomic structural equation modeling 1264

Board 5115T: Integrating massive-scale GWAS and multi-omic datasets reveals disease relevant cell types and effector genes for inflammatory bowel disease ⭐ 1265

Board 5116T: Integrative analysis of tandem repeat expansions in pediatric heart diseases 1266

Board 5117T: Inter-eye clinical asymmetry and molecular discrepancies in the corneal epithelium in Keratoconus 1267

Board 5119T: Investigation of rare recessive effects on quantitative and binary phenotypes in a cohort of 39,152 British Pakistanis and Bangladeshis enriched for autozygosity 1268

Board 5120T: Large scale analysis of patient-derived induced neurons to identify early developmental changes in schizophrenia ⭐ 1270

Board 5121T: Large scale exome sequencing in South-Asian populations identifies an enriched missense variant in *HNF4A* protective from type 2 diabetes 1271

Board 5122T: Latent classes analyses as a tool for revealing variable genetic association patterns in COVID-19 clinical subgroups 1272

Board 5123T: Leveraging EHR Phenotyping and Genome-Wide Association Studies to Identify Genetic Predisposition Factors for Systemic Lupus Erythematosus Within *All of Us* Research Program 1273

Board 5124T: Longitudinal blood pressure is associated with risk of all-cause dementia 1274

Board 5125T: Lower fasting insulin and higher vigorous physical activity independently associate with the same direction of cell-type-specific gene expression in human skeletal muscle fibers ⭐ 1275

Board 5126T: Mapping RNA isoform diversity and dysregulation in neuropsychiatric and control human brains 1277

Board 5127T: Metabolic Syndrome-Associated Variants and Orofacial Cleft Risk in Puerto Rican Hispanics: Impact of Maternal BMI 1278

Board 5128T: Metabolomic Profile of Excessive Daytime Sleepiness in the Hispanic Community Health Study/Study of Latinos 1279

Board 5129T: Meta-Prediction of Coronary Artery Disease Risk 1280

Board 5130T: Molecular analysis of a trans interaction between *Ret* and *Ednrb* in Hirschsprung disease 1281

Board 5131T: Multi-ancestry *APOE*-stratified genome-wide association study and identification of genetically at high-risk individuals for AD among ε4 non-carriers 1282

Board 5132T: Multi-ancestry stratified GWAS of alcohol use disorder and comorbid depression identifies novel loci..... 1283

Board 5133T: Multi-disciplinary Genomic Studies Inform the Etiology of Congenital Heart Disease 1284

Board 5134T: Multiethnic genome-wide association meta-analyses of primary open-angle glaucoma identifies novel cross-ancestry and ancestry-specific loci..... 1285

Board 5135T: Multi-omic characterization of persons living with T2D in a vulnerable Hispanic/Latinx population 1287

Board 5136T: Multi-omics profiling of angiographically confirmed coronary artery disease highlights genetic regulation of the human plasma proteome ⭐ 1288

Board 5137T: Multi-Trait Genome-Wide Association Analysis of Psychiatric Traits Identified New Loci 1289

Board 5139T: Next generation sequencing study to unravel the genetic basis of congenital bile-duct dilatation 1290

Board 5140T: Novel contributors to orofacial cleft risk identified at 7q33 and 10q24.32 through multi-ancestry meta-analysis..... 1291

Board 5141T: Novel loci mapped for blood pressure trait trajectories in diverse adolescence and early adulthood. 1292

Board 5142T: Overlap between COPD genetic association results and transcriptional quantitative trait loci 1294

Board 5143T: Pathogenic variation in *SNCA* causes allelic biases in transcript use that are identified by long read RNA sequencing and have the potential to drive Parkinson's disease pathology. 1295

Board 5144T: Pathway of Genetic Load Analysis Within a High-Risk Population in Nigeria to Analyze Genetic Risk for Developing Kernicterus 1296

Board 5145T: Polygenic predictors of incident cognitive impairment and dementia-associated mortality in the REasons for Geographic and Racial Differences in Stroke (REGARDS) study 1297

Board 5146T: Predicting and communicating Alzheimer's Disease genetic risk..... 1299

Board 5147T: Predicting coronary artery disease using an exome-based estimate of KIV-2 repeat count 1300

Board 5148T: Prevalence of *RFC1* repeat expansions in an American idiopathic and diabetic neuropathy patient population 1301

Board 5149T: Protection against motor deficits in a Parkinson Disease mouse model, achieved by CRISPR-mediated deletion of a neuronal α-synuclein (*Snca*) intronic enhancer..... 1302

Board 5150T: PsychGenCOV19 : Mapping Molecular Signatures of COVID-19 Impact on Mental Health 1303

Board 5151T: Rare protein-coding variants mapping on multi-organ imaging-derived phenotypes ⭐ 1304

Board 5152T: Refining cross-population polygenic risk scores to optimize trait prediction in diverse populations 1305

Board 5153T: Revealing the developmental-genetic basis for the human cranial vault 1306

Board 5154T: Risk factors for coronary artery disease in individuals with high plasma Lp(a): A Mendelian randomization study 1307

Board 5155T: Severe COVID-19 enriched in participants with G6PD deficiency in the All of Us Research Program dataset..... 1308

Board 5156T: Sex-dependent genetic architecture of tobacco-related biomarkers ... 1309

Board 5157T: Sex-Specific genetic insights into Alzheimer's Disease: integrative analysis of genomic, transcriptomic, and phenotypic Data 1310

Board 5158T: Sex-specific genetic underpinnings of Alzheimer's disease and body mass index 1311

Board 5159T: Sex-Stratified Genome-Wide Association Study of Sulfonylurea Drug Response in Individuals with Type 2 Diabetes 1312

Board 5160T: Shared genetic architecture of brain shape and IBD is specifically enriched for inflammatory response genes 1313

Board 5161T: Single-cell profiling of proliferating T-cells reveals division-state-specific gene expression dynamics and links immune-disease variants to cell division states ⭐ 1314

Board 5162T: Single-cell transcriptomics of nasal swabs identifies distinct viral signatures of pediatric respiratory syncytial virus infection 1315

Board 5163T: Spatial transcriptome profiling of the pancreas uncovers novel targets and pathways in *Clec16a* knockout mice. 1316

Board 5164T: Targeted Resequencing of Otosclerosis Patients from Different Populations Replicates Results from a Previous Genome-Wide Association Study 1317

Board 5165T: The Brain on Bilirubin - Genetic Analysis of Hyperbilirubinemia and Kernicterus Susceptibility in Nigerian Newborns 1318

Board 5166T: The clinical benefits of genetically guided population health management ⭐ 1319

Board 5167T: The impact of donor and recipient polygenic burden on kidney transplant outcome ⭐ 1320

Board 5168T: The interactions of mosaic loss events with genetic risk are strongly associated with schizophrenia and bipolar disorders. ⭐ 1321

Board 5169T: The landscape of structural variation in bipolar disorder and schizophrenia across the frequency spectrum of diverse populations ⭐ 1323

Board 5170T: Trans-ancestry Meta-analysis of HIV Acquisition Reveals Novel African American-specific Risk Loci 1324

Board 5171T: Transcriptome prediction and BMI polygenic risk score performance varies across African genetic ancestries populations from four diverse environments 1325

Board 5172T: Transcriptome-wide Association Study Identifies Novel Risk LociAssociated with Age-Related Macular Degeneration 1326

Board 5173T: Transcriptomic Effects of Mood Stabilizing Drugs on the Gene Regulatory Network in Cell Villages of Human iPSC-derived Cortical Neurons 1328

Board 5174T: Transcriptomic profiling analysis during tolerance induction in dendritic cells from women with Systemic Lupus Erythematosus 1329

Board 5175T: Uncovering novel genetic associations for lung, liver, and renal-system fibrosis through multi-trait analysis. 1330

Board 5176T: Understanding the impact of genetic variants on Alzheimer's disease with mass spectrometry proteogenomics 1331

Board 5177T: Unraveling the human brain cell-cell interactome: Shared mechanisms across prevalent neuropsychiatric and neurodegenerative diseases 1332

Board 5178T: Unravelling the genetic architecture of inflammatory bowel disease multiplex families with rare and common variant polygenic risk scores 1334

Board 5179T: Unveiling Gene Regulatory Programs Linked to Autism Spectrum Disorder Through Multimodal Sequencing of Human iPSC-Derived Neurons 1335

Board 5180T: Validation of Multi-ancestry Polygenic Scores for Lipid Levels in 3,119 Participants from Samoa and American Samoa 1336

Board 5181T: Variability in predictive and stratification performance of epilepsy polygenic risk scores in Taiwanese population biobank and clinical cohorts 1337

Board 5182T: Variants in *CHCHD2* on European and Amerindian ancestral background in Hispanic PD patients 1338

Board 5183T: Whole exome sequencing analysis to identify rare variation in polycystic ovary syndrome (PCOS) 1339

Board 5184T: Whole genome sequencing analysis reveals rare variant contribution to coronary artery disease in the Japanese population 1340

Board 5185T: Whole genome sequencing association testing in 520,000 individuals identifies rare variants and non-coding regulatory elements associated with HbA1c and Type 2 diabetes 1341

Board 5186T: Whole-genome identification of short tandem repeats underlying idiopathic peripheral neuropathy using short-read and long-read sequencing 1343

Session Title: Epigenetics Poster Thursday Session 1344

Board 1001T: A flavonoid associated with slower epigenetic aging 1344

Board 1002T: A Modified Atkins Diet is Feasible, Tolerable, and Shows Early Signs of Efficacy in Kabuki Syndrome Type 1 1345

Board 1003T: A probabilistic model for RNA-binding QTL discovery 1346

Board 1004T: Advantages in modeling neuropsychiatric disorders in iPSC derived cell types to functionally evaluate genetic risk factors ⭐ 1347

Board 1005T: Association of BMI with Whole Blood DNA Methylation: An Epigenome-Wide Study of Bangladeshi Adults 1348

Board 1006T: Associations between depressive symptoms and DNA methylation among Black women in the InterGEN study 1349

Board 1007T: Autism spectrum disorder specific longitudinal patterns of DNA methylation from birth to age 5 1350

Board 1008T: Clonal Hematopoiesis Paradox Resolved by Profiling Allele-Specific Methylation and Hydroxy-Methylation 1352

Board 1009T: Comparative DNA methylation profiling in PBMCs of cirrhosis with and without hepatocellular carcinoma 1353

Board 1010T: Comprehensive analysis of epigenome and transcriptome-based correlation network analysis in severe asthma risk prediction 1354

Board 1011T: DeepFace: Deep learning-based framework to contextualize orofacial cleft-relatedvariants during human embryonic craniofacial development 1355

Board 1012T: Differential DNA Methylation Enables Accurate Prediction for *C9orf72* Pathogenic Repeat Expansion 1356

Board 1013T: Distance-dependent synergy among neighboring genomic regulatory elements 1357

Board 1014T: DNA methylation and type 2 diabetes in West Africans: A Mendelian randomization analysis 1358

Board 1015T: Elucidating DNMT3A instability mutation in clonal hematopoiesis 1359

Board 1016T: Epigenetic Age in Veterans with Gulf War Illness in the Gulf War Era Cohort and Biorepository (GWECB) 1360

Board 1017T: Epigenetic and genetic diagnostic innovations for differentially methylated regions in cell-free DNA from cancer patients and high-throughput urine samples, respectively 1361

Board 1018T: Epigenetic Evidence that Heavy Alcohol Consumption is a Common, Large Risk Factor for Mortality in Patients with Acute Coronary Syndrome 1362

Board 1019T: Epigenetic scores of plasma proteins as biomarkers of Amyotrophic lateral sclerosis 1363

Board 1020T: Epigenome-wide association study in multi-ethnic Asian population identifies novel markers for skin physiology measures and atopic dermatitis 1364

Board 1021T: Evaluating Methylation Analysis on Microbial Community Standards to Advance the Study of the Human Microbiome 1365

Board 1022T: Extensive methylation array variant-under-the probe (VUP) binding disruption: evidence for probe binding disruption for VUPs up to 50 bp from the CpG 1366

Board 1023T: Genome Organization with CUT and Tag (GO-CaT) identifies substructure of lamina-associated domains in developing human brain neurons 1367

Board 1024T: Genome-wide dysregulation of R-loops in Ataxia Telangiectasia neurological pathogenesis 1368

Board 1025T: Genomic architecture of lowly expressed bidirectionally transcribed enhancers and their links to complex traits 1369

Board 1026T: Identification of a shared transcriptomic signature in sex chromosome aneuploidies: A study on 47,XYY and 47,XXY iPSCs ⭐ 1370

Board 1027T: Integrated analysis of genetic variation, gene expression and methylation changes in temporal lobe epilepsy 1371

Board 1028T: Integrative analysis of RNA-Seq, long-read DNA sequencing, and clinicopathological profiles of LINE-1 retroelement in non-small cell lung cancer..... 1372

Board 1029T: Investigating the Role of Small Cajal Body-Associated RNAs (scaRNAs) in Congenital Heart Defects: A Study of SNORD94 Regulation of Alternative Splicing ... 1373

Board 1030T: Machine learning reveals the diversity of human 3D chromatin contact patterns 1374

Board 1031T: Metabolomic analysis indicates adenine, oleic acid, and glutamic acid are associated with flare remission, DNA methylation changes, and clinical subtypes in systemic lupus erythematosus 1374

Board 1032T: MeVars (DNA Methylation Variants): Identification, tissue-specificity, reprogramming effects and disease susceptibility. 1376

Board 1033T: Molecular cross-talk in vascular calcification 1377

Board 1034T: Multi-Ancestry Epigenetic Investigation of Generalized Anxiety Disorders in the Million Veteran Program ⭐ 1377

Board 1035T: Multilocus Imprinting Disturbances (MLID) analysis in a cohort of Beckwith-Wiedemann and Silver-Russell Syndromes: Comparison between specific MS-MLPA and methylation microarray 1378

Board 1036T: Nuclear regulatory disturbances precede and predict the development of Type-2 diabetes in Asian populations 1379

Board 1038T: PTSD and Epigenetic Aging: A Longitudinal Meta-analysis..... 1381

Board 1039T: The nuclear epigenome mediates the effect of mitochondrial DNA variation on cardiovascular disease risk 1382

Board 1040T: To stop or not to stop? An epigenomic landscape of transcription termination alteration networks implies their functional role in solid cancers 1383

Session Title: Evolutionary and Population Genetics Poster Thursday Session 1384

Board 3002T: A Diversity-First Approach to Understanding Primary Open-Angle Glaucoma 1384

Board 3003T: A pangenome approach to study short read data for the population genetics of HLA ⭐ 1385

Board 3004T: Australian Indigenous Population Structure and Its Impact on Unique Genetic Variation 1387

Board 3005T: Characterizing evolutionary constraint of different time scale on genome function and diseases 1387

Board 3006T: Confirmatory testing for Hemoglobin Disorders Included in the Philippine Newborn Screening Program 1389

Board 3007T: Efficient adaptive target detection and characterization with minimal training by transfer learning 1390

Board 3008T: Evaluating Functional and Evolutionary Genomic Predictors of Polygenic Score Portability Using Machine Learning and Explainable AI 1391

Board 3009T: Evolutionary dynamics of intrinsically disordered regions of human transcription factors 1392

Board 3010T: Founder variants in the Mexican population: a Systematic Review 1392

Board 3011T: Functional and evolutionary relationships of chitinase-like protein YKL-39 with chitotriosidase 1394

Board 3013T: Functional impact of 2.7M structural variants across global populations 1394

Board 3014T: Genetic ancestry reveals European population contribution in Brazilian communities of African descent in Central Brazil 1396

Board 3015T: Genetic similarity groups detected by Identity-By-Descent clustering reveals disparities in disease prevalence in a Quebec population-based biobank ⭐ 1397

Board 3016T: GenLib.jl: Unveiling fine population structure through large-scale kinship coefficients' visualization 1398

Board 3017T: GenoBERT: A full-attention small language model for enhanced genotype imputation 1399

Board 3018T: GrafAnc: Enhanced Genetic Ancestry Inference at Subcontinental Levels 1400

Board 3019T: Gut microbiomes from diverse human populations induce unique transcriptional host response 1401

Board 3021T: Haplotype-sharing among 860K individuals in Britain and Denmark reveals migration pattern across the North Sea ⭐ 1402

Board 3022T: Hidden Population Structure in the All of Us Research Project ⭐ 1403

Board 3023T: Increased representation of human genetic ancestries improves genetic intolerance metrics..... 1404

Board 3024T: Initial analysis of 160,000 whole genomes from the Emirati Genome Program 1405

Board 3025T: Integration of ribosome profiling experiments identifies a large pool of novel translated open reading frames neutrally evolving in the human genome 1406

Board 3026T: Local ancestry analysis identifies signatures of selection in the admixed Luo population from Kenya..... 1407

Board 3027T: Long-read *de novo* assembly and comparative analysis of six howler monkey genomes within genus *Alouatta* 1408

Board 3028T: Massive expansion of a Denisovan-introgressed tandem repeat in MUC19 in Americans ⭐ 1409

Board 3029T: Peruvian Population Genomics: Diversity, Admixture Dynamics, and the Impact of Indigenous American Ancestries on Complex Phenotypes ⭐ 1410

Board 3030T: *PGG*.PGP: an academic web server for personal genome analysis 1411

Board 3032T: Robust detection of ancient trans-species balancing selection of varied footprint sizes 1412

Board 3033T: Structural variation hotspots in the human genome and their evolutionary implications 1413

Board 3034T: Surprising source of confounding impacts the sign of GWAS effect estimates on immune function 1414

Board 3035T: The Genetic History of the Qiang People..... 1415

Board 3036T: The Impact of Associative Overdominance on Genetic Diversity 1416

Board 3037T: Understanding founder variants in Kuwait to expedite diagnosis and reduce burden of genetic diseases. ... 1416

Board 3038T: What can admixture tell us about the genetics of health disparities? ... 1418

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session ... 1419

Board 7052T: A Comprehensive System for Monitoring Pediatric Genetic Testing Outcomes Using EHR Data: Insights from the eMERGE IV Study ... 1419

Board 7053T: A Framework for Integrating Polygenic Risk Scores in the Electronic Health Record: The eMERGE Phase IV Study ... 1420

Board 7054T: A scoping review investigating the impact of genetic counselor involvement in genetic and genomic test order review ... 1421

Board 7055T: A stakeholder informed ethical framework for epigenetic editing ⭐ ... 1422

Board 7056T: Assessing access to clinical genetic services among young children with birth defects in Texas based on parental recall ... 1423

Board 7057T: Barriers to genetic research enrollment in community settings: Lessons from SeqFirst recruitment ... 1424

Board 7058T: Bridging the gap: Patient and healthcare professional perspectives on accessing and using cancer genetics services by racialized and ethnic communities 1425

Board 7059T: Cancer Epigenetics through a Sex and Gendered Lens: Reflections on a Comprehensive Scoping Review ... 1427

Board 7060T: Community Engagement in Practice: A Case Study with the Pukapuka Community in Aotearoa, New Zealand ... 1428

Board 7061T: Comparative impact of rapid whole genome sequencing in pediatric and intensive care units in Quebec pediatric hospitals: the PRAGMatIQ study ⭐ ... 1429

Board 7062T: Delayed diagnosis in neurogenetics disorders in Peru ... 1430

Board 7063T: Democratization and cloud-based genomic data sharing: views of platform developers and early adopters ... 1431

Board 7064T: Disparities in VUS among ethnoracial groups in the National Brain Gene Registry ... 1432

Board 7065T: Evaluating patient-centric care in cancer genetics counselling for HBOC syndrome using a measure of informed choice ... 1434

Board 7066T: Family Information Disclosure and Testing Practices within the Context of a Population Biobank: an Analysis of Survey Results from the Colorado Center for Personalized Medicine Biobank (CCPM Biobank) 1435

Board 7067T: Forced resilience: Indigenous people's perspectives on humanizing diagnostic genomic testing in British Columbia, Canada 1436

Board 7068T: Improving Access to High-quality Healthcare for Incarcerated Patients: The Role of the Genetic Counselor ⭐ 1437

Board 7069T: Outcomes of clinical genetic testing for patients with kidney cancer referred to genetic counseling 1438

Board 7070T: Pathways to precision cancer care: integrating genetic services into a nationwide cancer program at scale ⭐ 1439

Board 7071T: Patient perspectives in precision medicine: Identifying knowledge gaps to inform cancer care 1440

Board 7072T: Protocol for a type 3 hybrid implementation cluster randomized clinical trial to evaluate the effect of patient and clinician nudges to advance the use of genomic medicine across a diverse health system ⭐ 1441

Board 7074T: Psychological effects of a predictive neurodegenerative disease diagnosis 1443

Board 7076T: Receiving Negative Results from DNA-based Population Screening for Adult-onset Diseases: Participants' Experience ⭐ 1444

Board 7077T: Research Participants and Stakeholders views on feedback of Genetics Research Findings: A qualitative study of the H3Africa Kidney Disease Research Network, Ghana. 1445

Board 7078T: Return processes and one-on-one disclosure uptake of a genome informed risk assessment which includes polygenic risk scores in eMERGE IV 1446

Board 7079T: SPOT: Sanger post-orthogonal tracker 1448

Board 7080T: State of Affairs of Genomic Education in US Allopathic Medical Schools 1449

Board 7081T: The Genome Health Analyst: a new role in genomic medicine 1450

Board 7082T: Understanding How Perception of Uncertainty Affects Decision Making for or Against Familial Carrier Testing in Families Affected by Disorders of Purine and Pyrimidine Metabolism 1451

Board 7083T: Utilizing *ABCC11* genotyping in high school genetics education: A hands-on approach 1452

Board 7084T: Utilizing web analytics to assess patient engagement and prioritize educational targets for population-scale biobanks 1453

Board 7085T: What should you do when molecular lab results do not acknowledge likely allele status or current literature regarding resulting disease? 1454

Session Title: Genetic Therapies Poster Thursday Session 1455

Board 8081T: A drug repurposing screen identifies N-acetyl-L-leucine as a candidate therapeutic molecule for SYNGAP1-related developmental disorder 1455

Board 8082T: A knockdown-replacement system to target pathogenic DEAF1 mutations 1456

Board 8083T: An antisense oligonucleotide (ASO) strategy for restoring IP_3R1 expression in spinocerebellar ataxia type 15/16 1457

Board 8084T: Antisense oligonucleotides targeting linked-SNPs provide allele-specific knockdown to a dominant-negative *SPTAN1* pathogenic variant in a complex genetic region 1458

Board 8085T: Bone-targeted anti-TGFβ antibody approach to improve efficacy and safety for the treatment of osteogenesis imperfecta ⭐ 1459

Board 8086T: Development and application of antisense oligonucleotide (ASO) assessment guidelines for loss-of-function variants causing autosomal recessive and X-linked monogenic disorders 1460

Board 8087T: Exon Specific U1 snRNAs as a therapeutic approach to rescue optic neuropathy in Familial Dysautonomia ⭐ 1462

Board 8088T: Expanding the Landscape of Drug Targets for 112 Chronic Diseases Using a Machine Learning-Assisted Genetic Priority Score 1463

Board 8089T: G-quadruplex-based splice switching as a therapeutic approach in Duchenne muscular dystrophy 1464

Board 8090T: High-dose pyridoxine prevents glomerulosclerosis and other disease phenotypes in a patient and knock-in mouse model of $SGPL1^{R222Q}$ variant sphingosine phosphate lyase insufficiency syndrome 1465

Board 8091T: Mitochondrial dysfunction heightens the integrated stress response to drive ALS pathogenesis 1466

Board 8092T: Molecular mechanism and gene augmentation therapy for inherited retinal dystrophy caused by a novel mutation *PROM1* c.619G>T ⭐ 1466

Board 8093T: Prediction of *P.falciparum adenylosuccinate lyase* inhibitors from african natural product databases 1468

Board 8094T: Selective JAK2 specific ASO based potential treatment for MDS and PV 1469

Board 8096T: The RNAs contained in extracellular vesicles from the retinal pigment epithelium show potential as circulating biomarkers of retinal degeneration..... 1469

Board 8097T: Using the zebrafish to model*KRAS*-related complex lymphatic anomaly 1471

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session 1472

Board 1042T: A compendium of somatic mutations across genomes, tissues, and individuals 1472

Board 1043T: A genome-wide association study of multiple epigenetic age accelerations identifies novel genetic variants within *ENO2* in an East Asian population ⭐ 1473

Board 1044T: A global reference map of STR and VNTR composition using vamos 1474

Board 1045T: A population-specific pangenome for Saudi Arabia 1475

Board 1046T: A Public Resource of gnomAD Local Ancestry Inference Data for Diverse Human Genomes 1476

Board 1048T: Accelerating Medicines Partnership® Common Metabolic Diseases (AMP CMD): Building resources to generate prioritized targets for all metabolic diseases .. 1477

Board 1049T: Accurate and robust cell type deconvolution with multiple single cell reference panels in spatial transcriptomics studies 1478

Board 1050T: Accurate genotype calls in complex genetic regions using microarrays 1479

Board 1051T: An Integrated Database of Experimentally Validated Major Histocompatibility Complex Epitopes for Antigen-specific Cancer Therapy..... 1480

Board 1052T: Analysis Strategy for Whole Genome Data in a National Genome Project in Korea 1481

Board 1053T: Bioinformatic analysis of tuberculosis to identify new therapeutic targets through drug repurposing approaches..... 1482

Board 1054T: Carrier Screening in the Middle Eastern Population: A Study of the Qatar Genome Programme ⭐ 1483

Board 1055T: Characterization of AZF Deletions on the Y Chromosome Using Nanopore Long-Read Sequencing: Advances and Diagnostic Applications..... 1484

Board 1056T: Characterizing the natural history of *SLC6A1* disorder: Insights from the Ciitizen database..... 1485

Board 1057T: CHIP Pleiotropy and Shared Burden in Heart Failure and CAD..... 1486

Board 1058T: Clinical Genome Resource Neurological Disorders Clinical Domain Working Group:Delineating gene/variant perplexities in common neurological disorders 1487

Board 1059T: Comparative Analysis of Antihypertensive Medications Stimulating vs. Inhibiting Angiotensin II Type 2 and 4 Receptors in Relation to Type 2 Diabetes Incidence in VUMCSD Cohort 1488

Board 1060T: Comparative analysis of structural variant calling: Short-read alignment-based vs. long-read assembly-based methods 1489

Board 1061T: Construction of a Korean pan-genome based on long-read sequences 1491

Board 1062T: Copy number variant associations with plasma protein levels in the UKB Biobank ⭐ 1492

Board 1063T: CRISPR CuT Visualizer (CCTV) Discovers High Efficiency Designs for Cas9 Multi-Guide Mutagenesis Experiments..... 1493

Board 1065T: deciphEHR: Towards Genomic Medicine at NYU Langone Health..... 1494

Board 1067T: Developing a natural history resource to inform decisions about when to screen for rare monogenic conditions 1495

Board 1068T: Establishing a Community Advisory Board for a Genomic Medicine Research Program in the Somali Community of Southeast Minnesota 1496

Board 1069T: Evaluating Large Language Models on Medical, Lay Language, and Self-Reported Descriptions of Genetic Conditions..... 1497

Board 1070T: *De novo* assemblies and variant calls for 625 samples from the Million Veteran Program cohort using HiFi reads 1498

Board 1071T: Exploring Gut-Brain Axis in Alzheimer's Disease using Computational Host-Microbiome Interaction Analyses..... 1499

Board 1072T: Genome wide association studies of 1,728 traits in a multi-ancestry cohort from the Michigan Genomics Initiative ⭐ 1500

Board 1073T: Hybrid Cloud and Local Data Lifecycle Management at a Medium-Scale Genomics Center 1501

Board 1074T: Identifying Somatic Mutations Utilizing GPU Accelerated DeepSomatic 1502

Board 1075T: Investigation of Genetic Determinants of Cystatin C serum levels in the Korean Cohort 1503

Board 1076T: Landscape of 25,000 Korean whole-genomes in the Korea National Genome Project 1504

Board 1077T: Large scale characterization of disease expressivity in carriers of inherited cardiovascular disease pathogenic variants in the UK Biobank. 1505

Board 1078T: Leveraging Explainable AI for prioritizing genomic variants across monogenic, digenic, and oligogenic frameworks 1506

Board 1079T: Long read analysis predicts RNA binding proteins with isoform-specific functions ⭐ 1507

Board 1080T: Long-read sequencing of hundreds of ancestrally diverse brains provides insight into the impact of structural variation on gene expression and methylation ⭐ 1508

Board 1081T: Loss of Y chromosome associates with lung function decline in men with a history of smoking 1509

Board 1082T: MetaSTAARlite: An all-in-one tool for biobank-scale whole-genome sequencing meta-analysis ⭐ 1510

Board 1083T: Mitochondrial genetics in the Taiwanese population: Insights from the Taiwan Biobank on mtDNA variations and their phenotypic implications..... 1512

Board 1084T: MitoNet1.0: Bridging core and auxiliary mitochondrial processes using genetic interaction profiles 1513

Board 1085T: Native Hawaiian and Pacific Islander reference panel improves genome-wide imputation quality and accuracy in the population 1514

Board 1086T: Quality management strategy for genomic data of a National Genome Project in Korea 1515

Board 1087T: Race and ethnicity biases introduced by filtering genomics-connected electronic health records for patients with complete data 1516

Board 1088T: STimage-1K4M: A histopathology image-gene expression dataset for spatial transcriptomics 1517

Board 1089T: The ClinGen Alport Syndrome Variant Curation Expert Panel: Guiding variant classification in *COL4A3*, *COL4A4*, and *COL4A5* 1518

Board 1090T: The gene expression of rev erb alpha and ror alpha differs in individuals with different chronotypes when it comes to late night eating or disturbed eating schedule 1519

Board 1091T: The NIH Comparative Genomics Resource (CGR): Maximizing the impact of research organisms to human health 1520

Board 1092T: The PhenoDB Ecosystem: Open-Source Web-based Tools for Genomic Discovery 1521

Board 1093T: The Single Cell Aging Library With A Genome Browser: SCALIWAG 1522

Board 1094T: Too many transcripts? Managing transcript complexity with GENCODE Primary 1523

Board 1095T: Tracing the evolutionary path of the CCR5delta32 deletion via ancient and modern genomes 1524

Board 1096T: Underdiagnosis and Health Disparities in Hereditary Hemorrhagic Telangiectasia 1525

Board 1097T: Understanding Willingness to Share SBEDOH Information: Cultural Sensitivity and Community Engagement to enhance Alzheimer's disease research in the understanding of AD risk. 1526

Board 1098T: Variant Impact Predictor database (VIPdb), version 2, with ClinGen PP3/BP4 calibrated threshold recommendations, and CAGI assessments 1527

Board 1099T: Whole genome sequencing analysis of Alzheimer's disease in a west African cohort: preliminary results from the DAWN study 1528

Session Title: Mendelian Phenotypes Poster Thursday Session 1530

Board 2002T: A compound heterozygous variant in *SLC52A2* causes BROWN-VIALETTO-VAN LAERE syndrome in a Malian family 1530

Board 2003T: A *Drosophila* model of mucopolysaccharidosis IIIB 1531

Board 2004T: A Novel Variant in *ARSA* Causes Unusual Phenotype of Infantile Metachromatic Leukodystrophy 1532

Board 2005T: A novel variant in *CLN5* of Japanese girl with Neuronal Ceroid Lipofuscinosis 1533

Board 2006T: A prevalent *MOCS2* variant in the Roma population is associated with a mild form of molybdenum cofactor deficiency 1534

Board 2007T: An effective frameworkpredicting splicing single nucleotide variants in exome sequencing ⭐ 1535

Board 2008T: An Oral-Facial-Digital Syndrome Patient Who Expands the Phenotypic Landscape of *KIAA0753* Related Disorders Suggesting a Continuum of Disease 1536

Board 2009T: Analysis of a novel complex structural abnormality of the *FLNA* gene in a patient with periventricular nodular heterotopia and thrombocytopenia using HiFi read sequencing 1537

Board 2010T: Ankylosing spine-brachydactyly syndrome: A possible novel skeletal dysplasia 1538

Board 2012T: Assessment of suspected candidate gene variants in hypermobility Ehlers-Danlos syndrome patients from the HEDGE study cohort 1539

Board 2013T: Candidate variants in regulatory and non-coding regions of coding genes identified in 4 large UK families from an NGS study - 2x HTAAD [Hereditary Thoracic Aortic Aneurysm and Dissection] and 2x hEDS [hypermobile Ehlers-Danlos syndrome] 1540

Board 2014T: *CASZ1* gene: From a promising GWAS risk locus to a possible new monogenic etiology for primary aldosteronism 1542

Board 2015T: Clinical and Genetic Spectrum in Chinese Hereditary Spastic Paraplegias patients and Identification of a new candidate gene for HSP 1543

Board 2016T: Compound heterozygous*TXNDC15*variants associated with Joubert syndrome phenotype and forebrain malformation 1544

Board 2017T: Comprehensive assessment of recessive, pathogenic *AARS1* alleles in a humanized yeast model reveals loss-of-function and dominant-negative effects 1545

Board 2018T: Comprehensive Genetic Analysis of the Mexican Familial Hypercholesterolemia Registry Participants 1545

Board 2019T: *CSTB* null disease demonstrates the exquisite dosage sensitivity of cystatin B in humans 1546

Board 2020T: *De novo* variants in *TMEM184B* cause a neurodevelopmental syndrome underscored metabolic signaling deficits ⭐ 1548

Board 2021T: Detection of deletions as second variants by whole-genome sequence analysis overlooked by whole-exome sequence analysis in autosomal recessive disorders 1549

Board 2022T: Developing a pipeline to identify rare and novel pathogenic variants for inborn errors of metabolism 1550

Board 2023T: Diagnosing the undiagnosed: Lessons from a large heterogeneous cohort of patients with Cornelia deLange and related diagnoses 1551

Board 2024T: Discovery of cryptic splice variants prompts recalculation of Van der Woude syndrome and popliteal pterygium syndrome mutation rates 1553

Board 2025T: Distinct founder haplotypes across global *CHCHD10* p.Arg15Leu amyotrophic lateral sclerosis patients 1554

Board 2026T: Driver gene ontology analysis highlights key pathways for male infertility 1555

Board 2027T: Early Diagnosis and Optimized Care of Vascular Ehlers-Danlos Syndrome Through AI-Powered Facial Analysis: Results from the Montalcino Aortic Consortium 1556

Board 2028T: Efficacy of whole genome sequencing in the diagnosis of rare neurodevelopmental disorders: experience from a tertiary health care center and the Danish National Genome Center 1557

Board 2029T: Elevated Iron Levels In Tears of Patients Diagnosed With WDR45 X-Linked Optic Atrophy 1559

Board 2030T: Elucidating the biophysical properties of TANGO2 protein; an essential step towards future therapies for TANGO2 deficiency disorder 1560

Board 2031T: Encoding diseases and phenotypes using large language models reveals linkages between Mendelian disorders and complex diseases 1561

Board 2032T: Enhanced detection and genotyping of disease-associated tandem repeats using HMMSTR and targeted long-read sequencing 1562

Board 2033T: Expansion of the NDST1 syndrome and evidence for abnormal heparan sulfate 1563

Board 2034T: Frequency of *FGF14* Intronic GAA Repeat Expansion in Patients with Multiple System Atrophy and Undiagnosed Ataxia in the Japanese Population 1564

Board 2035T: Gaucher disease type 1 patient treated succesfully with Taliglucerase alfa and Eliglustat: a case report 1565

Board 2036T: Generating and sharing valuable rare disease data via the GREGoR Consortium ⭐ 1566

Board 2037T: Genetic epidemiology and clinical characteristics of SCA424 patients with mutations in *CACNA1G*based on J-CAT study 1567

Board 2038T: Genetic Factors in Incontinentia pigmenti patients with subclinical autoimmunity..... 1568

Board 2039T: Genetic testing diagnostic yield by age and test modality across 856 patients in a cohort of adult and pediatric patients with epilepsy at a tertiary care center 1569

Board 2040T: Genome Sequencing in Early-onset Psychosis..... 1570

Board 2041T: Genomic analyses of 317 consanguineous kindreds from the Middle East and North Africa facilitate the discovery of novel autosomal recessive neurodevelopmental rare disease traits 1572

Board 2043T: GSK3β inhibition is a candidate pharmacological mechanism for treating NGLY1 deficiency 1573

Board 2045T: HNRNP-Related Neurodevelopmental Disorders: Expanding knowledge of known conditions and prompting novel studies into candidate genes. 1574

Board 2046T: Identification of *ERCC2* as a Novel Waardenburg Syndrome Gene Through Analysis Human Neural Crest Developmental Regulatory Networks ⭐ 1575

Board 2047T: Integrative analysis of genotype and phenotype data from the UK Biobank identified over 50-fold higher-than-expected prevalence of hereditary pancreatitis .. 1576

Board 2048T: Investigating the molecular function of *THAP12*, a novel gene associated with a rare epileptic encephalopathy..... 1577

Board 2049T: Investigating the role of the GABA transporter Solute carrier 6 member 1 (*SLC6A1*) in Mendelian disease traits 1578

Board 2050T: Is Classic Galactosemia a Progressive Disease? The First Longitudinal Natural History Data in Children with Classic Galactosemia. 1579

Board 2051T: Jacobsen and Beckwith-Wiedemann Syndromes in a Child with Neonatal Encephalopathy and De Novo 11pter Duplication and 11qter Deletion 1580

Board 2052T: Laboratory experience with primary immunodeficiency panel testing.. 1581

Board 2053T: Large-scale whole-genome sequencing analysis reveals a higher than expected prevalence of C9orf72 repeat expansion alleles associated with specific haplotype structure in the general population. 1582

Board 2054T: *Lat*^{Y136F} mutant mice as a preclinical model of IgG4-RD enable to assess the efficacy and safety of therapies *in vivo*. 1583

Board 2055T: Leveraging population datasets to identify highly penetrant variants in monogenic obesity genes ⭐ 1584

Board 2056T: Lung sensitizing monogenic variant carrier frequency in idiopathic bronchiectasis 1585

Board 2057T: Lysinuric protein intolerance: Clinical manifestations in a compound heterozygote with a novelpathogenic variant..... 1587

Board 2058T: Mendelian variation in a Mexican early onset psychosis cohort analyzed using an AI interpretation algorithm 1588

Board 2059T: Mitochondrial variant calling from genome sequencing identifies new diagnoses in a rare disease cohort 1589

Board 2060T: Molecular characterization of the pathogenic exon 13-15 duplication in *PHEX*: investigating the intervals in families with and without the *231A-G variant 1590

Board 2061T: *MSH3* is a modifier of repeat expansion and age at onset in X-linked dystonia parkinsonism (XDP)..... 1591

Board 2062T: Multi-omic Sequencing Approaches Identify Molecular Causes of Genetically Unsolved Developmental and Epileptic Encephalopathies..... 1592

Board 2063T: *NKX6-2*- related leukodystrophy: natural history and novel genetic therapeutic approaches..... 1593

Board 2065T: Novel candidate gene associated in Dilated Cardiomyopathy (DCM) in a Malian family..... 1594

Board 2066T: Partial methylation of a pathogenic *XYLT1* repeat expansion associated with intrafamilial variation in severity of Desbuquois dysplasia 2..... 1595

Board 2067T: Pathogenic variants in ADTKD-*MUC1* can be detected using short-read sequencing data..... 1596

Board 2068T: Pathological Short Tandem Repeats Analysis and Imprinting Disorders by Long-Read Sequencing in Affected Individuals 1597

Board 2069T: PD GENEration: a genetic analysis beyond the classical 7 Parkinson's disease genes 1598

Board 2070T: Phenotypic expansion and genotype-phenotype correlations in Kleefstra syndrome using data from a large clinical cohort 1600

Board 2071T: *POLD3* haploinsufficiency is linked to non syndromic sensorineural late onset progressive hearing and balance impairment. 1601

Board 2072T: Rare homozygous variants in *INSR* and *NFXL1* are associated with severe treatment-resistant Psychosis ⭐ 1601

Board 2073T: Rare variants modulating phenotypic outcome in Neurofibromatosis type 1 1603

Board 2074T: The Australian Genomic Health Alliance (AGHA) Mitochondrial Flagship: a national approach to delivering mitochondrial disease diagnoses. ⭐ 1604

Board 2075T: The utility of exome sequencing in the seemingly obvious neurological diagnoses 1605

Board 2076T: TMC-CZI: Integrative Genomic and Functional Modeling Approaches for Novel Gene Discovery in Children 1606

Board 2077T: Understanding rare variant contribution to cleft palate using multiplex families 1607

Board 2078T: Universal Exome Sequencing in Critically Ill Adults:A Diagnostic Yield of 25% and Race-Based Disparities in Access to Genetic Testing ⭐ 1608

Board 2079T: Variants in *FGF20* underlie a novel breast malformation 1609

Board 2080T: Whole exome sequencing trio identifies a novel copy number variation linked to Canavan disease 1611

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session 1612

Board 3041T: A 13q14.11q21.33 deletion involving *RB1* and *ATP7B* implicating hereditary retinoblastoma and Wilson disease: A 26-year follow-up 1612

Board 3042T: A retrospective study of Sanger sequencing reduction in clinical genetic testing based on whole genome sequencing 1613

Board 3043T: A *ZFHX3* nonsense variant in a girl with developmental delay and arrhythmia 1614

Board 3044T: Autosomal recessive variant reporting in clinical CMA, a three-year retrospective analysis 1615

Board 3045T: Building an interpretable Bayesian quantitative variant classification system 1616

Board 3046T: Characterization of a complex genomic rearrangement by optical genome mapping reveals BIRC6 as a candidate gene for intellectual disability 1617

Board 3047T: Complex chromosomal rearrangements involving 66 breakpoint junctions on 7 chromosomes in a patient with atypical symptoms 1618

Board 3048T: Complex rearrangement including 7q21.3 region detected in the patient with split hand/foot malformation. 1619

Board 3049T: Copy number variation detection without using external baseline 1620

Board 3050T: Detection of a balanced translocation between chromosomes 11p and 14q by fluorescent in situ hybridization- A Case Report 1621

Board 3051T: Evaluating consensus policy in rare disease diagnosis using automated HPO match 1622

Board 3052T: Expanding the diagnostic toolkit using long-read RNA-sequencing 1623

Board 3053T: Experience of germline genetic testing for inborn errors of immunity: using multigene panel testing compared to exome sequencing at a diagnostic laboratory . 1624

Board 3054T: Genetic causes of autism spectrum disorder in twins: Molecular delineation and variability in autism parameters ⭐ 1626

Board 3055T: Identification of diagnostic and potential causative variants from critically ill neonates through rapid WGS 1626

Board 3056T: Improvement of Copy Number Variant Analysis from Exome Sequencing Using a Backbone Probe Enhancement 1627

Board 3057T: Long-molecule genomics analyses elucidating the hidden structural variations associated with *MECP2* duplication syndrome 1628

Board 3058T: Long-read sequencing of a neurodevelopmental disorder patient reveals complex rearrangement involving the *ARID1B* gene 1629

Board 3059T: Molecular genetic testing for rare diseases in Peru ´s National Pediatric Institute: A 9-year journey 1630

Board 3060T: Molecular genetic testing of congenital adrenal hyperplasia with massive parallel long-read sequencing 1631

Board 3062T: Racial disparities in access to a precise genetic diagnosis are not due to differences in diagnostic yields 1632

Board 3063T: Rapid Genome Sequencing Augments Time to Diagnosis in Prader-Willi Syndrome 1633

Board 3064T: Reducing uncertainty in variant classification by leveraging the capabilities of evidence modeling 1634

Board 3065T: Revealing Hidden Splicing Variants in Inherited Retinal Degeneration .. 1635

Board 3067T: Sensitivity of next generation sequencing platforms in the detection of polyalanine repeat mutations in *PHOX2B* 1636

Board 3068T: SimPheny: Matching similar phenotypes across undiagnosed disease patients for genetic diagnosis 1637

Board 3069T: The development of a diagnostic test method for Prader-Willi syndrome and Angelman syndrome based on targeted long-read sequencing 1638

Board 3071T: The twist of *TP53*: Rare case report of 17p13.1 multigene deletion and Li-Fraumeni syndrome 1640

Board 3072T: Two intronic variants in the SDCCAG8 gene end a patient journey 1641

Board 3073T: Unearthing clinical genetic test data concealed in the electronic health record to quantify progress in genetic medicine 1642

Board 3074T: Utilizing whole exome sequencing to uncover genetic diagnoses in prolonged neonatal/infantile intrahepatic cholestasis 1643

Board 3075T: Variant classification is impacted by the presence of biochemical data in published patient reports: A case study for Familial Hypercholesterolemia 1644

Board 3076T: Whole exome sequencing resulted in a definitive diagnosis for individuals exhibiting common global developmental delay and intellectual disability at early stages 1645

Board 3077T: Whole-genome, single-cell DNA sequencing to identify de novo copy number mutations in individual sperm cells 1646

Board 3078T: Williams Beuren Syndrome, 11 year clinical follow up, cytogenetic study, and a neuropsychological approach: case report 1647

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session 1648

Board 6002T: A concordance study between molecular and HPLC-based methods in diagnosing sickle cell disease 1648

Board 6003T: A framework to characterize rare recessive diseases in diverse populations using biobank-scale data 1649

Board 6004T: A Functional Analysis of Ramp Sequences in Data from the Clinical Proteomic Tumor Analysis Consortium 1650

Board 6005T: A genome-wide association study identifies genetic regulators of metabolism at birth 1651

Board 6006T: A genome-wide association study of human blood metabolite levels in the Japanese population elucidates genetic metabolic mechanisms 1652

Board 6007T: A novel ancestry-specific eQTL resource of healthy lung tissue from African Americans 1653

Board 6008T: A zebrafish crispant model of *crim1* loss of function shows similarity to mouse models 1654

Board 6009T: Accurate polygenic association of disease to individual cells in single-cell RNA seq data 1655

Board 6011T: An Integrative Single-Cell Allele-Specific Gene Expression and Open Chromatin Analysis Method ⭐ 1656

Board 6013T: Application of Deep Learning Algorithms reveals Copy Number Variations (CNVs) Outside of Chr21 that associate with Congenital Heart Defects in Down Syndrome Children 1657

Board 6014T: Application of PEXDETEX to predict poison exons in publicly available databases 1658

Board 6016T: Cell type-specific RNA-sequencing analysis identifies KCNJ6 as a novel gene associated with Alzheimer's disease in East Asian populations 1659

Board 6017T: Cell-type-specific effects of chromosome 15q11-q13 duplications in human brain 1660

Board 6018T: Cell-type-specific mapping of enhancer-gene pairs from multimodal single cell data ⭐ 1661

Board 6019T: *CNTN2* is a candidate modifier in a PIGA-CDG pedigree with reduced penetrance 1662

Board 6020T: Colocalization of a bipolar disorder type 1 GWAS signal with expression and splice QTLs suggests a causal role of increased expression and cryptic splicing of *CACNA1B* in disease risk 1663

Board 6021T: Comparison of BabySeq's novel gene selection approach to other newborn genomic sequencing programs' gene lists 1664

Board 6022T: Comprehensively comparison and application of ancestry-specific eQTL from an admixed population. 1665

Board 6023T: Context-aware single-cell multiome approach identified cell-type specific lung cancer susceptibility genes ⭐ 1667

Board 6024T: CRISPR screens to unravel the role of candidate genes associated with blood pressure by GWAS on endothelial cell functions ⭐ 1668

Board 6025T: Cryptic donor splice-site in non-coding region of *HADHA* leads to diagnosis of Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (LCHAD) 1669

Board 6026T: De novo deletion variant in the GTPase effector domain (GED) of DNM1L leads to Gross motor delay, hypotonia, axonal sensory neuropathy with peroxisomal and mitochondrial abnormalities 1670

Board 6027T: Detection of a pathogenic *PRKN* inversion by a novel DNA-RNA combined analysis of structural variants 1671

Board 6028T: Discovering the molecular basis of *TBR1*-associated autism 1672

Board 6029T: Discovery of noncoding mobile element insertions associated with aberrant splicing signatures ⭐ 1673

Board 6030T: Distinguishing environmental variance from measurement noise in plasma protein levels 1674

Board 6031T: Effect of the rs1050828 G6PD variant on HbA1c-based risk assessment for diabetes and its complications in a model incorporating polygenic scores..... 1675

Board 6032T: Efficient Demultiplexing and Dynamic Allele-Specific Expression Quantification for Single-Cell RNA-seq and ATAC-seq Data 1677

Board 6033T: Endocrine and reproductive phenotypes in the *Sox2* heterozygous knockout mouse..... 1678

Board 6034T: Estimating the proportion of disease heritability mediated by molecular traits..... 1679

Board 6035T: Evolutionary loss of chitinolytic activity in YKL-40 (HC-gp39, CHI3L1) is caused by multiple sequence modifications that did not affect its affinity to chitin... 1680

Board 6036T: Exploration in a Latin American sample of the genotype-specific effect of eQTLs on expression of *PM20D1*, a gene linked to Alzheimer's disease..... 1681

Board 6037T: Expression of Alzheimer's genetic risk genes in glial cells correlates with amyloid beta in human brain ⭐ 1682

Board 6038T: Functional analysis of the chr13q12.2 pancreatic cancer risk GWAS locus reveals allele-specific regulatory effects of rs9581943 on *PDX1* gene expression 1683

Board 6039T: Functional assessment of Dishevelled pathogenic variants in Robinow syndrome..... 1684

Board 6040T: Functional insights into the divergent role of *PIGF* in adipogenesis and osteogenesis..... 1685

Board 6041T: Genetic Editing of Archaic Alleles into Human Stem Cell Models: NOTO and Cortical Development 1686

Board 6042T: Genetic insights in FCGR2/3 locus: Role in susceptibility to Kawasaki Disease 1687

Board 6043T: Genetic risk for obesity predicts weight loss outcome in bariatric surgery through open chromatin variants regulating adipocyte gene expression..... 1688

Board 6044T: Genome-wide association study identifies genetic variants associated with induced pluripotent stem cell growth rates..... 1689

Board 6045T: Genomic deletions of candidate enhancers at QT interval associated loci 1690

Board 6046T: GSAP: A Comprehensive Pipeline Integrating Genomics and Structural Biology for Gene and Variant Impact Analysis 1691

Board 6047T: Heritability of transcript splicing and the development of splicing genetic scores for biobank-scale TWAS analyses..... 1692

Board 6048T: *HGSNAT, KDM6B, LMNA* and *WFS1* genes pathophysiological heterogeneity for neurodevelopmental disorders reveals novel, rare and ultra-rare variants in four unrelated probands of consanguineous families. 1693

Board 6049T: Human Inner Ear Organoids Validate Coding and Enhancer Variants in Cochlear Malformations 1694

Board 6050T: Human synonymous codon variation is constrained within multivalent coding regions..... 1696

Board 6051T: Identifying cell state-specific genetic effects on gene expression in human microglia..... 1697

Board 6052T: Identifying risk variants in the 3' untranslated regions affecting Alzheimer's Disease by integrative analysis of eQTL, microRNA, and protein binding profiling 1698

Board 6054T: Integrating Machine Learning and Functional Genomics to Prioritize CVD-associated Variants Altering Cardiac TF Function ⭐ 1699

Board 6055T: Integration of proteogenomics in cerebrospinal fluid and plasma with multiple GWAS identifies tissue-specific mechanisms dysregulated in Alzheimer's and other neurological traits ⭐ 1700

Board 6056T: Integrative omics reveals the genetic and molecular landscape of immune responses and comorbidities in people living with HIV 1701

Board 6057T: Interpretable GWAS by linking clinical phenotypes to quantifiable immune repertoire components 1702

Board 6058T: Investigating the impact of non-coding structural variants downstream of epilepsy-associated gene, *FOXG1* on chromatin architecture and gene regulation ... 1703

Board 6059T: Investigating the mechanisms of cell-type-specific alternative polyadenylation ... 1704

Board 6060T: KIR associations with plasma protein levels and phenotypes in the UK Biobank ... 1705

Board 6061T: Loss of EPHB4 function leads to hyperproliferative phenotypes in model systems of central conducting lymphatic anomaly ⭐ ... 1706

Board 6062T: Metabolome-wide association of carotid intima media thickness identifies FDX1 as a determinant of cholesterol metabolism and cardiovascular risk in Asian populations. ... 1707

Board 6063T: Metabolomic profiling of the entire UK Biobank enables interpretation of Mendelian and complex disease mutations ⭐ ... 1708

Board 6064T: Modeling heterogeneity in single-cell perturbation states enhances detection of response eQTLs ... 1709

Board 6065T: Multi-ancestry brain pQTL fine-mapping and integration with GWAS of 21 neurologic and psychiatric conditions ⭐ ... 1710

Board 6066T: Multiple target proteins with altered-recycling result in varieties of pathological cellular processes, establishing the complex phenotypes of Ritscher-Schinzel syndrome 3 ... 1711

Board 6067T: Multi-study pQTL analysis of circulating aptamer-based proteomics in multi-ancestry TOPMed Cohorts ... 1712

Board 6068T: Non-linear genetic regulation of the blood plasma proteome ... 1714

Board 6069T: Optical Pooled Phenotyping at Nucleotide Resolution ⭐ ... 1715

Board 6070T: Pervasive human mitochondrial-nuclear interactions and their implications for disease identified across >600,000 individuals ... 1716

Board 6071T: Predicting missense mutation effects with biophysics-informed protein language model ... 1717

Board 6072T: Predicting the functional impact of missense variants using a multi-gene classification model ... 1718

Board 6073T: Predicting, understanding, and rescuing pathogenic vasopressin 2 receptor variants ... 1719

Board 6074T: PreMode predicts mode-of-action of missense variants by deep graph representation learning of protein sequence and structural context 1720

Board 6075T: Quantifying the effects of rare genomic variants affecting cell types on Autism spectrum disorders across time thanks to brain single-cell burden analysis . 1721

Board 6076T: Revealing disease mechanisms driven by BAF chromatin remodeling complexes using computational structural and functional genomics 1722

Board 6077T: Saturation genome editing based functional evaluation and clinical classification of BRCA2 single nucleotide variants ⭐ 1723

Board 6078T: Sex-stratified Association of Variants in the Serotonin 1A Receptor Gene with Acute Crisis Pain among African American Sickle Cell Disease Patients 1724

Board 6079T: Shift augmentation for improved indel scoring in DNA sequence-based machine learning models 1725

Board 6080T: Single-nuclear profiling of chromatin accessibility and gene expression in human liver reveals genetic effects on cell-type gene regulation 1726

Board 6081T: Small but potentially clinically-relevant contribution of CNVs to idiopathic nephrotic syndrome 1727

Board 6082T: Splice rescue of protein-truncating variants is associated with reduced selection pressure and attenuated disease phenotype 1729

Board 6083T: Systematic targeting and inhibition of bone mineral density GWAS loci in single cells identifies novel regulatory networks for human bone formation ⭐ 1730

Board 6084T: Telomere length in Colombian Centenarians (COOLCEN cohort) 1731

Board 6085T: Tissue specific modifiers of stenosis in Williams-Beuren syndrome 1732

Board 6088T: Unproductive splicing QTLs implicate post-transcriptional regulatory variants in Alzheimer's Disease 1733

Board 6089T: Unraveling Rare Diseases: The NIH Undiagnosed Diseases Program's Integrated Approach to Genetic Diagnosis and Treatment 1734

Board 6090T: Unraveling the role of WNT2B in ocular development 1735

Board 6091T: Utilizing large-scale whole genome sequencing data from the All of Us research program to discover novel genes and endogenous retroviral loci associated with female infertility and recurrent pregnancy loss 1736

Board 6092T: Variability of variant pathogenicity prediction accuracy across genes associated with orofacial clefts 1737

Board 6093T: VISTA: An integrated framework for structural variant discovery 1738

Session Title: Omics Technologies Poster Thursday Session 1740

Board 1100T: A comparison between traditionally used variant detection techniques and long-read whole genome sequencing in a clinical setting 1740

Board 1101T: A curated research catalogue of structural variation detected by optical genome mapping 1741

Board 1102T: A graph-based method to characterize tandem repeats and methylation for long read genome sequencing ⭐ 1742

Board 1103T: A Method for genome-wide analysis of short tandem repeats' impact on human gene expression using massively parallel reporter assays ⭐ 1743

Board 1104T: A Modular Workflow with Minimal QC Steps Provides Improved Capture Performance for Intact gDNA/total RNA as well as Low Input FFPE. 1744

Board 1105T: A multi-dimensional approach to map disease relationships challenges classical disease views 1745

Board 1106T: A streamlined enzymatic method for whole genome methylation detection 1746

Board 1108T: An Integrative Approach to Estimate the Adverse Impacts of Butylparaben on Othologous Human Genes using *Daphnia magna*via a Two-step BLAST Strategy and Network Analysis 1747

Board 1109T: Assessment and comparative analysis of novel single cell RNASeq platforms 1748

Board 1110T: Automated and scalable single technology telomere-to-telomere assembly of human genomes 1749

Board 1111T: Automated Spatially Resolved Laser-Activated Cell Sorting (SLACS) for In-depth Analysis of Heterogeneity in Mouse Brain 1749

Board 1112T: Automation of time-consuming WES steps and use of LLMs 1750

Board 1113T: Benchmarking of Straglr for the detection of short-tandem repeats using high-coverage Cas9-enriched nanopore sequencing data 1752

Board 1114T: Beyond genomics: using RNAseq to unlock the clinical relevance of splicing variation. 1752

Board 1116T: Brain-regional gene expression imputed from the blood transcriptome by *BrainGENIE* recapitulates dysregulation observed in the *postmortem* brain in Alzheimer's disease 1753

Board 1117T: Causal inference of drug effects on metabolomics and proteomics in UK Biobank 1754

Board 1118T: Caution when using network partners for target identification in drug discovery 1755

Board 1119T: Charting inflammasome-driven immune infiltration in homologous recombination deficient high grade serous ovarian cancer 1757

Board 1120T: Circulating cell-free DNA as a novel genomic biomarker of immune and organ dysfunction during acute infection 1758

Board 1121T: Clinical blended genome exome (cBGE) for translational genomics 1759

Board 1122T: Comprehensive Study and Implementation of Enhanced Healthcare Data Management Practices in Korea 1760

Board 1123T: Context-Dependent Embedding and Interpretation of Genomic Data Using Large Language Models 1761

Board 1124T: Current genomic deep learning models display decreased performance in cell type specific accessible regions 1762

Board 1125T: De novo mutation identification without parental data: Implications for rare disease gene discovery and heritable effect of ionizing radiation 1763

Board 1126T: Deep learning-based multi-omics data integration framework identifies possible drug-repurposing gene targets for sarcopenia 1764

Board 1127T: Designing Cell-Type-Specific Promoter Sequences Using Conservative Model-Based Optimization 1765

Board 1128T: Detecting differential RNA editing sites using single-cell RNA sequencing data 1766

Board 1129T: Detection of repeat expansions with PureTarget 1767

Board 1130T: Differential diagnosis of rare genetic diseases aided by large language model 1768

Board 1131T: Elucidating genetic mechanisms directing in vitro blood cell formation with single cell multiomics 1769

Board 1132T: Enabling low cell input and high throughput processing of cell and nuclei fixation for single-cell RNA profiling using combinatorial barcoding 1770

Board 1133T: Enhanced Library Preparation Methods for Targeted Oxford Nanopore Technologies (ONT) Long-Read Sequencing 1771

Board 1134T: Estimating gene regulatory noise by measuring gene expression dispersion to assess robustness across cell types using single-cell gene expression data 1772

Board 1135T: Evaluating transcriptomic signatures in post-mortem brain tissue of ALS and related disorders using innovative long-read sequencing technologies..... 1773

Board 1136T: Exploring Pediatric Disease Genomics with the Upgraded Variant WorkBench in the Kids First Data Resource Portal 1774

Board 1137T: Feasibility of a Complete PCR-free Whole Exome Sequencing Using a New Integrated Library Preparation and Sequencing Workflow 1775

Board 1138T: From pixels to genes: leveraging spatial transcriptomics for single-cell resolution gene expression prediction in histology images 1776

Board 1139T: From Validation to Diagnosis: Metadata Analysis of Long-Read Sequencing for Clinical Insights 1777

Board 1140T: Haplotype calculation with SNV and Structural Variations for long-read sequencing data 1778

Board 1141T: Human pancreatic islet single cell transcriptomes reveal convergent T2D genetic and environmental β-cell gene expression effects. 1779

Board 1142T: In vivo-in vitro genomic comparison reveals TNFa as a missing factor that supports in vitro hematopoiesis..... 1780

Board 1143T: Infrared-laser based single cell sorting for advancing spatial omics..... 1781

Board 1144T: Integrating single-cell transcriptome and epigenome information with spatial analysis at scale with massively parallelized barcoding 1782

Board 1145T: Interplay of genetic and environmental perturbations on the human plasma proteome 1783

Board 1146T: JARVIS 2.0: A non-coding variant effect predictor based on 500,000 UK Biobank WGS samples that surpasses existing benchmarks for prioritisation of common variants ⭐ 1785

Board 1147T: Leveraging deep neural networks to identify single-cell gene expression profiles that characterize Parkinson's disease 1786

Board 1148T: Lifespan cellular dynamics and disease associations in the human prefrontal cortex ⭐ 1787

Board 1149T: Long-read sequencing enables structural variant discovery and identification in haptoglobin. 1788

Board 1150T: MAJIQ-CLIN: A novel tool for the identification of Mendelian disease-causing variants from RNA-seq data ⭐ 1789

Board 1151T: Maternal and Fetal Genetic Associations between the Circulating Proteome and Preeclampsia ⭐ 1790

Board 1152T: Modeling nascent transcription from chromatin landscape and structure 1791

Board 1153T: Mosaic Variant Detection from Whole Genome Sequencing 1792

Board 1154T: Multi-year follow-up analysis of infants and children post-rapid whole genome sequencing 1793

Board 1155T: MuVEH and mitoMuVEH improve discovery of genetic variation from single cells 1794

Board 1156T: needLR: A structural variant filtering and prioritization tool for long-read sequencing data 1795

Board 1157T: Neural transcription factors identified driving brain organoid maturation 1796

Board 1158T: Novel Drivers of α1A Adrenergic Receptor-mediated Heart Failure Identified through Correction for Cell-Type Shifts in Bulk Cardiac Transcriptomes 1797

Board 1159T: Optimization and quality control of library preparation in bulk and single-cell ATAC-sequencing 1798

Board 1160T: Optimized workflows for custom targeted enrichment, from oligo probe design to NGS data analysis 1799

Board 1161T: Orthogonal validation of sequencing in the clinical setting: Lessons from assessing concordance between whole genome sequencing and genotyping arrays with 133,000 samples 1800

Board 1162T: Piloting the 48-hour sequencing workflow for BadgerSeq: a decentralized model for ultra-rapid long-read diagnostic genome sequencing 1801

Board 1163T: PRISM: Refraction of Causal Analyses upon Proteomics Data Into Bands of Efficacious Drug Targets 1802

Board 1164T: Retrieval Augmented Generation Greatly Improves the Ability of Large Language Models to Precisely Assign Human Phenotype Ontologies to Clinical Data 1803

Board 1165T: Saliva DNA can support long read sequencing technologies to improve resolution of high complexity genomics regions 1805

Board 1166T: Sample and Cell-type specific inference from bulk RNA-sequencing data via EPIC-unmix 1806

Board 1167T: SCUBA enables consistent visualization of single-cell sequencing data across storage formats and supports the development of more flexible analysis tools 1807

Board 1168T: Sex differences in the human plasma proteome ⭐ 1808

Board 1169T: ShinyMagnifying: A Comprehensive Shiny-Based Toolbox for Single-Cell RNA-Seq Analysis 1809

Board 1170T: Simultaneous single cell sequencing of genetic and epigenetic bases. 1810

Board 1171T: Single-cell based integrative analysis of transcriptomics and genetics reveals robust associations and complexities for inflammatory diseases 1811

Board 1172T: Single-cell multi-omics analysis reveals mucosal immune cell profiles in ulcerative colitis 1812

Board 1173T: Single-cell RNA sequencing analysis of malignant transformation and carcinogen response heterogeneity in lung cancer 1814

Board 1174T: Single-cell RNA sequencing of plasmacytoid dendritic cells implicates inflammatory pathways in primary graft dysfunction after heart transplantation..... 1815

Board 1175T: SLACS (Spatially-resolved Laser Activated Cell Sorting) Empowers Primary Template-Direct Amplification 1816

Board 1176T: ST-COS: Simulator of Spatial Transcriptomic Data Preserving Gene-Gene Co-expression Patterns..... 1817

Board 1177T: Succinic semialdehyde dehydrogenase deficiency: Age-dependent metabolomic profile of diagnostic biomarkers in a cross-sectional study 1818

Board 1178T: Transgenic Humanized Mouse Models for Disease Research: Leveraging Bxb1 Integrase and New Safe Harbor Loci 1819

Board 1180T: Unbiased multiplex antigen screening of Cerebrospinal Fluid detects microbial and autoantigenic epitopes associated with Multiple Sclerosis ⭐ 1820

Board 1181T: Unraveling alveolar macrophage plasticity and metabolism in hypoxia-induced pulmonary hypertension. 1821

Board 1182T: Unraveling neurodevelopmental phenotypes in individuals with complex copy number and structural variants via integrated genomic, transcriptomic and epigenomic analyses 1822

Board 1183T: Unsupervised evaluation of pre-trained DNA large language models embeddings 1823

Board 1184T: Using donor-specific assemblies as a benchmark for short- and long-read sequencing platform comparisons within the SMaHT network 1824

Board 1185T: VariLink - a comprehensive assay for cancer genome characterization and its comparison to WGS 1825

Board 1186T: When Proteins Mislead: The Epitope Effect in pQTL Analysis 1826

Session Title: Pharmacogenomics Poster Thursday Session 1827

Board 7001T: Anthracyclines induce global changes in chromatin accessibility in cardiomyocytes that overlap with cardiovascular disease loci 1827

Board 7002T: Building Gene-drug Clinical Validity Framework for integrating Pharmacogenomics into Genomic Medicine 1828

Board 7003T: Characterizing UGT1A1 promoter TA repeat polymorphisms and searching for additional polymorphisms that regulate hepatic UGT1A1 expression 1830

Board 7004T: Chrono therapeutic effect of dietary intervention on modern lifestyle and association with circadian gene expression 1831

Board 7005T: Cross platform comparison of pharmacogenomics reporting from whole-genome sequencing 1832

Board 7006T: CYP2D6*71 is a poor metaboliser allele common in Polynesian and Māori people and absent from Europeans 1833

Board 7007T: Diagnosis and Treatment of rare cases of PH1 in elderly patient with chronic kidney disease 1834

Board 7008T: Drug Repurposing for Candidiasis-Associated Diabetic Foot Ulcer Based on Data-Driven Computational Approach 1835

Board 7010T: Genetic variants associated with adverse effects of NSAIDs in a Korean rheumatology cohort 1836

Board 7011T: Genetically enriched clinical trials for precision development of non-cancer therapeutics: A scoping review 1837

Board 7012T: Genome-wide profiling of TOP2B binding in human cardiomyocytes to understand Doxorubicin-induced cardiotoxicity ⭐ 1838

Board 7013T: Germline genetic factors impacting endocrine toxicities in patients treated with immune checkpoint inhibitors ⭐ 1839

Board 7014T: Identifying epigenetic influences on hepatic gene expression: MeQTLs in African Americans that associate to disease and drug response ⭐ 1840

Board 7015T: Integrating Muti-Omics Data to Decode the Heterogeneity in Antidepressant Response ⭐ 1841

Board 7016T: Opioid Prescription Pattern and Patient-level Disease Polygenic Risk Score 1842

Board 7017T: Pharmacogenetic insights into hypertension with intracranial aneurysm as an outcome: Implications for precision medicine..... 1843

Board 7018T: Pharmacogenetic study of metformin response in type 2 diabetes treatment..... 1845

Board 7019T: Pharmacogenomic Landscape of the UAE: Insights into Star Allele Frequencies and Actionable Variants Using a Large Cohort from the Emirati Genome Project..... 1846

Board 7020T: PharmGKB tools for accessing clinical pharmacogenomics prescribing guidance by genotype 1847

Board 7021T: Plausible genetic factors mostly do not explain heterogeneity in GLP1-RA and bariatric surgery-associated weight loss in 10 000 individuals from 9 multi-ancestry biobank studies..... 1848

Board 7022T: Resolution of *CYP2D6* pharmacogene complex haplotypes with end-to-end Oxford Nanopore sequencing workflows..... 1849

Board 7023T: The demographic burden of pharmacogenetic prescribing: the case for panel pharmacogenomic testing to ensure equitable healthcare 1851

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session 1852

Board 7025T: Advancements in non-invasive prenatal testing (NIPT): Integrating NGS for the diagnosis of congenital abnormalities 1852

Board 7026T: Analysis of Differential Gene Expression in Umbilical Cord Blood and Peripheral Blood Mononuclear Cells from Healthy Children..... 1853

Board 7027T: Deregulated MicroRNAs (miRNAs) as Promising Biomarkers for Monitoring Preeclampsia (PE) in Pregnancies 1854

Board 7028T: Disruption and Amplification of *PRKN* gene (505 kb) in a case of fetal demise with a normal karyotype 1855

Board 7029T: Enhancing Fetal Fraction in Non-Invasive Prenatal Screening (NIPS) Through Differential Methylation Status ⭐ 1856

Board 7030T: Exome-Based Genetic Screening for Preconception in the Mexican Jewish Community 1857

Board 7031T: Fetal carriage of variants near *GCLC* are associated with altered levels of multiple placentally-derived proteins in maternal serum across gestation ⭐ 1858

Board 7032T: Genes encoding inhibitory KIR contribute to pre-eclampsia risk in Uganda 1859

Board 7033T: Genetic burden of glycaemic dysregulation is associated with early-life cognitive development and neurology-related protein expression 1860

Board 7034T: Genetic Link Between Androgen Receptor Repeats and Sex Hormones Levels in Premature Ejaculation. 1861

Board 7035T: Genome-wide association study implicates genes related to development of cardiovascular system and immunology in predisposition to preeclampsia..... 1862

Board 7036T: Human placental gene expression at different collection sites..... 1863

Board 7037T: Incidental detection of maternal cancer in prenatal cell-free DNA screening: Clinical laboratory experience and qualitative approach to Multiple Structural Variants ⭐ 1864

Board 7038T: Investigation of clinical and genetic heterogeneity in human fetuses with arthrogryposis multiplex congenita ⭐ 1865

Board 7039T: Phenotypic characterization of Global Developmental Delay and Intellectual Disability of genetic etiology. 1866

Board 7040T: Placental dysfunction is associated with rare damaging variants in the stem cell differentiation pathway 1867

Board 7041T: Quantifying the proportion of women at risk of an FNAIT pregnancy in diverse populations 1868

Board 7042T: Reconstructing clonal dynamics of the human neural crest by genetic mosaicism 1869

Board 7043T: Sex differences in placental gene expression and intraplacental variation in normotensive versus hypertensive pregnancies 1870

Board 7044T: spatiotemporal transcriptomic map of non-human primates at early development from the dGTEx project 1871

Board 7045T: Study of the mutational landscape of premature ovarian failure by whole exome sequencing in Bangladeshi women 1872

Board 7046T: Use of single cell genomic strategies to prioritize candidate genes in *Sox10*^{Dom} aganglionosis genetic modifier intervals 1873

Board 7047T: Validation of fetal and maternal Spinal Muscular Atrophy (SMA) and Hemoglobin (Hb) Barts screening with FirstGene, a combined non-invasive prenatal cfDNA assay for fetal aneuploidy, recessive diseases, and serological screening 1874

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session 1876

Board 4002T: (less un)Realistic Simulation of Complex Trait Architectures for Open Benchmarking of Statistical Genetic Estimators 1876

Board 4003T: A Bayesian Approach to Correcting the Attenuation Bias of Regression Using Polygenic Risk Score 1877

Board 4004T: A community of networks approach for multi-omics integration..... 1877

Board 4005T: A computationally efficient pipeline to perform rare variant burden tests with imputed data in the 23andMe database 1878

Board 4006T: A deep generative network for genotype-phenotype prediction using high-content zebrafish imaging 1879

Board 4007T: A deep learning algorithm to predict risk of coronary artery disease using genetics and electronic Health record trajectory 1880

Board 4008T: A framework to maximise genetic diversity in genome wide association study meta-analyses 1881

Board 4009T: A knockoff-based framework to identifydisease-causing genes in conditionaltranscriptome-wide association studies 1882

Board 4010T: A meta-analysis of electronic health record-based cohorts identifies rare burden signals associated with amyotrophic lateral sclerosis 1883

Board 4011T: A new multi-trait fine-mapping method using a non-local prior, with applications in circulating metabolic biomarker level analysis 1884

Board 4012T: A new pipeline for Killer-cell Immunoglobulin-like Receptor (KIR) genotyping from Illumina short-reads 1885

Board 4013T: A novel method to build context-specific polygenic scores improves prediction in complex human traits ⭐ 1886

Board 4014T: A novel multivariable Mendelian randomization framework to disentangle highly correlated exposures with application to metabolomics 1887

Board 4015T: A Novel Varying Coefficient Model for Identification of Latent Effect Modification in Genomic Studies 1888

Board 4016T: A per-domain study of rare variant burden in Parkinson's Disease risk genes 1888

Board 4017T: A scalable approach to uncovering non-additive genetic associations in large biobanks. ⭐ 1889

Board 4018T: A unified genealogical measure of genetic similarity accounts for relatedness and admixture in GWAS ⭐ 1891

Board 4019T: Analysis of maturity-onset diabetes of the young pathogenic gene variants indicate potential oversight in diagnosis 1892

Board 4020T: Analysis of parent-of-origin effects in over 100,000 individuals in the UK Biobank ⭐ 1893

Board 4021T: Analysis of rare exome sequenced variants in UK Biobank to discover causal genes and fine-map causal variants for lung function 1893

Board 4023T: Benchmarking DNA foundation models for genomic sequence classification 1895

Board 4024T: Benchmarking of Long Read vs Short Read Low-Coverage Human Whole Genome (lcWGS) Sequencing for Genotype Imputation and Genetic Relationship Analysis 1895

Board 4025T: Calibration of polygenic risk scores in African Americans using estimated genetic ancestry 1896

Board 4026T: CaseControlAF: An R package for reconstructing case and control allele frequencies from GWAS summary statistics 1898

Board 4027T: Cell Type-Specific contributions of PRS of 34 complex traits 1899

Board 4028T: Challenges associated with the use of biobank data as external controls in genome-wide association studies 1900

Board 4029T: Characteristics of rare variants relevant for dilated cardiomyopathy depend on local ancestry: The DCM Precision Medicine Study 1901

Board 4030T: Characterizing Genetic Diversity in the All of Us Research Program in the Context of Genetic Association Studies 1902

Board 4031T: Chasing Shadows: Identifying genes for undiagnosed rare diseases through statistical analysis of a large hospital cohort ⭐ 1903

Board 4032T: Clinical Utility, Feasibility and Cost-Effectiveness of Polygenic Risk Scores for Breast Cancer Screening in Europe 1904

Board 4033T: Coding cis pQTLs from proteogenomic data allow evaluation of the performance of missense variant effect predictions and the utility of their application to rare variant association analyses. 1905

Board 4034T: Common Variants on Amerindigenous Local Ancestry Share Similar Causal Effects with Those on Other Ancestries in Complex Traits 1906

Board 4035T: Comparison of genetically regulated prediction models in lipidomics and transcriptomics in a Hispanic cohort..... 1907

Board 4036T: Comparison of methods for accurately estimating hereditary cancer prevalence in diverse U.S. biobanks in the presence of participation bias 1908

Board 4037T: Comprehensive evaluation of pleiotropy methods leveraging GWAS summary statistics 1910

Board 4038T: Considering confounding in rare variant genome wide association studies ⭐ 1911

Board 4039T: Cox proportional hazards mixed models accurately estimate the heritability of time to event traits 1911

Board 4040T: Deciphering causal protein biomarkers in Alzheimer's disease: Integrating a novel robust Mendelian randomization method for proteomics data analysis and AlphaFold3 for predicting 3D structural alterations 1913

Board 4042T: Decoding the shared genetic structure of heterogeneous complex diseases 1913

Board 4043T: Deep learning models of gene regulation and the application in Cross-ancestral PRS in Alzheimer's Disease..... 1914

Board 4044T: Detection of Large Somatic Copy Number Variation in Normal Tissue Through Single-Cell Multiome Sequencing 1916

Board 4045T: Detection of non-additive biallelic effects on trait variation through meta-analysis of 552,042 individuals ⭐ 1917

Board 4046T: Differentiating germline and somatic sources of apparent homozygosity 1918

Board 4047T: Discovery of core genes for type 1 diabetes via genome-wide aggregation of *trans*-effects using proteomics studies ⭐ 1919

Board 4050T: Dissecting cross-population polygenic heterogeneity of multimorbidity across respiratory & cardiometabolic diseases 1920

Board 4051T: Distinct Genetic Architecture in the Tails of Complex Traits 1921

Board 4052T: Effect of Polygenic Scores for Serious Mental Illness on Cell Type-Specific Gene Expression in The Brain 1922

Board 4053T: Efficient genealogical compression for biobank-scale datasets 1923

Board 4054T: Efficient saddlepoint approximation in generalized linear model for large-scale genetic association studies 1924

Board 4055T: Enhancing non-linear TWAS performance via trait imputation with applications to Alzheimer's disease 1925

Board 4056T: Epistatic SNPs land in regulatory regions and possess a similar risk as GWAS SNPs in a common liver disease, MASLD 1926

Board 4057T: Proteome-wide Mendelian randomisation of 2,167 disease outcomes in UK Biobank and FinnGen 1927

Board 4058T: Estimating the environmental effect of social determinants of health and race/ethnicity on human trait and disease outcomes 1929

Board 4059T: Evaluating common practices for Mendelian randomization studies in East Asians: current practices, challenges and recommendations 1930

Board 4060T: Exploring platform validation of modern microarrays with applications in polygenic risk scoring at-scale in a large institutional biobank 1931

Board 4061T: FastGxC: A fast and powerful statistical method for mapping context-specific regulatory variants in bulk and scRNA-seq studies ⭐ 1932

Board 4062T: Finding Structure In Gene Regulatory Networks: A Graph Neural Network Algorithm for Hierarchical Community Detection 1933

Board 4063T: FISHNET: Finding Significant Hits In Networks 1934

Board 4064T: Gene-Environment and Metabolite-Environment Interaction Effects on Intraocular Pressure in the Canadian Longitudinal Study on Aging 1935

Board 4065T: Genes, Machines, and Missteps: The Unseen Risks of Machine Learning Derived Phenotypes 1936

Board 4066T: Genetic & phenotypic associations of frailty with cardiovascular indicators & behavioral characteristics 1937

Board 4067T: Genetic overlap between impaired kidney function and respiratory function: a large-scale genome-wide cross-trait analysis 1938

Board 4068T: Genome-wide aggregated *trans*-effects analysis on the risk of polymyalgia rheumatica and giant cell arteritis in UK Biobank 1939

Board 4069T: Genome-wide association study of fibrotic disease across organs and systems 1940

Board 4070T: Identification of moderate effect size genes in autism spectrum disorder through a novel gene pairing approach 1941

Board 4071T: Identification of potential drug targets for cardiac congenital anomalies through proteome-wide mendelian randomization analysis 1942

Board 4072T: Identification of Target Patient Populations Using Genetic Proxies of Drug Targets and Time-to-event analyses 1943

Board 4073T: Identifying Rheumatoid Arthritis Gene-regulatory Enhancers with RNA+ATAC Multiome Data 1944

Board 4074T: Imp-Map: A novel biobank scalable approach to fine-mapping and imputation 1945

Board 4075T: ImputeMLP: An efficient and accurate deep-learning method for genotype imputation method 1946

Board 4076T: Incorporating additive genetic effects and full LD information to discover genome-level gene-environment interactions with summary statistics of complex traits 1947

Board 4077T: Integrating genetic data and electronic health records for health risk predictions using a graph convolution network 1948

Board 4078T: Integration of CNVs with SNPs enhances exome-wide association study analysis of 770,000 samples ⭐ 1949

Board 4079T: Integration of multi-omics data & eQTL summary statistics with omicsSTAAR boosts power in rare variant association tests of non-coding regions ⭐ 1950

Board 4080T: Interaction analysis of smoking, prostate specific antigen, particulate matter, and polygenic risk score on prostate cancer incidence in Korea 1951

Board 4081T: Intraplaque haemorrhage prediction and MACE risk modelling using attention based multiple instance learning 1952

Board 4082T: Investigating the determinants of frailty: The intersection of genetics and environment in Mexican Americans from South Texas 1954

Board 4083T: LEGO: Local ancestry Enable predicting Gene expression Optimally ... 1955

Board 4085T: Leveraging causal cell-type-specific regulatory annotations improves the accuracy of functionally-informed fine-mapping in complex traits 1956

Board 4086T: Leveraging genetic and proteomic data to decipher smoking-induced pathology ⭐ 1957

Board 4087T: Leveraging haplotype information in heritability estimation and polygenic prediction 1958

Board 4088T: Leveraging phenotypic, genetic, and environmental correlations between complex traits to improve risk predictions 1959

Board 4089T: Leveraging protein interaction networks for powerful rare-variant gene set association analysis 1959

Board 4090T: Linkage analyses identify novel loci for Alzheimer's disease in non-Hispanic white families 1960

Board 4091T: localQTL: A GPU-Accelerated Software for Accurate Molecular QTL Mapping in Admixed Populations ⭐ 1962

Board 4092T: Longitudinal changes in epigenetic age acceleration prior to T1D onset in the Diabetes Autoimmunity Study in the Young (DAISY) 1962

Board 4093T: Mapping aging and disease pathways through integrating single-cell gene expression and plasma proteins 1963

Board 4094T: Mendelian randomization borrowing strength via shared pleiotropic pathways across populations 1964

Board 4095T: Mendelian randomization identifies known and novel regulation between enzymes and metabolites 1965

Board 4096T: Mendelian randomization of sex-dependent traits 1966

Board 4097T: Meta-SAIGE: Scalable and accurate meta-analysis for rare variants 1967

Board 4098T: mixTWAS: a framework to identify genes from GWAS using mixed effects to account for relatedness and polygenicity-driven inflation 1968

Board 4099T: Multi-ancestry gene-based association tests identify novel loci and drug targets in minority populations 1969

Board 4100T: Multi-ancestry meta-analysis identifies genetic modifiers of age-at-onset of Alzheimer's disease 1970

Board 4101T: Multi-group causal-TWAS provides a unified framework to integrate multi-omic QTL data from multiple tissues with GWAS ⭐ 1971

Board 4102T: Multi-omics integrative risk model for Alzheimer's Disease in a large-scale biobank 1972

Board 4103T: Multi-trait inference of full genome-wide associations of type 2 diabetes subtypes informs distinctive biology and putatively causal genes ⭐ 1973

Board 4104T: MV-DeLIVR: deep learning based multivariable instrumental variable regression for nonlinear causal inference 1974

Board 4105T: Network Construction using Sparse Gaussian Graphical Model based on GWAS Summary Statistics 1975

Board 4106T: NeuroID: Target discovery engine for Parkinson's disease 1976

Board 4107T: New insights into kidney function: GWAS analysis of MRI-derived kidney volumes 1977

Board 4108T: Optimal Polygenic Risk Score Ancestry Adjustment Occurs Within Individualized Neighborhoods 1978

Board 4109T: Pairwise Likelihood with Latent Variables for Detecting Genetic Variants Associated with Human Complex Traits Using Multiple Types of Phenotypes 1979

Board 4110T: pecotmr: A unified framework for integration of genome wide association studies with molecular quantitative trait loci 1980

Board 4111T: PeReS:personalized reference sets for polygenic risk scores negates need for arbitrary groupings 1982

Board 4112T: PhenoSS: Phenotype semantic similarity-based approach for rare disease prediction and patient clustering 1983

Board 4113T: Powerful, robust and fast whole-genome regression for Time-to-Event Data in large-scale biobanks 1984

Board 4114T: PRS-GRID: A Cross and Within Ancestry Polygenic Risk Prediction Method Based on Individual Genetic Distance 1985

Board 4115T: *Quantile IV*: A nonparametric Mendelian randomization method to estimate heterogeneous and nonlinear causal effects 1986

Board 4116T: Rare and common genetic associations with internalising and externalising behaviours across childhood and adolescence 1987

Board 4117T: Rare variant analysis pipelines: a systematic review..... 1988

Board 4118T: Rare variant intensity estimation for genetic mapping of complex traits 1989

Board 4119T: Relationships Between Physical Activity And Overall Brain And Body Health Demonstrated By Multi-Organ Imaging Genetics 1989

Board 4120T: Removing array-specific batch effects in GWAS mega-analyses by applying a two-step imputation workflow 1990

Board 4121T: Removing confounding and refining mendelian randomization genetic instruments of dietary intake traits 1991

Board 4122T: Retrospective analysis frameworks to control for population admixture and sample relatedness in large-scale genome-wide association studies 1992

Board 4123T: Retrospective set-based gene-environment interaction tests for biobank data 1994

Board 4124T: Revealing ancestry-specific genetic effects on Alzheimer Disease in African Americans 1994

Board 4125T: scFM: An efficient statistical fine-mapping approach for eQTLs using large-scale single-cell data 1996

Board 4126T: Sex-Specific Genetic Overlap Between Multisite Chronic Pain and Asthma by Age of Onset 1997

Board 4127T: simGO: A Comprehensive Simulation Framework for Modeling Relationships Among Genotypes, Omics, and Phenotypes 1998

Board 4128T: SPLENDID improves polygenic risk prediction across the genetic ancestry continuum using diverse biobank data ⭐ 1999

Board 4129T: Spatial multi-modal atlas of bone tissue from human femur 2000

Board 4130T: Splicing-specific transcriptome-wide association uncovers genetic mechanisms for Schizophrenia 2001

Board 4131T: Symptom-level characterization of Gulf War Illness in deployed and non-deployed participants enrolled in the Million Veteran Program 2002

Board 4132T: Synthetic surrogates improve power for genome-wide association studies of partially missing phenotypes in population biobanks 2003

Board 4133T: Systematic Comparison of Type I Error and Power in EMMAX and GENESIS 2004

Board 4134T: The dual influence of reference panel size on genotype imputation accuracy and marker selection 2005

Board 4135T: The effect of population differences on two-sample Mendelian randomization. 2006

Board 4136T: The longitudinal effect of *APOL1* risk alleles on sickle cell disease-associated kidney damage 2007

Board 4137T: The polygenic risk score of biological age in nine human organ systems 2008

Board 4139T: The PRS-RV Causal Pivot: A Segmentation Approach to Complex Disease Genetics 2010

Board 4140T: Transcriptome-wide association analysis identifies novel genes underlying coronary heart disease and stroke using kidney tissue and brain cell type transcriptome in individuals with chronic kidney disease 2011

Board 4141T: Transferability of a European-derived Alzheimer's Disease Genetic Risk Score across Multi-Ancestry Populations. 2012

Board 4142T: Triangulation of the causal relationship between plasma proteins and systemic lupus erythematosus 2013

Board 4143T: TRUHiC: A transformer model incorporating the U-2 Net to enhance Hi-C data for 3D chromatin structure delineation 2013

Board 4144T: Uncovering disease-critical pathways for diabetes through global and local embeddings of CRISPR perturbation data 2015

Board 4145T: Unique Signatures of Highly Constrained Genes Across Publicly Available Genomic Databases 2016

Board 4147T: Univariate and Multivariate Proteome-Wide Association Studies to Identify Causal Proteins for Alzheimer's Disease in the Presence of Invalid Instruments with GWAS Summary Data 2017

Board 4148T: Unravelling the gene regulatory networks driving the polygenetic risk of human complex diseases 2018

Board 4150T: Whole genome sequence-based association study on the multiple system atrophy in the Japanese population 2019

Board 4151T: XPAT2: A Quality Control Framework for Sequencing-Based Association Studies Using Biobank Scale Data ⭐ 2020

Friday, November 8, 2024, 2:30 pm - 4:30 pm 2021

Session Title: Cancer Poster Friday Session 2021

Board 8001F: A deep learning-based precision oncology framework integrating tumor omic and functional characterization for personalized treatment selection 2021

Board 8002F: A Discovery of Novel Methylation Profiles and Race-Specific Methylated Sites for Breast Cancer Risk 2022

Board 8003F: A machine learning approach towards prediction of pancreatic cancer using gene expression and DNA methylation 2023

Board 8005F: A transcriptome-wide association study identifies *PHACTR4* as a potential mediator of prostate cancer survival. 2024

Board 8006F: Adiposity distribution and risks of 12 obesity-related cancers: a Mendelian randomization analysis 2025

Board 8007F: An exploration of sex differences in the genetic architecture of pancreatic cancer using GXwasR 2027

Board 8008F: An innovative approach for detecting cancer-associated germline variants using automated and multiplexed targeted HiFi read sequencing 2028

Board 8009F: Analysis of extensive data towards Genome in a Bottle benchmarks for a broadly-consented tumor normal pair 2029

Board 8010F: Approaches to returning germline and somatic test results across the Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network ⭐ 2030

Board 8011F: Association of polygenic-based breast cancer risk prediction with patient management ⭐ 2031

Board 8012F: Atavistic reversion in cancer through fusion genes and ancient protein structures 2032

Board 8013F: Breast Cancer Risk Polygenic Score optimisation through a novel SNP selection algorithm 2033

Board 8014F: Case Report: Prostate Cancer Gene *HOXB13* Involved in Metastatic Omentum Carcinoma 2034

Board 8016F: Comprehensive somatic mutation analysis of oral leukoplakias 2036

Board 8018F: Comprehensive tumor mutational profiling and associated impact on overall survival outcomes in young women (diagnosed ≤ 40 years old) with breast cancer carrying a pathogenic germline *BRCA1* or *BRCA2* variant in the POSH cohort. 2037

Board 8019F: Decoding *CDH1*: the hidden role of regulatory noncoding sequences in hereditary diffuse gastric cancer 2038

Board 8020F: Defining actionable co-mutations and mutually exclusive genes in*NF1*-altered sporadic breast cancer 2039

Board 8021F: Diverse Genetic Architecture of Neuroblastoma: Insights from a Multi-Ancestry Genome-Wide Association Study ⭐ 2040

Board 8023F: Finding cancer misdiagnosis windows through electronic health records 2041

Board 8024F: From desert flora to cancer therapy: systematic exploration of multi-pathway mechanisms using network pharmacology and molecular modeling approaches 2042

Board 8025F: FusionRank: A machine learning framework for annotating and prioritizing oncogenic gene fusions 2043

Board 8026F: Gene-Environment Interaction: Small Deletions and Transcriptomic Profile in Non-Melanoma Skin Cancer (NMSC) in a Population Exposed to Arsenic 2044

Board 8028F: Germline genetic analysis of 43,248 individuals across nine less-investigated cancer types in *BRCA1* and *BRCA2* 2045

Board 8029F: Germline structural variants as risk alleles for high grade serous ovarian cancer 2047

Board 8030F: Identification of Genetic Network unveils the pathologic mechanism of UTUC:A multi-omics and longitudinalcohortstudy 2048

Board 8031F: Immunoglobulin heavy chain variable region genes usage and mutational status in chronic lymphocytic leukemic patients in Nigeria ⭐ 2049

Board 8032F: Integrated multi-omics drive precision cancer care for pediatric CNS tumors: An institutional study of 582 tumors ⭐ 2050

Board 8033F: Integrative proteogenomic analyses reveal insights into subtype-specific glioma risk 2052

Board 8034F: Investigating the Association Between Clonal Hematopoiesis and Lung Cancer Risk 2053

Board 8035F: Investigating the cell-type specific regulation of IRF4 and its role in lung cancer via a lung cancer risk-associated pleiotropic variant 2054

Board 8037F: Isoform-specific transcriptome-wide association studies of 6 cancers uncover extensive risk mechanisms undetectable by gene-level methods 2055

Board 8038F: Longitudinal characterization of 1,763 mosaic chromosomal alterations identifies factors influencing clonal dynamics of leukocytes ⭐ 2056

Board 8039F: Long-range somatic structural variation calling from matched tumor-normal co-assembly graphs 2057

Board 8040F: Loss of Y chromosome as the only cytogenetic abnormality in myelodysplastic syndromes 2058

Board 8041F: Lynch Syndrome Pathogenic Variant Prevalence in a Population Biobank: an Analysis of Results from the Colorado Center for Personalized Medicine Biobank 2059

Board 8042F: Machine learning approaches to understand regulatory mechanisms of methylation dysregulation 2060

Board 8043F: MicroRNA signature predicts prognosis in patients with acute myeloid leukemia 2061

Board 8044F: Multi-ancestry transcriptome-wide association analyses on single-cell deconvoluted expression data to identify cell-type specific risk genes for colorectal cancer ⭐ 2062

Board 8045F: Nanoparticle-mediated CRISPR-Cas9 Delivery Targeting BCR/ABL for Chronic Myeloid Leukemia 2063

Board 8046F: Novel ETV6 fusion transcripts in acute leukemia were identified by Next Generation Sequencing (NGS) test 2064

Board 8047F: Pangenome references improve biomarker estimation from tumor sequencing data 2065

Board 8048F: Patterns of hepatitis b virus mutations associated with hepatocellular carcinoma is controlled by HLA genotype. 2066

Board 8049F: $PM_{2.5}$ exposure is associated with EGFR somatic mutations in non-small cell lung cancer 2067

Board 8050F: Prediction of anthracycline-induced cardiomyopathy in long-term childhood cancer survivors: the power of polygenic risk scores 2069

Board 8052F: *TERT*p.K1050N: A founder variant in Ashkenazi Jewish populations associated with telomere biology disorders 2070

Board 8053F: The Multiethnic Cohort: A Resource for the study of Genetic and non-Genetic Cancer Risk Across Populations 2071

Board 8054F: The Palestinian founder allele TP53 p.R181C predisposes to high risk of breast cancer but not to the pediatric cancers of Li-Fraumeni or Li-Fraumeni-like syndrome 2072

Board 8055F: The role of pre-diagnostic circulating metabolites in prostate cancer risk: A cross-population meta-analysis of untargeted metabolomic studies 2073

Board 8056F: Towards personalised medicine for Caribbean breast cancer: the Bermuda cohort 2074

Board 8057F: Trio exome sequencing identifies increased incidence of clonal hematopoiesis of indeterminate potential in Bloom Syndrome probands and carriers. 2076

Board 8058F: Unveiling the Gut Microbiome Network Fingerprint: A Deep Learning Approach to Predict Immunotherapy Response in Melanoma Patients. 2077

Board 8059F: Use of national genome strategy to determine rate of germline pathogenic mutations among Emirati women with early onset breast Cancer 2078

Board 8060F: Variation in bulk RNA-seq and estimated cell type proportion using deconvolution when comparing pancreatic cancer samples within the same individual ⭐ 2079

Board 8061F: Whole-genome RNAi screen to identify modulators of prometastatic gene ASAP1 2080

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session 2081

Board 5001F: A Comparative Study of Molecular Landscape of Prion Diseases in Humans Suggests Novel Drug Repurposing Opportunities 2081

Board 5002F: A compendium of splicing variation in the Alzheimer's disease risk gene *TREM2* 2082

Board 5003F: A comprehensive MHC-class-II region association-scan for the study of DQ allotypes and race influence on the risk of FVIII inhibitor development in hemophilia-A patients 2083

Board 5004F: A genome-wide analysis of pleiotropy between morning circadian preference and BMI reveals the tissue-specific rhythmicity of ADCY3 in adipose tissue. 2084

Board 5005F: A genome-wide association study of cancer-related cognitive impairment in patients who received chemotherapy for non-metastatic breast cancer 2085

Board 5006F: A meta-predictor for causal gene identification in GWAS overcomes limitations of existing computational approaches ⭐ 2086

Board 5007F: A multi-omics study of cerebral vasospasm in peripheral blood from patients with aneurysmal subarachnoid hemorrhage 2087

Board 5008F: A phenome-wide association study reveals the impacts of germline copy number variations on multiple traits in Taiwan 2088

Board 5009F: Almost all rare genetic associations identified through whole-genome sequencing occur at loci identified from GWAS ⭐ 2089

Board 5010F: Altered TNF-α response in Down syndrome 2090

Board 5011F: Alzheimer's Disease Problem History Heterogeneity and the Relationship with Mitochondrial Genetic Risk 2091

Board 5013F: Analyses of GWAS and subthreshold loci lead to the discovery of dendrite development and morphology dysfunction underlying schizophrenia genetic risk 2092

Board 5014F: Are GWAS maximizing potential discoveries with non-European ancestry data? 2093

Board 5015F: Assessing Non-Communicable Disease Risk Factors in an Admixed Population: Insights from the UAE Healthy Future Study 2094

Board 5016F: Assessing the role of rare pathogenic variants in heart failure progression by exome sequencing in 8,190 patients 2095

Board 5017F: Association of Genetic Susceptibility Factors with Coronary Heart Disease in a Diverse Cohort: an eMERGE Network Study 2096

Board 5018F: Benchmarking ancestry-aware PRS strategies for optimal prediction of PTSD in admixed samples 2097

Board 5020F: Blood Cell-Type Specific eQTL Analysis in a Multi-Ethnic Cohort Provides Novel Insights into Cardiovascular Disease Risk 2098

Board 5021F: Building the path to clinical polygenic risk scoring for adult-onset autoimmune diabetes across diverse ancestries in the biobank at the Colorado Center for Personalized Medicine 2099

Board 5022F: Characterising genetics of gynecological pain conditions and interstitial cystitis 2100

Board 5023F: Characterizing the Genetic Landscape of Indigenous Australians: Association of Regional-wide Autozygosity with ACR Reveals a novel region in *FTO* Gene 2101

Board 5024F: Charting the molecular and spatial transcriptomic landscape of crypt variability in Inflammatory Bowel Disease 2102

Board 5025F: Common and rare variant associations with diabetic retinopathy in 18,175 cases ⭐ 2103

Board 5026F: Common genetic risk variants in clinical single nucleotide polymorphism data in autism spectrum disorder 2105

Board 5027F: Common polygenic risk for asthma is associated with multiple atopic phenotypes in individuals with inborn errors of immunity and immune dysregulation 2106

Board 5028F: Common variants at 22q12.2 are associated with susceptibility to Tuberculosis 2107

Board 5029F: Context specific cell-cell communications reveal Aβ-related signals in human microglia for Alzheimer's disease 2108

Board 5030F: Contribution of common and rare variants to schizophrenia risk in East and South Asian ancestries ⭐ 2109

Board 5031F: Deciphering the complex cases of Early-Onset Parkinson's Disease: Advanced Insights from Next Generation Sequencing Technologies 2110

Board 5032F: Determination of the function of ZBTB40 in osteoblast maturation 2111

Board 5033F: Development of a Polygenic Risk Score for Accurate Differentiation of Type 1 and Type 2 Diabetes in Adults Using At-Home DNA Testing 2112

Board 5034F: Discovering stimulatory state specific type 2 diabetes GWAS mechanisms with single-cell multi-omics on iPSC-derived fibro-adipogenic progenitors 2113

Board 5035F: Discovery of novel loci for blood pressure from genome-wide meta-analysis accounting for gene-alcohol interactions in African population cohorts 2114

Board 5036F: Dose Optimization of Probucol in Rescuing Clec16a Hypofunction Phenotype in KO Mice with Disrupted Mitophagy 2115

Board 5037F: Elevated Prevalence of MODY Genetic Variants in Gestational Diabetes from the UK Biobank 2116

Board 5038F: Environmentally-Driven gene expression variation influences NAFLD risk in Mexican Americans 2117

Board 5039F: Evaluating genetic and proteomic support for cytokine-targeting therapies of prevalent immunological diseases in the UK Biobank 2118

Board 5040F: Evaluating kidney size as a marker for renal and hepatic health 2119

Board 5041F: Evaluatingthe impact of biological sex on ADHD presentation, prevalence, and genetic risk 2120

Board 5042F: Examination of unreported *NOD2* variants in unsolved autoinflammatory cases 2121

Board 5043F: Excess of rare variants in MODY genes in T2D families from Endogamous Ethnic Groups in India 2122

Board 5045F: Expression and Association study of microRNA 192 levels in an urban *Punjabi* sub-population group with Hypercholesterolemia 2123

Board 5046F: Expression of *APOE4* is greater than *APOE3* but differs between Ancestries. 2124

Board 5047F: Functional dissection of heart enhancers implicated in inherited cardiomyopathies 2125

Board 5048F: Gene Discovery and Biological Insights into Anxiety Disorders from a Multi-Ancestry Genome-wide Association Study of >1.2 Million Participants ⭐ 2126

Board 5049F: Gene dosage architecture of Autism and neurodevelopmental related traits 2128

Board 5050F: Gene prioritization analyses on the multi-ancestry Neuroticism GWAS point to novel risk loci. 2129

Board 5051F: Genetic analysis of variable number of tandem repeats (VNTR) in the Japanese population 2129

Board 5052F: Genetic architecture of white matter hyperintensity volume in diverse Hispanic/Latino adults 2130

Board 5053F: Genetic architecture of white matter revealed by Deep unsupervised learning derived endophenotypes from DTI fractional anisotropy brain images in UK Biobank 2132

Board 5054F: Genetic association analysis of 24-hour rest-activity rhythm 2133

Board 5055F: Genetic Determinants of Lung function Decline: A Multi-Level Analysis of Gene Expression 2133

Board 5056F: Genetic diversity & epidemiology of human adenovirus among hospitalized children with severe acute respiratory tract infection in Guangzhou, China 2135

Board 5057F: Genetic Modifiers of Motion Sickness: Whole Genome Sequencing Analysis of Individuals with a Significant History of Motion Sickness Participating in Clinical Studies 2136

Board 5058F: Genetic relationships and biobehavioral pathways between suicidal behaviors and comorbid mental disorders 2137

Board 5059F: Genetic risk stratification in the Leuven familial pulmonary fibrosis cohort 2138

Board 5060F: Genetic study of Crohn's disease in the US population 2139

Board 5061F: Genetic subtypes of type 2 diabetes are distinguished through the lens of abdominal MRI ⭐ 2140

Board 5062F: Genetic Underpinnings of Coronary Heart Disease in Hispanic/Latino Populations 2141

Board 5063F: Genetic variants associated with diurnal intraocular pressure fluctuation in primary open-angle glaucoma 2142

Board 5064F: Genetic variants in known genes in diverse Frontotemporal Dementia cohort, enriched for Caribbean Hispanic patients 2143

Board 5065F: Genetically-informed drug repurposing analysis identifies hydralazine as a therapeutic candidate for uterine fibroids 2144

Board 5066F: Genome wide assessment of homozygosity in Parkinson's disease across diverse ancestral populations 2145

Board 5067F: Genome-wide admixture and association studies of allergic rhinitis in African American population 2146

Board 5068F: Genome-wide association analysis of blood pressure traits in up to 79,402 Hispanic/Latino participants ⭐ 2147

Board 5069F: Genome-wide association study for abnormal forms of Covid-19 2149

Board 5071F: Genome-wide association study meta-analysis reveals novel loci associated with polyps of the female genital tract 2149

Board 5072F: Genome-wide association study of childhood dental caries in a multi-ethnic sample. 2150

Board 5073F: Genome-wide mapping of the *de novo* index ratio: quantifying and translating intolerance to gene dosage into neurodevelopmental effect sizes ⭐ 2152

Board 5074F: Genome-wide survival study identified variants associated with disease progression in IgA nephropathy 2153

Board 5075F: Genomic Rare Variant Mechanisms for Congenital Cardiac Laterality Defect: A Digenic Model Approach. 2154

Board 5076F: Genomic Risk Profiles for 11 Common Diseases in the electronic MEdical Records and GEnomics (eMERGE) Network Phase IV Study 2155

Board 5077F: GWAS meta-analyses of fasting glucose, fasting insulin, and HOMA-IR in Samoans 2157

Board 5078F: GWAS of overweight, not obesity, identifies new genes associated with overweight in two datasets 2158

Board 5079F: High PRS for hypercholesterolemia is equivalent to Familial Hypercholesterolemia risk in a large prospective study in primary care settings 2159

Board 5080F: Higher African ancestry proportion attenuates the increase of serum phosphorylated tau181 in a longitudinal study of aging African Americans 2160

Board 5081F: Host genome-wide association study of the Swedish gut microbiome. 2161

Board 5082F: Host genomics is associated with longitudinal infant gut microbial changes in childhood asthma and atopy 2163

Board 5083F: Identification of candidate genes associated with congenital hypopituitarism through *PROP1* phylogenetic tree and transcriptomic analysis 2164

Board 5084F: Identification of differentially expressed genes in postmortem tissue from African American Alzheimer disease cases and controls 2165

Board 5085F: Identifying functional type 1 diabetes risk variants in T cells and beta cells using a massively parallel reporter assay 2166

Board 5087F: Impact of BMI-driven proteins on nonalcoholic fatty liver disease (NAFLD) 2167

Board 5088F: Improved Diagnostic Accuracy by Trio Whole Exome Sequencing in Autism Spectrum Disorder - Experience from a single quaternary care centre in India..... 2168

Board 5089F: Improved prediction of hepatic steatosis in the UK Biobank using multimodal predictors 2169

Board 5090F: Improving Polygenic Risk Prediction of Kidney Function by removing Biomarker Specific Genetics 2170

Board 5091F: Improving the predictability of polygenic risk score 2171

Board 5092F: INHBC influences the effect of Type 2 diabetes on chronic kidney disease 2172

Board 5093F: Insights into causes and consequences of repeat expansions from biobank sequencing data 2173

Board 5094F: Insights into mosaic chromosomal alterations from whole-genome sequencing in UK Biobank 2174

Board 5095F: Insulin resistance measure TG:HDL-C ratio adjusted for BMI polygenic risk score associates with metabolic syndrome sequalae in the Michigan Genomics Initiative 2175

Board 5096F: Integrating functional genomic annotations improves prediction performances of colorectal cancer polygenic risk score across different racial and ethnic groups 2176

Board 5097F: Integrating Methylation and Transcriptomic Analyses with Whole Genome Sequencing YieldsNovel Insights into SARS-CoV-2 in African Americans..... 2177

Board 5098F: Integrating Single-Cell & Bulk Multi-Omics Data to Uncover Novel Drug Targets for Neurological and Psychiatric Disorders 2178

Board 5099F: Integration of GWAS and multi-omic QTLs identifies uncharacterized Atopic Dermatitis gene-biotype associations and pinpoint keratinocyte-linked causal candidates 2180

Board 5100F: Integration of GWAS and multi-omic QTLs identifies uncharacterized COVID-19 gene-biotype and phenotype associations 2181

Board 5101F: Integrative Dissection of Causal Variants at Four Novel Alzheimer's Disease Risk Loci 2182

Board 5102F: Interplay between genetic predisposition and plasma metabolome in relation to dementia risk and healthy aging..... 2183

Board 5104F: Investigating genetic effects on height shrinkage in the East Asian population 2184

Board 5105F: Investigating sex-specific genetic susceptibility in autoimmune diseases 2185

Board 5106F: Investigating the association between vitamin D levels and uterine fibroids using polygenic scores 2186

Board 5108F: Investigating the genetic link between bipolar disorder and gut microbial taxa through polygenic risk score analysis 2188

Board 5109F: Investigating the role of *APOE* and retinal biomarkers for predicting Alzheimer Disease 2189

Board 5110F: Investigating the Role of ETFDH and UQCRC1 Mutations in Hypertrophic Cardiomyopathy Using a Zebrafish Model ⭐ 2190

Board 5111F: Investigation of the molecular mechanism underlying IBD phenotypes through a large meta-analysis of scRNAseq data 2191

Board 5112F: Large-scale Multi-Ancestry Epigenome-wide Analysis of BMI in the Million Veteran Program 2192

Board 5113F: Late-Stage Skeletal Muscle Transcriptome in Duchenne muscular dystrophy shows a BMP4-Induced Molecular Signature 2193

Board 5114F: Leukodystrophy and Cancer in Sisters Suggests Genetic Pathogenesis: Case Series 2194

Board 5115F: Leveraging multi-omics data to explore genetic etiology for peripheral artery disease and coronary artery disease 2195

Board 5116F: Leveraging the *All of Us* Research Program to better understand the genetic architecture of opioid use disorder spectrum behaviors 2196

Board 5117F: Linkage and association of cognitive SuperAging to known protective Alzheimer's locus at *WDR12* in the Midwestern Amish 2197

Board 5119F: Linking R-loop dysregulation to anxiety-related behaviors in fragile X-associated neuropsychiatric disorders (FXAND) 2199

Board 5121F: Local genetic covariance analysis with lipid traits identifies novel loci for early-onset Alzheimer's Disease 2200

Board 5122F: Longitudinal proteomics of body mass index (BMI) in the Cameron County Hispanic Cohort (CCHC) ⭐ 2201

Board 5123F: Mapping multivariate associations among polygenic factors, brain, behaviors and psychopathologies in children 2202

Board 5124F: Meta-analysis of Genome Wide Association Studies identifies genetic loci that affect steatosis and steatohepatitis 2203

Board 5125F: Mitochondrial gene signatures across the clinical spectrum of age-related macular degeneration 2204

Board 5126F: Modeling the genomic architecture of anthropometrics and adiposity across the lifespan 2205

Board 5127F: Multi-Factor Analysis of Nicotine Metabolism, Genomics, and Epigenomics in American Indians from the Strong Heart Study 2206

Board 5128F: Myocardial Infarction Metabolomic Signatures in a Middle Eastern Cohort 2207

Board 5129F: Noncoding de novo mutations in SCN2A are associated with autism spectrum disorders 2208

Board 5130F: Non-human reads in human whole-genome sequencing identify endogenous human herpes virus (HHV)-6B and blood anellovirus virome associated with autoimmune diseases and COVID-19 risk 2209

Board 5131F: Novel Chromosome *2* locus associated with post-traumatic osteoarthritis of the knee in 47,396 cases and comparators 2210

Board 5132F: Novel genetic loci associated with knee osteoarthritis in 284,184 cases and controls in the Million Veteran Program and the All of Us Research Network 2212

Board 5133F: Novel rare variant associations with late-life cognitive performance ... 2213

Board 5134F: Objective Evaluation of Clustering Methods for Resolving the Heterogeneity of PCOS ... 2215

Board 5135F: OPRD1 encoding delta opioid receptor: a new culprit in type 2 diabetes and insulin secretion ... 2216

Board 5136F: Paternal age, *de novo* mutation, and age at onset among co-affected schizophrenia sib-pairs ... 2217

Board 5137F: Pediatric Case of Hereditary Polyneuropathy: Charcot-Marie-Tooth Disease Type 2Q and CMT1A gene mutation. ... 2218

Board 5138F: Peroxisome Biogenesis Disorder in a Patient with heyerozygous PEX6 gene mutation and 3'UTR TAAA deletion representing allelic expression imbalance ... 2218

Board 5139F: Phenotypic consequences of variation in elastic fiber genes ... 2219

Board 5141F: Polygenic Risk for Substance Abuse in the ABCD Study ... 2220

Board 5142F: Polygenic Risk Scores can predict gestational diabetes and weight gain among pregnant women in Taiwan ... 2221

Board 5143F: Polygenic risk scores for incident dementia in the Multi-Ethnic Study of Atherosclerosis ... 2222

Board 5144F: Polygenic risk scores of rheumatoid arthritis associated with seropositivity & bone erosions in a Taiwanese population ... 2223

Board 5146F: Prediction of clinical diabetes subclasses using polygenic risk scores considering enrichment of diabetes pathways ... 2224

Board 5148F: Prevalence and Impact of CYP1A1 and GST Gene Polymorphisms on Aplastic Anemia in Northern India ... 2225

Board 5150F: Quantifying extent of overlap in gene expression regulation between common and rare variants in psychiatric traits ... 2226

Board 5151F: Rare and common variants associated with alcohol consumption identify a conserved molecular network ... 2227

Board 5152F: Rare variant analysis for post- traumatic stress disorder using whole-exome-sequencing data. ... 2228

Board 5153F: Rare variant analysis of developmental stuttering identifies novel genes ... 2229

Board 5154F: Rare variation in *LMNA* underlies polycystic ovary syndrome (PCOS) pathogenesis in two independent cohorts ... 2230

Board 5155F: Recruitment biases substantially impact GWAS results and their out-of-sample portability 2231

Board 5156F: Risk of Parkinson's and Alzheimer's disease co-localises with expression of *SNCA*, *LRRK2* and *BIN1* in peripheral innate immune cells in an ancestry-dependent manner. 2232

Board 5157F: Sex-specific transcriptional differences in Gulf War Illness patients under stressful conditions. 2233

Board 5158F: Shared genetic basis of obesity, type 2 diabetes, and Alzheimer's disease: Novel insights into sex and age-specificity 2234

Board 5159F: Single cell multi-omic profiling in hiPSC-derived neurons reveals regulatory mechanisms of neuron activity dependent response and genetics of neuropsychiatric disorders ⭐ 2235

Board 5160F: Sleep inertia drives the causal association of evening chronotype with psychiatric disorders: genetic and epidemiological evidence 2237

Board 5161F: Spectrum of genetic variants in a cohort of patients with long QT interval 2238

Board 5162F: *Staphylococcus aureus* bacteremia is associated with *HLA DRB1* amino acid changes 26Y and 71K in a multi-ethnic sample 2238

Board 5163F: Structural Haplotypes in the 17q21.31 *MAPT* Region and Cognitive Resilience to Alzheimer's Disease Pathology 2239

Board 5164F: T-cell receptor repertoire sequencing at single-cell resolution in endemic pemphigus foliaceus 2241

Board 5165F: The Association of a Polygenic Risk Score for Idiopathic Pulmonary Fibrosis with Diagnosis and Mortality in Biobank Cohorts 2242

Board 5166F: The Genetic Etiology of Alzheimer's Disease Mediated by Regional Brain Gray Matter Densities 2242

Board 5167F: The role of genetics in predicting progression to severe obesity in midlife 2244

Board 5168F: The role of *Tumor necrosis factor-alpha* and *Interleukin-6* gene polymorphisms in genetic susceptibility of juvenile idiopathic arthritis in Kuwaiti Arab children 2245

Board 5169F: Trajectory analysis reveals islet cell depletion exacerbated in individuals of African ancestry with Type 2 diabetes ⭐ 2246

Board 5170F: Transcriptomic characterization of induced pluripotent stem cell-derived neural cells from patients with Williams Syndrome 2247

Board 5171F: Transferability of tobacco and alcohol use polygenic scores across populations 2248

Board 5172F: Type 2 Diabetes protective effects of rare loss of function variants in African Americans: an All of Us cohort study 2249

Board 5173F: Understanding Lecithin-cholesterol acyltransferase Deficiency 2249

Board 5174F: Understanding sleep and metabolic traits in ME/CFS 2250

Board 5175F: Unlocking Knee Osteoarthritis: Whole-Exome Sequencing Identifies Key Susceptibility Genes 2251

Board 5176F: Unveiling Mechanistic Overlaps in Opioid and Nicotine Addiction through Integrative Systems Biology: Incorporating Single-Cell RNA-Seq, Multiplex Networks, and Explainable-AI 2252

Board 5177F: Using polygenic risk score to assess the role of common genetic variants in phenotypic variability of inflammatory bowel disease in a rare immunodeficiency cohort 2253

Board 5178F: Utilizing prescription data from electronic health records to better characterize chronic pain patients and uncover novel genetic associations 2255

Board 5179F: Ventricular arrhythmia risk is predicted by a polygenic risk score for atrial fibrillation 2256

Board 5180F: Vitamin D, cognition, and Alzheimer's disease:Observational and two-sample Mendelian randomization studies. 2257

Board 5181F: Waist-to-hip ratio as a sex-specific genetic proxy for visceral adipose tissue: Apples and (pine)apples? 2258

Board 5182F: Whole genome sequencing analysis of mitochondrial genetic variation associated with cachexia in COPD 2259

Session Title: Epigenetics Poster Friday Session 2260

Board 1001F: A Cell-Type-Specific DNA Sequence Model for Understanding the Role of Noncoding Variants in Chromatin Accessibility 2260

Board 1002F: A universal spike-in normalization strategy for CUT&RUN, CUT&Tag, and ATAC-Seq 2261

Board 1003F: Allelic imbalance in chromatin accessibility in the *KCNQ1* SNP region that associates with type 2 diabetes 2262

Board 1004F: Brain epigenetic regulatory atlas reveals context-dependent trans-regulatory programs in Alzheimer's disease ⭐ 2263

Board 1005F: Characterizing the effect of flowcell and basecaller upgrades on 5mC and 5hmC methylation calling in ONT long read sequencing technologies 2264

Board 1006F: Cross species epigenomic analysis to assess long-term molecular responses to environmental chemicals: from conserved mechanisms to the One Health framework ⭐ 2265

Board 1007F: Differentially Methylated Regions in Neurexin 3 are Associated with Chronic Heavy Ethanol Use in Rhesus Macaques..... 2266

Board 1008F: Dissecting chromatin regulatory landscape during B cell activation 2267

Board 1009F: DNA methylation and genetic regulation in natural populations of East Asian and mixed Eurasian ancestry..... 2268

Board 1010F: Epigenetic signatures of age-related changes in total gene expression and splicing in human skin 2269

Board 1011F: Epigenetics for genome-wide target discovery of cardiometabolic diseases 2270

Board 1012F: ERG impacts the shear stress response in human aortic endothelial cells 2271

Board 1013F: Functional annotation of risk loci for genetically complex disease using epigenomic liquid biopsy 2273

Board 1014F: Functional interactions between LINE-1 retrotransposon signaling and the innate immune system in lung cancer 2274

Board 1015F: Genetic and epigenomic cross talk between subcutaneous adipose tissue cell-types contributes to abdominal obesity risk ⭐ 2275

Board 1016F: Genotype-epigenotype-phenotype correlations in Cornelia de Lange Syndrome: Insights from a case with a novel intragenic deletion in the *NIPBL* gene... 2276

Board 1017F: High-resolution CNS-specific spatial architectures of chromosomes in Drosophila reveal higher order organization patterns of gene regulation. 2277

Board 1018F: Hyperosmotic stress induces complete rewiring of chromatin interactions ⭐ 2278

Board 1019F: Identification of the *hsa-circ-0018905* as new biomarker for Multiple Sclerosis..... 2279

Board 1020F: Identifying Biomarkers Associated with Chronic Pain in Hidradenitis Suppurativa 2280

Board 1021F: Inferring genome organisation and gene regulation from 6-base sequencing data 2281

Board 1022F: Integrative analysis of omics summary data reveals putative causal involvement of cardiac DNA methylation in dilated cardiomyopathy 2282

Board 1023F: KMT2D-deficiency in chondrocytes induces hypoxia responses in normoxia and causes precocious differentiation 2283

Board 1024F: MeSS: a DNA methylation smoking score to quantify individual-level long-term tobacco exposure and its health impacts 2284

Board 1025F: Neuroepigenomic Profiling of Opioid Use Disorder in Human Neurons of the Orbitofrontal Cortex 2285

Board 1026F: Novel Angelman and Prader-Willi Syndrome Mouse Models with a 6 Mb Deletion homologous to 15q11.2-q13 Deletion generated by CRISPR editing 2286

Board 1027F: Novel chromatin QTL mapping approach provides insight into the cistrome regulation and complex traits 2287

Board 1028F: Novel epigenetic mediators for Barker's hypothesis: A Mendelian randomization study 2288

Board 1029F: Schizophrenia methylation profile score in a Brazilian antipsychotic-naive first-episode psychosis cohort 2290

Board 1030F: Single cell multiomic profiling of the human lung in health and disease to understand genetic risk of complex respiratory disease 2291

Board 1031F: Single-cell multi-omics analysis of transcription factor cooperativity in oligodendrocyte gene regulation 2292

Board 1032F: Sleep, Epigenetics, and Childhood Obesity: Unveiling Novel DNA Methylation Patterns 2293

Board 1033F: Structural Genomics Reveals the Effects of Rare Disease Mutations on RAG2 Epigenetic Reader Functions 2294

Board 1034F: The Multi-Ancestry Genomics, Epigenomics, and Transcriptomics of Alzheimer's (MAGENTA) Project: Understanding Alzheimer's Disease through DNA Methylation Analysis 2295

Board 1035F: Tracing Early Fetal Chromatin Signatures in Schizophrenic Brains 2296

Board 1036F: Understanding the functional consequences of rapid evolution of sperm protamines 2297

Board 1038F: Variants at adipocyte open chromatin act in *cis* and interact in *trans* on abdominal obesity risk ⭐ 2297

Board 1039F: Whole-blood DNA methylome sequencing identifies genes associated with asthma in a family cohort 2299

Session Title: Evolutionary and Population Genetics Poster Friday Session 2300

Board 3002F: 3023 human genomes from mainland Southeast Asia disclose hidden genetic diversity and signatures of tropical adaptation ⭐ 2300

Board 3003F: A preliminary study on polygenic adaption in the Han Taiwanese people 2301

Board 3004F: Admixed X-linked hemophilia A patients from the Brazilian Study of Induction of Immune Tolerance (BrazIT) study do not show the pervasive African sex-ancestry bias of Latin American populations 2302

Board 3005F: Ancient DNA provides evidence for related selection pressures at *ABO* and *FUT2* loci over the past 10,000 years in West Eurasia ⭐ 2303

Board 3008F: Characterizing and elucidating the mechanistic basis of context-dependent CpG mutation rate variation 2304

Board 3009F: Cross-comparison of genomic approaches for ancestry estimation 2305

Board 3010F: Endogamy and high prevalence of deleterious mutations in India: evidence from strong founder events ⭐ 2306

Board 3011F: Evaluating Genetic Diversity in the All of Us Research Program 2307

Board 3013F: Evaluating the evolutionary mismatch hypothesis in the electronic health record: A phenome and lab-wide investigation of known malaria variants in >35k individuals of primarily African ancestry in the Vanderbilt University biobank 2308

Board 3014F: Evaluation of imputation methods for ancient African DNA 2309

Board 3015F: Exploring structural variants and adaptation in the Andean highland and coastal populations using long-read sequencing ⭐ 2310

Board 3016F: Exploring the demographic histories of rainforest hunter-gatherers and neighboring Bantu-speaking agriculturalists as well as the genomic basis of short stature in Africa ⭐ 2312

Board 3017F: Familial long-read genome sequencing identifies early embryonic new mutations and an accelerated SNV mutation rate in high identity segmental duplications. 2313

Board 3018F: Genealogy-informed variant filtering and genotype calling on large scale human Y chromosome call sets 2314

Board 3019F: Genetic Prevalence of PEX1, PEX3, PEX5, and PEX6 Variants in Nephrolithiasis and Nephrocalcinosis within Zellweger Spectrum Disorders 2315

Board 3021F: Inference of recent population histories by combining genetic and archeological data 2316

Board 3022F: Insights into copy-number evolution and polygenic score transferability from *LPA* ⭐ 2317

Board 3023F: Integrated haplotype score IHS and signals of positive selection in suicide associated SNPs..... 2318

Board 3024F: Long-read sequencing on diverse populations identifies an association of a haplotype with two CAA interruptions in the CAG trinucleotide repeats of the ATXN2 gene 2319

Board 3025F: Neanderthal introgressed variants are associated with lung function .. 2320

Board 3026F: Non-parametric representations of structure in genetic ancestry through time 2321

Board 3027F: Overlap of adaptive genetic signatures between Andeans and Tibetans demonstrates convergent signals of natural selection 2322

Board 3028F: Paleolithic gene duplications primed adaptive evolution of human amylase locus upon agriculture 2323

Board 3029F: Phenotypic and genotypic distribution of ABO and RHD blood group antigens in Senegal. 2324

Board 3030F: Selection signatures in Pacific Islanders and associations with Native American ancestry 2325

Board 3031F: Selective dynamics of interruptions at short tandem repeats differ between coding and noncoding loci 2326

Board 3032F: SNP heritability estimation in admixed populations with Haseman-Elston regression 2327

Board 3033F: Temporal estimation of Gene-specific ω identifies genes for the emergence of novel SARS-CoV-2 strains having super infectivity and virulence 2328

Board 3034F: The battle of the sexes in humans is highly polygenic 2329

Board 3035F: The effect of long-range linkage disequilibrium on allele frequency dynamics under stabilizing selection 2329

Board 3036F: The genomic population history of the Aleutian Islands before and after European contact 2330

Board 3037F: The impact of background selection on deleterious alleles and its consequences on complex trait variation 2331

Board 3038F: Whole genome sequencing of 1,795 individuals from large pedigrees in Mongolian populations using Ultima Genomics' UG100 2332

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session 2333

Board 7052F: A study of practical application of Instrument to Assess Children's Assent (IACA): Explanation support ICT materials and other factors related to assent of children participating in genome analysis research 2333

Board 7053F: A systematic approach to community-engaged recruitment for genetics research 2334

Board 7054F: Adaptation of the Witness Project® to include genetic testing for hereditary breast and ovarian cancer risk 2335

Board 7055F: Breaking barriers: identifying challenges in accessing genetic diagnostics for at-risk Puerto Rican individuals 2336

Board 7056F: BYOB (Beat Your Own Bots): Mitigating automated responses in genetics survey research 2337

Board 7057F: Catalyzing biobank expansion: Empowering clinical staff champions to amplify biospecimen collections for a population-scale biobank 2338

Board 7058F: Clients' experiences of empathy in genetic counseling for hereditary breast and ovarian cancer: A qualitative study in Japan 2340

Board 7059F: Clinical genetic testing outcomes of 189 cancer patients: Importance of no-cost patient germline genetic testing access 2340

Board 7060F: Complex Medical Management and Ethical Considerations in Foster Care Twins with Congenital Disorder of Glycosylation Type 1A 2342

Board 7062F: Elective genome screening, clinical experience in 1300 healthy individuals 2343

Board 7063F: Exploring Uterine Fibroid Disparities among Black and Hispanic Women: Insights from the *All of Us* Workbench 2344

Board 7064F: Genetic Counseling Services for Hereditary Breast and Ovarian Cancer: Patients' Experience and Satisfaction with Different Service Models ⭐ 2345

Board 7065F: Genomic newborn screening, first clinical experience in Spain 2346

Board 7066F: How to become a genomics GENie: Presenting GenQA's online education platform that offers training and ongoing competency assessment 2347

Board 7067F: Increasing the Diversity of the Genomic Workforce : Designing Evaluation Plans to Maximize Impact 2348

Board 7068F: Limitations in access to genetic counseling training programs across the globe 2349

Board 7069F: Modernizing a comprehensive measure of genetic literacy 2350

Board 7070F: Online Support for Individuals at Risk of Genetic ALS: A Digital Platform for Education, Counseling, Testing, and Peer Support 2351

Board 7071F: Parents' responses to their children's PRS-based genomic risk results ⭐ 2352

Board 7072F: Piloting anew resiliency training program to support biomedical graduate students: The BioGREAT curriculum 2353

Board 7074F: Policymaking in the Real-World: Assessing Key Informant Perspectives on Prenatal Gene Editing 2354

Board 7075F: Polygenic-Risk Score Based Identification and Tracking of Pediatric Patients at High Risk for Type 1 Diabetes in the eMERGE Network ⭐ 2355

Board 7076F: Population Grouping Schemas related to race and ethnicity in a large, contemporary young American population: Comparing Medical Record, Caregiver Survey, and Genetic Similarity 2356

Board 7077F: Precision Oncology Workflow Modifications in a Community Cancer Center Led to Greater Adoption of Genetic Counseling Referral and Testing in Patients Diagnosed with Ovarian Cancer ⭐ 2357

Board 7078F: Public perceptions of the development and use of polygenic scores for social and behavioral traits 2359

Board 7079F: RareBoost: European Horizon 2020 Project for Rare and Undiagnosed Disease Research Capacity Building by an Interdisciplinary Approach 2360

Board 7080F: RareInsight: From diagnosis to dashboard; innovations in rare disease reporting 2361

Board 7081F: State Medicaid coverage for genetic counseling with CPT Code 96040 2362

Board 7082F: Summary and insights from the first genetic services clinic in public sector pediatric healthcare in Sindh, Pakistan. A joint collaboration 2363

Board 7083F: Technical and ethical considerations and guidelines for return of RNA-seq results for rare disease research participants 2364

Board 7084F: The Clinician-reported Genetic testing Utility InDEx for Newborn Screening (C-GUIDE NBSeq): Development and content validity 2365

Board 7085F: The feasibility of returning reclassified genetic test results through electronic patient portals 2366

Board 7086F: Women's healthcare providers' knowledge and practices surrounding fragile-X associated primary ovarian insufficiency (FXPOI) 2367

Session Title: Genetic Therapies Poster Friday Session 2368

Board 8081F: A drug repurposing screen reveals dopamine signaling as a critical pathway underlying potential therapeutics for the rare disease DPAGT1-CDG 2368

Board 8082F: Adults with acid sphingomyelinase deficiency show continued or sustained improvements in hepatosplenomegaly, respiratory outcomes, and lipid profile after 4 years of olipudase alfa enzyme replacement therapy: Final results of the ASCEND trial..... 2369

Board 8083F: Advanced Therapeutic Strategies for Fibrodysplasia Ossificans Progressiva: siRNA-Mediated Gene Silencing and sgRNA-Guided Gene Editing Using CRISPR-Cas9 2371

Board 8084F: An expedited workflow for the development of personalized corrective editing therapies for inborn errors of metabolism ⭐ 2372

Board 8086F: From bench to bedside: Lamotrigine therapy demonstrates improvements in attention, anxiety, school performance, and body mass index in a patient with MYT1L-associated neurodevelopmental disorder 2373

Board 8087F: Generation and characterization of human iPSC derived*NPC1*$^{I1061T/I10161T}$¡^{3}Neurons as a model for NPC1 disease..... 2374

Board 8088F: Identification of Small Molecule Inhibitors against *Plasmodium falciparum* Bromodomain Protein 1 using Pharmacophore Modeling, Molecular Docking, and Molecular Dynamic Simulation 2375

Board 8089F: Neuron-targeted lipid nanoparticles for miR-137-3p therapeutic delivery to Alzheimer's disease 2376

Board 8090F: Personalized antisense oligonucleotides modulate splicing to restore SDCCAG8 expression in patient-derived fibroblasts with biallelic intronic mutations 2377

Board 8091F: Preliminary results from the ongoing APHENITY extension study: Sepiapterin reduces blood Phe with improved dietary Phe tolerance in participants with phenylketonuria 2378

Board 8092F: Rescuing *KANSL1* Haploinsufficiency as a Therapeutic Strategy for Koolen-de Vries Syndrome 2379

Board 8093F: Rescuing lethality in the Homozygote *Vcp*$^{R155H/R155H}$mouse model provides a robust disease model for testing AAV gene therapy. 2380

Board 8094F: Restoring CHAMP1 expression and action in modeling studies using AAV9-mediated gene therapy 2381

Board 8096F: The natural history of NGLY1 deficiency, a rare neurodevelopmental disorder, and preliminary safety of GS-100, an AAV9 gene therapy for its treatment ⭐ 2382

Board 8097F: Treatment paradigms for studying functional & molecular characterization of the Phospholamban L39 Stop mutation (PLN L39X) & its relevance to cardiomyopathy using patient-derived pluripotent stem cell models (hiPSCs) 2383

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session 2384

Board 1042F: A lay-translation of the human phenotype ontology aids diagnostics... 2384

Board 1043F: A reference panel for linkage disequilibrium and genotype imputation using whole-genome sequencing data from 2680 LASI-DAD participants across India..... 2385

Board 1045F: All models are wrong, but which ones are useful? Characterizing broad and recurrent transcriptional profiles in neuronal cell models 2387

Board 1046F: Assessing the value of the human pangenome reference for trait mapping 2388

Board 1047F: Automated large-scale quantification of sex bias in biomedical studies 2389

Board 1048F: AVITI Sequencing for high-quality, low-cost genomic analysis at the NWGC 2390

Board 1049F: Building phenotype trajectories for rare disease: A pilot focusing on Noonan Syndrome 2391

Board 1050F: Challenges and lessons learned in developing ClinGen initial modified ACMG/AMP criteria for variant curation in X-linked inherited retinal disease genes RPGR, RS1, CHM and RP2 2392

Board 1051F: Clinical application of whole exome sequencing in the diagnosis of Pediatric neurological disorders: A single center study from Nepal 2394

Board 1052F: Comparison of variant callers using60,532 diverse genome sequences 2395

Board 1053F: Comprehensive identification and characterization of genetic variants within a pangenome reference 2395

Board 1054F: Comprehensive network modeling approaches unravel dynamic enhancer-promoter interactions across neural differentiation 2397

Board 1055F: Connecting with the Impact of Genomic Variation on Function (IGVF) Consortium 2397

Board 1056F: Creating a virtual education & thought leadership webinar program to engage the UK Biobank Research Analysis Platform community 2398

Board 1057F: Cross-population replication analysis of STR-trait associations in the All of Us dataset 2400

Board 1058F: Custom genome assembly of the reference IPSC line KOLF2.1J improves annotation and mapping for genomic data 2401

Board 1059F: Deciphering Genetic Associations in Autoimmune Thyroid Diseases for Drug Repurposing 2402

Board 1060F: Divergent genetic influences on body mass index and body fat percentage: Implications for musculoskeletal and cardiometabolic health 2402

Board 1061F: DNA Methylation levels within human genome variable number of tandem repeats (VNTRs) ⭐ 2404

Board 1062F: Empower Users with Programmatic Access to dbGaP Data Using FHIR 2405

Board 1063F: Enhancing LRRK2 Detection in Parkinson's Disease through Direct-to-Patient Genetic Testing and Community Engagement Strategies 2406

Board 1064F: Expression landscape of TE-derived transcripts in human adult tissues 2407

Board 1065F: GENE TARGET 2.0: A refined tool for evaluating monogenic neurodevelopmental disorders for therapeutic development 2407

Board 1066F: Genetic Determinants of Type 1 Diabetes: A Comprehensive GWAS and Local Biobank Initiative in Ukraine 2408

Board 1067F: Genetic diversity and structural complexity of the killer-cell immunoglobulin-like receptor gene complex: A comprehensive analysis using human pangenome assemblies 2409

Board 1068F: Haplotype-resolved characterization of repeat expansions and patterns of methylation from 1000 Genomes ONT Consortium data 2410

Board 1069F: How Open Targets resources use Genetics data to facilitate target identification and prioritisation 2411

Board 1070F: Impact and evolution of a philanthropic global network providing clinical genomic testing 2413

Board 1071F: Impact of Bias in Variance Analysis: From SNPs to Pathways 2414

Board 1072F: Impact of COVID-19 vaccination status and strain-specific infection on risk of developing dementia in the National COVID Cohort Collaborative (N3C) Data Enclave 2415

Board 1073F: Impact of Gene-Disease Validity on Variant of Unknown Significance rates in hereditary cancer panels 2416

Board 1075F: Insights from Comprehensive Fertility Biobank with Deep Phenotyping and Genomic Data in East Asian Population ⭐ 2417

Board 1076F: Introducing Human Induced Pluripotent Stem Cells (iPSCs) to the NHGRI Sample Repository for Human Genetic Research 2419

Board 1077F: Investigating methylation patterns associated with obesity in *Danio rerio* 2419

Board 1078F: Investigation of febrile seizure susceptibility and motor function in *Magel2* mutant mice as a model for Schaaf-Yang syndrome 2420

Board 1079F: Leveraging the *All of Us* platform to enable rapid genomic analysis of primary autoimmune hypothyroidism in diverse populations 2421

Board 1080F: Long-read based detection of large copy number variants with potential functional significance using the ContextSV structural variant caller 2422

Board 1081F: Mining for gold standard gene sets: maximizing the accuracy and utility of effector gene predictions from GWAS results 2423

Board 1083F: Pipeline palooza: Streamlining multi-omic discoveries across biobanks 2425

Board 1085F: Polygenic risk scores and risk of drug-induced liver injury following initiation of antiretroviral therapy in people living with HIV 2426

Board 1086F: PopCNVBrowser: a web-based Platform for CNV analysis and visualization using 1000 Genomes Project Data 2427

Board 1087F: Predicting GWAS causal variants from individually assayed regulatory effects 2428

Board 1088F: Quadrupia: Derivation of G-quadruplexes for organismal genomes across the tree of life 2429

Board 1089F: Reproducible single cell annotation of programs underlying T-cell subsets, activation states, and functions 2430

Board 1090F: Systematic Evaluation of Polygenic Score Performance Across Diverse Population Characteristics in the All of Us Research Program 2431

Board 1091F: The Helmholtz Munich imputation server: an EU based, GDPR compliant imputation server ⭐ 2432

Board 1092F: The NeuroBioBank Whole-Genome Catalog : sequencing from human brain donorswith nervous system disorders 2433

Board 1093F: The repertoire of short tandem repeats across the tree of life 2434

Board 1094F: Unraveling Cardiovascular Disorders: Integrative Analysis of GWAS-Phewas and Single-cell Data 2435

Board 1095F: Unveiling the shared genetic architecture and causality between kidney dysfunction and bladder cancer: A genome-wide cross-trait analysis 2436

Board 1096F: Upcoming version of Integrated Genome Browser will better support visualizing and exploring personal genome sequence data 2437

Board 1097F: Visual analysis of RNA-Seq data in the Integrated Genome Browser; why every researcher needs to view their data 2438

Board 1098F: WebGWAS: A web server for instant GWAS on arbitrary phenotypes ... 2438

Board 1099F: Investigation of Immune Related Gene Transcription Differences Associated with Obesity in Danio rerio (Zebrafish) 2439

Session Title: Mendelian Phenotypes Poster Friday Session 2440

Board 2002F: A comprehensive analysis of structural variations in 2,733 Orofacial Clefts (OFC) trios reveals novel genetic drivers ⭐ 2440

Board 2003F: A *de novo WDR45* gene deletion in a pediatric patient with atypical beta-propeller protein-associated neurodegeneration (BPAN) features 2442

Board 2004F: A drug repurposing screen reveals dopamine signaling as a candidate modifier pathway in *PIGA*-CDG ⭐ 2443

Board 2005F: A genotype-first approach in over 170,000 participants from a large healthcare population identifies treatable Mendelian conditions 2444

Board 2006F: A multi-omics approach improves the diagnosis and understanding of inborn errors of immunity 2445

Board 2007F: A Novel Homozygous Nonsense Variant in the LRRC56 Gene in a Neonate with Situs Inversus, Complex Heart Malformations, and Complete AV Block 2446

Board 2008F: A novel *SCN1A* variant in a patient with neonatal-onset developmental and epileptic encephalopathy with arthrogryposis multiplex congenita..... 2447

Board 2009F: A pilot project moving the initiation of exome sequence panel testing for patients with neurodevelopmental disorders into the developmental and behavioral pediatric clinic identifies both actionable diagnostic findings and a high rate of variants of uncertain significance. 2448

Board 2010F: A quick diagnosis of Alagille syndrome 1 of a Romanian patient with cholestatic liver disease..... 2449

Board 2013F: Arimoclomol upregulates expression of genes belonging to the coordinated lysosomal expression and regulation (CLEAR) network 2450

Board 2015F: Biallelic null variants in *PNPLA8* cause microcephaly by reducing the number of basal radial glia 2451

Board 2017F: Biallelic variants in consanguineous families causing neurodevelopmental disorders 2452

Board 2018F: Case report of PPOX mutation within intron-exon boundary junction inducing aberrant splicing in a patient with porphyria variegata 2453

Board 2019F: Characterization of hearing loss and skin cancer in individuals with Rothmund-Thomson syndrome 2454

Board 2020F: Characterization of pathogenic variant disease-phenotypes throughout the human genome 2455

Board 2021F: Clinical and Molecular Genetic Studies of a Filipino Family with a Peculiar Muscle Disease 2456

Board 2022F: Clinical features of HNRNP-related disorders 2458

Board 2023F: Clinical, genetic and bioinformatic analysis of Saudi families with Joubert Syndrome and Related Disorders 2458

Board 2024F: Complete nanopore repeat sequencing of SCA27B (GAA-FGF14 ataxia) in Japanese 2459

Board 2026F: Comprehensive Observational Study on the Genetic Basis of CADASIL: Insights into a Complex Neurological Disorder 2460

Board 2027F: De novo variants in GTF2H1 underlie variable syndromic developmental delay 2462

Board 2028F: Drosophila Models Uncover Substrate Channeling Effects on Phospholipids and Sphingolipids in Peroxisomal Biogenesis Disorder ⭐ 2463

Board 2029F: Elucidating the functional relationship between ATAXIN1 and ATAXIN1-Like in the brain 2464

Board 2030F: Establishing prevalence and phenotypic spectrum of purine and pyrimidine disorders using genotype-first approach to mining electronic health record biobanks 2465

Board 2031F: Establishment the transgenic pig model and exploration the molecular mechanisms of rare disease white sponge nevus 2466

Board 2032F: Estimating the disease burden due to genetic variants associated with hypocalcemia 2467

Board 2033F: Exploring the genetic architecture of orofacial clefts in Filipino trios identifies GRHL2 as a novel gene associated with clefting 2468

Board 2034F: Familial TBX5 duplication associated with Currarino-Silverman syndrome 2469

Board 2035F: Female carriers of X-linked myotubular myopathy (XLMTM) in Germany: Clinical characterization of a cohort extends the phenotypic spectrum and knowledge about the impact of heterozygous variants in the *MTM1* gene 2470

Board 2036F: *FGF14*, *ZFHX3*, *NOTCH2NLC* gene repeat variant found in clinically diagnosed multiple system atrophy. 2471

Board 2037F: Generating pituitary cells from iPSCs for pituitary disease modeling and drug discovery 2472

Board 2038F: Genetic Analysis of Patients with Intellectual Disability and Congenital Anomalies Presenting with Hypomelanosis of Ito 2474

Board 2040F: Genetic etiologies and diagnostic yield of exome sequencing in pediatric motor speech disorders 2474

Board 2041F: Genetic Investigation of Consanguineous Pakistani Families Segregating Rare Neurological Disorders 2476

Board 2042F: Genome Sequencing of 400 rare disease patients through the Korean Regional Rare Disease Diagnostic Support Program 2476

Board 2043F: Genome-wide association study reveals novel loci associated with acute chest syndrome in sickle cell disease patients 2478

Board 2044F: Homozygous variants of *ELOVL1* presenting with severe liver diseases 2479

Board 2045F: Identification of a novel splicing variant in *LEPR* associated with severe early onset obesity 2480

Board 2046F: Identification of de Novo and novel mutations in LTBP2 2481

Board 2047F: Identification of novel pathogenic variants in the neurodevelopmental gene *LNPK* 2481

Board 2048F: Interpretation of neurological disorders cohort through in-house developed tool, FestiVAR, showed an improved diagnostic yield: data from over 1000 Indian patients 2482

Board 2049F: Investigating Missing Heritability in Inherited Retinal Dystrophy 2483

Board 2050F: Investigating the genetic and phenotypic landscape of ectodermal dysplasia 2485

Board 2051F: Investigating the genetic etiologies underlying septo-optic dysplasia (SOD) 2486

Board 2052F: Leveraging large scale biobank data to enrich phenotypic understanding of Mendelian disorders 2487

Board 2053F: Low concordance of variant effect predictors to determine the pathogenicity of genetic variants in core immunity genes 2488

Board 2054F: Management of inborn errors of metabolism in Ecuador. 2489

Board 2055F: Mental health conditions, physical functioning, and health-related quality of life in adults with OI: a cross-sectional study 2490

Board 2057F: Novel compound heterozygote variants in *AARS1* cause developmental and epileptic encephalopathy in a Malian family 2491

Board 2058F: Novel genetic findings in Malian families with tuberous sclerosis complex. 2492

Board 2059F: Novel germline mutation in *BCORL1* gene identified in a family with three Shukla-Vernon syndrome patients 2493

Board 2060F: Novel pathogenic variant of *TBCK* reveals broadened phenotypic spectrum of rare TBCK syndrome 2494

Board 2061F: Novel Strategies and Analytical Methods in Tackling Unresolved Rare Diseases: The Approach of the BCM-GREGoR Program ⭐ 2495

Board 2062F: Rare coding variants in epilepsy: A study of 9,205 individuals using NGS dataset. ⭐ 2496

Board 2063F: SATB2 Variants in Glass Syndrome: Discussing the Clinical Impact of Novel Variants in the context of Misleading Diagnoses. 2497

Board 2065F: Severe hemolytic anemia, seizures, and respiratory failure in a patient with triosephosphate isomerase deficiency: a case report 2498

Board 2066F: Siblings with Okur-Chung syndrome (OCS) due to a pathogenic mutation of *CSNK2A1* gene on chromosome 20p13: p.(R21*), and a 51 Kb duplication at chromosome Xp11.4 with variable expression. 2499

Board 2067F: Tandem repeats in genes associated with synaptic functions are frequently expanded in individuals with suspected genetic epilepsies ⭐ 2500

Board 2068F: Targeted panel sequencing: Advancing molecular diagnosis for patients with vascular anomalies 2501

Board 2070F: The role of multidisciplinary care for a family without a clear cardiac phenotype identified to have an *SCN5A* variant on exome sequencing 2503

Board 2071F: The unusual pattern of hypomethylation at IC1 and IC2 in Beckwith-Wiedemann syndrome and Russell-Silver syndrome 2504

Board 2072F: *UBE2I* mutations unveil the critical role of SUMOylation pathway in syndromic autism 2505

Board 2073F: Unlocking the genetic code of cerebral palsy: Exploring clinical and genetic insights in modern era 2506

Board 2074F: Unraveling the genetic mimics of cerebral palsy: Treatable disorders should not be missed 2507

Board 2075F: Unraveling the Genomic Landscape of Anti-Epileptic Drug Resistance: A Bioinformatics approach ⭐ 2508

Board 2076F: Using Whole Genome Sequencing to improve diagnostic and care in Critically Ill pediatric patients in Africa: A Preliminary Study from Kinshasa. 2510

Board 2077F: Variable Expressivity of Rare Familial 14q13.3 Deletions That Include NKX2-1, PAX9, and SLC25A21 2511

Board 2078F: Whole exome sequencing and metabolomics screening of a Kobberling Familial Partial Lipodystrophy cohort identifies new candidate genes and dysregulated pathways ⭐ 2512

Board 2079F: Whole genome analysis in consanguineous families reveals new loci for speech sound disorder (SSD) 2513

Board 2080F: Xanthine Stones and Molybdenum Cofactor Deficiency: A Genetic Exploration of MOCS1, MOCS2, and GPHN 2514

Board 2081F: Zhu-Tokita-Takenouchi-Kim Syndrome (ZTTK): a Novel Neurodevelopmental Disorder caused by pathogenic variants in *SON* 2515

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session 2516

Board 3041F: A 15-Year Review of Cytogenetic Studies and Analyses Done at the National Institutes of Health, University of the Philippines Manila from 2008-2022 ⭐ 2516

Board 3042F: Advancing mitochondrial DNA analysis: Utility of PacBio long-read sequencing technology in accurate and precise variant detection ⭐ 2517

Board 3043F: An exploratory prospective clinical trial of 120 neonates in BeginNGS, an artificial intelligence-enabled newborn genome sequencing system of 412 childhood genetic disorders. ⭐ 2518

Board 3044F: An interstitial duplication of 2q11.2q13 in a neonate with multiple congenital anomalies: A case report 2520

Board 3046F: Clinical relevance of mosaic variants detected from exome sequencing data in a cohort enriched in primary immune disorders. 2521

Board 3047F: Clinical significance of imprinting center 1 (IC1) microdeletions in Beckwith-Wiedemann syndrome 2522

Board 3048F: Clinical validation of RNA sequencing for Mendelian disorder diagnostics ⭐ 2523

Board 3049F: Comparison of prenatal cell-free DNA screen positive for monosomy X with diagnostic maternal and fetal cytogenetic studies: a single-center retrospective study ⭐ 2524

Board 3050F: Copy number Variants in Clinical Whole Exome Sequencing of First 1000 cases 2525

Board 3051F: Cytogenetics in the era of genomic medicine: a retrospective study of 750 patients 2526

Board 3052F: De novo variant calling from duo long-read sequencing: improving equitable variant interpretation for diverse family structures 2527

Board 3053F: Downgrading Partial Duplications of Haploinsufficient Genes from VUS to Likely Benign 2528

Board 3054F: Elucidating the diagnostic yield and allelic characteristics of FGF14 repeat expansion for adult ataxia through whole genome sequencing 2529

Board 3055F: Evaluating Dosage Sensitivity Predictions for Multigenic Copy Number Variants to Facilitate Clinical Interpretation 2531

Board 3056F: Genetic characterization and diagnostic yield of next-generation sequencing in Colombian patients with suspected skeletal dysplasias 2532

Board 3057F: Genetic insights into spina bifida: a study of 205 families 2533

Board 3058F: Hidden yet dangerous: searching for occult Y chromosome in patients with Turner syndrome 2534

Board 3059F: Integrating algorithms to optimize clinical identification of structural variation using whole-genome sequencing 2535

Board 3060F: Karyotype findings in recurrent pregnancy loss: a path to better outcomes 2536

Board 3062F: Long-read whole genome sequencing increases diagnostic yield for short-read negative pediatric sensorineural hearing loss patients ⭐ 2537

Board 3063F: Microduplication in chromosomal region 2p16.1p15: Further characterization of the phenotype 2539

Board 3064F: Molecular diagnostic findings of 166 patients with Pachyonychia Congenita ⭐ 2540

Board 3065F: Multicolor metaphase FISH analyses using selected BAC clones after CNV detection in a patient revealed that two types of chromosomal abnormalities involving chromosomes 16: a balanced translocation [t(16q;20q)pat], and a submicroscopic unbalanced translocation [der(22)t(16p;22q)dn]. 2541

Board 3066F: Multiple bioinformatic tool strategy for implementation of non-coding variant analysis in the genome diagnostics 2542

Board 3067F: PCH Gene Kids: Comprehensive WGS-based healthcare for the Intermountain West 2543

Board 3068F: Research genomics strategies to end the diagnostic odyssey for challenging rare disease patients ⭐ 2544

Board 3069F: Searching for answers: Continued utility of dynamic re-analyses in rare disease diagnostics 2546

Board 3071F: STRchive: a dynamic resource detailing population-level and locus-specific insights at tandem repeat disease loci 2547

Board 3072F: Structural Variation Detection from Short-Read Whole Genome Sequencing Data: An Approach that Increases Diagnostic Yield for Rare Disease 2548

Board 3073F: The fourth Korean patient to receive a gene therapy for *RPE65*-related LCA caused by a paternal UPiD 2549

Board 3074F: The Impact of Trisomy 21 on Dental Development - A Focus on Tooth Agenesis 2550

Board 3076F: Unraveling undiagnosed rare disease families by HiFi long-read genome sequencing 2551

Board 3077F: Utility of liquid biopsy in cutaneous mosaicims and hemihypertophy .. 2552

Board 3078F: X chromosome CNVs: Candidate gene search by reanalysis of 18,666 chromosomal microarrays at the Greenwood Genetic Center ⭐ 2553

Session Title: Molecular Effects of Genetic Variation Poster Friday Session 2555

Board 6002F: A high-throughput screen to enable identification of pleiotropic transcriptional programs in asthma 2555

Board 6003F: A humanized yeast model to test neuropathy-associated *AARS1* variants for dominant-negative effects and activation of the integrated stress response 2556

Board 6006F: A transcriptome-wide association study of amyotrophic lateral sclerosis 2557

Board 6007F: A unified framework for cell-type-specific eQTLs prioritization by integrating bulk and scRNA-seq data..... 2557

Board 6008F: Assessment of Longitudinal Trajectory Proteomic QTLs in the Multi-Ethnic Study for Atherosclerosis (MESA) 2558

Board 6009F: Benchmarking and characterizing *FMR1* mosaicism in the human brain 2560

Board 6010F: Biomarker discovery for Hypophosphatasia 2560

Board 6011F: Borzoi-guided fine mapping improves variant and gene prioritization in GWAS 2561

Board 6013F: Cell type-specific effects of common germline variation on gene expression and pancreatic cancer risk 2562

Board 6014F: Cell-type specific and temporally dynamic genetic regulatory effects in response to environmental perturbations 2564

Board 6015F: Characterization of inter-individual and cell-type-specific responses to treatment with *ALK* inhibitors 2565

Board 6016F: Comparative, cross-population analysis of the genetic regulation of plasma proteins in the Million Veteran Program ⭐ 2565

Board 6017F: Context-dependent proportional amplification of genetic and environmental effects on gene expression 2567

Board 6018F: Coronary artery disease SNPs regulate transcript diversity through splicing quantitative trait loci in endothelial cells 2568

Board 6019F: De novo assembled genomes from Middle Eastern trios unveil high-resolution diversity and effect of inbreeding 2569

Board 6020F: Deep Cellular Phenotyping of genetic variants in Hunter Syndrome 2570

Board 6021F: Defining tissue- and isoform-specific function of *OCRL* using patient-derived tissues 2571

Board 6022F: Diagnosis of *CHKB* associated muscular dystrophy, congenital, megaconial type after variant phasing and RT-PCR 2572

Board 6023F: Discovering disease variants and mechanistic networks for small-cohort GWAS with geometric deep learning ⭐ 2573

Board 6024F: *Drosophila* model of *de novo PHACTR1* variant demonstrates alterations to tubulin rich structures and synaptic function. 2574

Board 6025F: Dysregulated GluA2-Y876 PhosphorylationContributes toLossof Synaptic Upscaling,Impaired Sociability and Increased Repetitive Behaviors in Grip1 Mutant Mice 2575

Board 6026F: Enhanced mapping of gene environment interactions for vitamin D through variability quantitative trait loci 2576

Board 6027F: eQTL Discovery and Analysis in Individual Pancreatic Cell Types from the Human Pancreas Analysis Program ⭐ 2577

Board 6028F: E-Selectin as a Possible Biomarker for Vaso-Occlusive Crisis in Sickle Cell Disease 2578

Board 6029F: First Exon Splice Donor Variants: a Meta-Analysis 2579

Board 6030F: Folate as a Potential Treatment for Lethal Ventricular Arrhythmias in TANGO2-Deficiency Disorder 2580

Board 6031F: Founder Effect Maintains Rare Genetic Disease Variants in Pakistani Punjabi Population 2581

Board 6032F: Functional effects of*BRCA1*and*BRCA2*mutations in iPSC derived breast and ovarian cancer organoids 2582

Board 6033F: Genetic association of *APOL1 and GSTM1* with diabetic kidney disease (DKD) in Ghanaian and Nigerian populations ⭐ 2583

Board 6034F: Genetic Predictors that Modify the Association Between Hippocampal Volume and Memory Performance 2584

Board 6035F: Genetic Regulation of Vitamin B12 Concentrations in Plasma and Human Milk Across 4 Study Sites: Data from the MILQ Study 2585

Board 6036F: Genetically regulated, cell-type-specific transcriptional responses to anticancer drugs in heterogeneously differentiating cultures of cardiac cells 2587

Board 6037F: Genome-Wide Association Study of Individual Vitamin D Metabolite Levels and Ratios Replicates Known Variants on Chromosome 6 and Identifies Novel Variants in the Diabetes Autoimmunity Study in the Young (DAISY) 2588

Board 6038F: Genome-wide CRISPR screen identifies genes that affect hepatic steatosis and implicate new targets for MASLD 2589

Board 6039F: Genome-wide glomerular allele-specific expression in human proteinuric kidney tissue 2590

Board 6040F: Genome-wide quantitative trait locus mapping of serum proteins using whole-genome sequencing in a Japanese population 2591

Board 6041F: Haploinsufficiency versus Gain of Function in AHDC1 and the Xia-Gibbs Syndrome 2592

Board 6042F: Harnessing virtual Capture C to identify GWAS target genes for the African ancestry-specific variant at *GRIN3B* in Alzheimer's disease 2593

Board 6043F: High-throughput genotype-phenotype association data aid variant classification in a diverse cohort 2594

Board 6044F: Identification of variants associated with kidney dysfunction in individuals of African ancestry with sickle cell disease 2595

Board 6045F: Integrated computational and experimental biophysics reveal shared mechanisms for genetic variants across cancer and RASopathies 2596

Board 6046F: Investigating the functional relationships between mitochondrial DNA variation and nitric oxide regulation through multi-omics integration 2598

Board 6047F: Lentivirus-based MPRA allows fine-mapping of transcription factor binding sites within the promoter of a cardiomyopathy-associated gene. ⭐ 2599

Board 6048F: Leveraging bulk to interpret adipose single nucleus RNA-seq data: adipocyte subtypes reveal adipocyte genes for insulin sensitivity 2600

Board 6049F: Leveraging genome-wide effects on gene expression to identify new disease-critical genes with *trans*-genetic components 2601

Board 6050F: Leveraging long-read Oxford Nanopore Technologies and a national genome program to better understand structural variation inheritance and medical impact 2602

Board 6051F: Leveraging population cohorts to identity new pathogenic variants 2603

Board 6052F: Linking regulatory variants to target genes by integrating single-cell multiome methods and genomic distance 2604

Board 6053F: Monocyte Inflammation and Resilience to Alzheimer's disease: Novel Genetic Risk Genes Identified by Transcriptome-wide Association Study 2606

Board 6054F: Natural variants in SEL1L modify lethality, ERAD, and proteasome function in a model of NGLY1 deficiency 2607

Board 6055F: New multi-omics xQTL colocalization method improves the discovery of causal variants for Alzheimer's disease 2608

Board 6056F: Non-coding variants in *SCN1A* reveal novel poison exons and therapeutic targets 2609

Board 6057F: *NSD2* New variant leading to Global Developmental Delay and Autism 2610

Board 6058F: pcQTLs: Identifying variants impacting the expression of multiple genes 2610

Board 6059F: Pervasive gene overlap across pathways leads to false positives in GWAS pathway enrichment testing 2611

Board 6060F: Polygenic disease association of individual cells in single-cell ATAC-seq data 2612

Board 6061F: Population level-diversity of human IGHC haplotypes is linked to inter-individual variation in antibody isotype levels in the expressed repertoire 2613

Board 6062F: Population-scale Single-nucleus Cell-type-specific Chromatin Dynamics and Nucleosome Positioning in Skeletal Muscle ⭐ 2614

Board 6063F: pQTL discovery using the SomaScan®high-plex proteomics platform in the COPDGene study 2615

Board 6064F: Proteome-wide association study of blood mitochondrial DNA copy number 2617

Board 6065F: Rare and Common Variants Risk of Thoracic Aortic Aneurysm and Dissection 2618

Board 6066F: Reducing Variant Classification Disparities Using Multiplexed Functional Data 2619

Board 6067F: Regulatory variation in MTOR dampens Neutrophil-T cell crosstalk during sepsis 2620

Board 6068F: Responsible genes for triplet repeat diseases are involved in the ubiquitin proteasome system 2621

Board 6069F: RNA-based analysis of potential splicing variants facilitates genome diagnostics and reveals limitations of in silico prediction tools 2622

Board 6070F: Role of Trisomy 21 in Hirschsprung Disease 2623

Board 6071F: Scalable single-cell models for robust cell-state-dependent eQTL mapping 2624

Board 6072F: Single cell RNA sequencing-based modeling of the mismatch repair pathway to support clinical variant classification for Lynch Syndrome ⭐ 2625

Board 6073F: Single-cell long-read sequencing of normal lung tissues from Asian women characterizes transcript heterogeneities and facilitates splicing-QTLs 2626

Board 6074F: Single-neuron somatic mutation in the context of Alzheimer's disease tau pathology 2627

Board 6075F: Somatic Variants in Focal Cortical Dysplasia Create Disease-Associated Transcriptional Programs without Altering Cell Identities 2628

Board 6076F: Systematic cell-specific assessment of childhood obesity GWAS loci using massively parallel reporter assay 2630

Board 6077F: TET2 clonal haematopoiesis is associated with improved antibody responses to SARS-CoV-2 vaccination and potentiates a vaccine-response endotype 2631

Board 6078F: The allelic spectrum of *SCN4A*-associated disease in over 7000 individuals: lessons learned from genetic testing. 2632

Board 6079F: The effect of rare gonadotropin pathway variants on PCOS pathogenesis. 2633

Board 6080F: The human host genetic determinants of immune function across ancestry and disease 2634

Board 6081F: The influence of trinucleotide repeats in the androgen receptor gene on androgen-related traits and diseases 2635

Board 6082F: The Trisomy 21 Model Atlas 2636

Board 6083F: Trans-pQTL network highlights the role of the NLRC4 inflammasome in regulating circulating IL-18 level 2637

Board 6084F: Triplication of HSA21-encoded genes modulates Alzheimer's Disease-related phenotypes in vivo via interferon signaling 2638

Board 6085F: *TSC2* loss in neural progenitor cells suppresses translation of ASD/NDD-associated transcripts in an mTORC1- and MNK1/2-reversible fashion 2639

Board 6086F: Uncovering the Genetic Basis of Systemic Lupus Erythematosus in Diverse Populations 2640

Board 6087F: Uncovering the Role of an African-specific *ABCA7* Frameshift Deletion on Lipid Metabolism and Alzheimer’s Disease 2641

Board 6088F: Unravelling the effect of multiple sclerosis risk loci on splicing using targeted single-cell long-read sequencing 2642

Board 6089F: Using AllofUs to investigate genetic factors influencing the measles vaccine response 2643

Board 6090F: Using DOLPHIN protein domain information to classify human variations in rare diseases 2644

Board 6091F: Utilizing cortical organoids to understand cellular mechanisms in patients with the developmental delay-associated 16p12.1 deletion 2645

Board 6092F: Variability in Self-reported Race Among Individuals with Middle Eastern Inferred Genetic Ancestry in the UK Biobank, All of US and MGB Biobank 2646

Board 6093F: Versatility of variant libraries and deep mutational scanning in diagnostics and therapeutics for transmembrane proteins 2647

Session Title: Omics Technologies Poster Friday Session 2648

Board 1100F: "CellmarkerAI", an artificial intelligence tool for Single-cell transcriptomics gene marker identification and annotation 2648

Board 1101F: 2022 Trans-Antarctic Expedition: Microbiome and multiomic health tracking during a 2,150km unsupported kite-ski expedition across Antarctica, with robust control and twin pair participant comparison ⭐ 2650

Board 1102F: 39 unique proteins have differential responses associated with ancestry during severe COVID-19 infection 2651

Board 1103F: A comprehensive comparison on clustering methods for multi-slices spatially resolved transcriptomics data 2652

Board 1104F: A computational approach to capturing infant gut microbiome and metabolic changes with diet..... 2653

Board 1105F: A heterogeneous differentiating cell culture system for studying gene regulatory dynamics in endoderm-derived cell types..... 2654

Board 1106F: A Novel Bioinformatic Tool for Large-Scale Primer Design and Specificity Analysis 2655

Board 1107F: A novel NGS-based sample tracking system for robust sample chain of custody and quality control in population-scale genomics studies..... 2656

Board 1108F: A web-based application for integrative and comparative analysis of multi-omic data..... 2657

Board 1109F: Advancing ACMG/AMP variant classification through computational facial analysis for variants of unknown significance 2658

Board 1110F: CasNovo: De-novo Cas-9 protein engineering with transformer large language models 2659

Board 1111F: Characterization of the rat osteosarcoma UMR-106 using long read methods identifies shared patterns of amplification, mutation, methylation and expression with human disease 2660

Board 1112F: Classification of continental ancestry in the Million Veteran Program from high-throughput proteomics data 2661

Board 1113F: Cohort-based Integrated Multiomics from Single Sample Blood Draws Provides Comprehensive Holistic Phenotypic Insights 2662

Board 1114F: Comprehensive analysis of Cas9 off-target effects using Ultima Genomics deep whole genome sequencing 2663

Board 1115F: CRISPR-BE Screen Design Server: a free web-based service for designing accurate sgRNA libraries for CRISPR base editing screens. 2664

Board 1116F: Cross platform comparison of HLA typing from whole-genome sequencing 2665

Board 1117F: Cross-Ancestry Comparison of Aptamer and Antibody Proteomics Measures 2666

Board 1118F: Cross-Phenotype Genome-wide Association Study of Brain Morphology Features Extracted from Brain MRI Images 2667

Board 1119F: CUT&RUN-EM: An ultra-sensitive, multiomic method that directly links chromatin features to DNA methylation ⭐ 2669

Board 1120F: Defining a Clinical Data Ecosystem for Genomic Health 2669

Board 1121F: Development of a pipeline for RNA editing analysis in single cell RNA-Seq data 2670

Board 1122F: Digital Pathology Image Analysis Enhances 10-Year Risk Stratification for Early Stage ER+ Breast Cancer 2671

Board 1123F: Direct detection and counting of trinucleotide repeat expansion using CRISPR-Cas9 and PacBio HiFi Sequencing 2672

Board 1124F: DNA-Diffusion: Leveraging Generative Models for Controlling Chromatin Accessibility and Gene Expression via Synthetic Regulatory Elements ⭐ 2674

Board 1125F: DNAfusionDetect: A Clinical Integrative Pipeline for Detection and Annotation of Gene Fusions from DNA Sequencing Data 2675

Board 1126F: Enabling workflows for determining mechanisms of action in drug discovery with flexFS 2676

Board 1127F: Ethidium bromide treated cell lines reveal nuclear gene expression responses to changes in mitochondrial DNA copy number 2677

Board 1128F: Evaluating the dynamics of germline mutation at repetitive loci with AVITI sequencing and multi-generational Utah/CEPH pedigrees 2678

Board 1129F: Exposomic, proteomic, and metabolomic analyses reveal environmental toxicant exposures in the Rio Grande Valley 2679

Board 1130F: Flexible, Production-Scale, Human Whole Genome Sequencing on a Bench-top Sequencer 2680

Board 1131F: Formaldehyde Inhalation Toxicity: A Comparative Study of Acute and Chronic Exposure Using Transcriptomic Data 2681

Board 1132F: Functional validation of sickle cell disease alloimmunization susceptibility locus in the spleen 2682

Board 1133F: Harnessing the power of 3D genomics for personalised medicine and drug discovery 2683

Board 1134F: Identification of immunogenetic parameters associated with cancer health disparities through TCGA data analysis 2684

Board 1135F: Identifying Key Regulators of Alzheimer's Disease Risk Using Single-Nucleus RNA-Sequencing Analysis 2685

Board 1136F: igh-Coverage Spatial Genomics with Barcoded Multiple Displacment Amplification 2686

Board 1137F: Impact of IL-1 and IL-6 gene polymorphisms on the severity of COVID-19 2687

Board 1138F: Improving Transcriptome Variant Detection Using PacBio Isoform Sequencing 2688

Board 1140F: InterSpatial: Bi-directional Integration of Single Cell RNA-seq and Spatial Transcriptomic Data 2689

Board 1141F: Is there anything in the swamp? A credibility assessment of biologic signals from non-secreted low-abundance proteins measured in human plasma 2689

Board 1142F: Leveraging machine learning to predict Parkinson's disease development using proteomics data from presymptomatic patients ⭐ 2691

Board 1143F: Long-read sequencing to diagnose Mendelian diseases ⭐ 2692

Board 1144F: Managing multi platform long read sequencing data at scale 2693

Board 1145F: Metabolomic Analysis of a Rare Disease Cohort Identifies Disease-Specific Diagnostic Panels 2694

Board 1146F: mitoScope: A New efficient de novo workflow for processing long Reads for mitochondrial analysis ⭐ 2695

Board 1147F: Multi-Omics Data from Short-Duration Space Travelers 2696

Board 1148F: Multiplexed full-length single-molecule sequencing of the human mitochondrial genome and its application for rare neuromuscular diseases and chronic lymphocytic leukaemia ⭐ 2697

Board 1149F: Net2Onto: Leveraging molecular networks and ontology structure for predicting novel genes that play a role in understudied functions, traits, and diseases 2698

Board 1150F: New Modular Hands-On Training Resources for Nextflow 2699

Board 1152F: Optimal resolution of differential expression of long read RNA-seq using transcript trees and data thinning 2700

Board 1153F: Optimizing recall of RNA-seq outliers in heterogeneous rare disease cohorts 2700

Board 1154F: Parent-of-origin-aware genomic analysis of hereditary cancer patients for parent-of-origin assignment to pathogenic variants without parental data 2701

Board 1155F: Phasing Accuracy of Long-Read Whole Genome Sequencing is Coverage and Reference Genome Dependent 2703

Board 1156F: Phasing of RNA isoforms reveals haplotype-specific expression 2704

Board 1157F: Polygenic Risk Scores as Knockoffs for Deep Learning Analysis of Exposome Data: Unveiling Nonlinear Disease Risk Factors 2705

Board 1158F: Predicting Loss- and Gain-of-Function in Ion Channel Variants Using Machine Learning 2706

Board 1159F: Proteomic Analysis Identifies Novel Protein Quantitative Trait Loci and Causal Proteins in Type 2 Diabetes Among Individuals of African Ancestry ⭐ 2708

Board 1160F: QuadST: A Robust and Powerful Approach for Identifying Cell–Cell Interaction-Changed Genes on Spatially Resolved Transcriptomics 2709

Board 1161F: Rare variant associations with the plasma proteome in individuals of South Asian ancestry..... 2709

Board 1162F: Rare-variant drop-out in biobank-scale multi-sample genotyping of next-generation sequencing data 2710

Board 1163F: RISE: Relative Impact of Splicing and Expression in RNA-seq studies .. 2711

Board 1164F: ROQ: a new measure for more accurately filtering mapped reads 2712

Board 1165F: Sawfish: Improving long-read structural variant discovery and genotyping with local haplotype modeling 2713

Board 1166F: scExploreR: a flexible platform for democratized analysis of multimodal single-cell data..... 2714

Board 1167F: scraps: an end-to-end pipeline for measuring alternative polyadenylation at high resolution using single-cell RNA-seq 2715

Board 1168F: Single Cell RNA-seq Cell Annotation with Generative Pre-trained Transformers (GPT) Models 2716

Board 1169F: Single-cell spatial RNA-sequencing on viral infections in human cerebral organoids reveal human-viral transcript interactions in Alzheimer's disease 2717

Board 1170F: Single-cell Total-RNA Profiling Unveils Regulatory Hubs of Transcription Factors ⭐ 2718

Board 1171F: Spatial Characterization of Isoform Landscapes in the Human Brain... 2719

Board 1172F: Speeding up Genetic Analysis with a Smaller Data Format 2720

Board 1173F: Streamlined, multimodal, epigenetic measurements using CUT&RUN combined with nanopore sequencing 2720

Board 1174F: Temporal and spatial dysregulation of immune and metabolic pathways of prefrontal cortex due to COVID-19 2721

Board 1175F: The Long Road to Long Reads: Challenges to Implementing a Long Read Sequencing Service 2722

Board 1176F: Uncovering the impact of splicing variants in intractable pediatric epilepsies ⭐ 2723

Board 1177F: Understanding the role of rare untranslated region (UTR) and long non-coding RNA variants on human plasma protein levels using ~50,000 UK Biobank whole-genome sequences 2725

Board 1178F: Unmasking HMC3 Cell Lineage: Insights from Comparative RNA Sequencing Analysis..... 2726

Board 1179F: Unraveling the Testicular Seminoma Tumor Microenvironment by Single Cell RNA-sequencing..... 2727

Board 1180F: Unveiling novel insights into the pathophysiology of TANGO2 Deficiency Disorder using integrated functional and multiomics studies 2728

Board 1181F: Unveiling the complex landscape of 16p11.2 deletion: a multifaceted examination of phenotypes, genotypes, and transcriptomic patterns 2729

Board 1182F: Unveiling the Hidden Rules: Enhancing NMD Prediction for Protein-Truncating Variants 2730

Board 1183F: Using transcriptomics to identify RNA biomarkers of autophagy status in primary human mononuclear cells 2731

Board 1184F: Whole genome sequencing in clinical psychiatry: *RFX3* complex rearrangement identified by long-read sequencing in a hospitalized adolescent with autism, intellectual disability and behavioral decompensation 2732

Board 1185F: Detecting Circadian Rhythmicity by Single-Cell Multiomics..... 2733

Session Title: Pharmacogenomics Poster Friday Session 2734

Board 7001F: Analysis and diagnostic interpretation of therapeutic factor VIII plasmid in hemophilia A using logistic and exponential regressions 2734

Board 7002F: Analysis of genetic variation among pharmacogenes in Sardinian population 2735

Board 7003F: Association of *SULT2A1* with abiraterone clearance in the Alliance A031201: Randomized phase III study of enzalutamide compared with enzalutamide plus abiraterone for metastatic castration-resistant prostate cancer 2736

Board 7004F: Deep mutational scan of the pharmacogene *CYP2D6* in *Saccharomyces cerevisiae* ⭐ 2737

Board 7005F: Distinct HLA Class I Allele Clusters with Shared Peptide-binding Specificities Predispose to Co-trimoxazole-induced Severe Cutaneous Adverse Reactions ⭐ 2738

Board 7006F: Exploring allelic variation at structurally variant and medically relevant loci in the *All of Us* Research Program long-read genomes ⭐ 2739

Board 7008F: *FTH1* gene knockout induce immunomodulatory drug resistance in multiple myeloma cells through disruption of endoplasmic reticulum stress and calcium signaling pathways. 2741

Board 7009F: Genetic predictors of statin effectiveness, adherence to treatment, and adverse clinical outcomes: longitudinal analysis of UK Biobank electronic health records 2741

Board 7010F: Genetic susceptibility of *HLA* alleles to beta-lactam antibiotics allergy in the Taiwanese population 2743

Board 7011F: Genetically proxied therapeutic inhibition of kidney function drug targets and type 2 diabetes in Africans: A Mendelian randomisation study ⭐ 2744

Board 7012F: Genetics of glucocorticoid-induced hyperglycemia 2745

Board 7013F: Germline homologous recombination deficiency influences the fitness impact of oncologic therapy on TP53 clonal hematopoiesis 2746

Board 7014F: Incorporating transcriptional signatures to improve genetic prioritization of drug indications. 2747

Board 7015F: Integrating WES and array data for improved pharmacogenetic typing: A comparative study with WGS 2748

Board 7016F: Polygenic prediction of immune checkpoint inhibitor-mediated hepatotoxicity in non-small cell lung cancer patients ⭐ 2749

Board 7017F: Polygenic risk of depression is correlated with treatment resistance, treatment nonresponse and antidepressant class specific response across ancestrally diverse populations 2750

Board 7019F: Rapid and scalable personalized therapeutic screening in patient-derived organoid models of genetic disease ⭐ 2751

Board 7020F: Reproduction of *HLA-A*32:01* and discovery of two loci near *IL-7* and *MAP2K6* associated with vancomycin drug reaction with eosinophilia and systemic symptoms in an electronic health record-based study 2752

Board 7021F: StarPhase: Comprehensive Phase-Aware Pharmacogenomic Diplotyper for Long-Read Sequencing Data 2753

Board 7022F: Timely Integration of Pharmacogenomics Information into the Electronic Health Records to Deliver a Clinical Decision Support Tool for Clinicians- an Agile Solution at the Veterans Health Administration Medical Centers 2754

Board 7023F: Using whole genome sequencing to call star alleles expands coverage and sensitivity for clinical pharmacogenomics decision-making 2755

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session . 2757

Board 7025F: A Family Study: A previously unreported heterozygous variant in *DCHS2* is associated with CAKUT. 2757

Board 7026F: A method of identifying risk genes using ultra-low coverage whole-genome sequencing - insights into embryo aneuploidy 2758

Board 7027F: Analysis of maternal-fetal genetic incompatibility in preeclampsia 2759

Board 7028F: Calculating Maternal Polygenic Risk Scores from Non Invasive Prenatal Testing Data 2760

Board 7029F: Differential RNA editing in the placenta identifies genes relevant to preterm birth 2761

Board 7030F: EHR phenotypes associated with genetic testing and diagnosisin the NICU ⭐ 2762

Board 7031F: Exome sequencing uncovers recurrent and novel genetic causes of spermatogenic failure in a large cohort of infertile men 2763

Board 7032F: Expansion of the phenotype spectrum due to a splice variant in *WDR44*-gene - severe prenatal phenotype in multiple male fetuses in a Finnish family 2764

Board 7033F: Functional assessment of *PASD1* gene variants in male infertility 2765

Board 7034F: Genome-wide identification of epigenomic elements in human embryonic palatal cells 2767

Board 7035F: Genome-Wide Meta-Analysis of Single Variant Associations in Uterine Fibroid Samples 2767

Board 7036F: High frequency of *de novo* variants and parental DNA damage response gene variants in couples with recurrent pregnancy loss 2769

Board 7037F: Highly variable expressivity of a CNV deletion involving *TBX4* in three siblings with lethal lung developmental disorder and their mildly affected mother and grandfather..... 2770

Board 7038F: Human data, machine learning, and mouse models provide evidence that *WNT4* plays a critical role in heart and diaphragm development. ⭐ 2771

Board 7039F: Identification of fetal nucleated red blood cells suitable for cell-based non-invasive prenatal genetic testing using single-cell multi-omics sequencing. 2772

Board 7040F: Identifying temporal postnatal modifiers of Hirschsprung Disease using transcriptomic, chromatin accessibility and cellular composition changes in the mouse gut..... 2773

Board 7041F: Incidence and sex ratio in children born with tetralogy of Fallot in the Czech Republic 2774

Board 7042F: Multi-omics framework discovers novel biomarker and genetic signatures for women's reproductive health..... 2775

Board 7043F: NGS-based genetic analysis of embryos for chromosomal abnormalities in Nepal 2776

Board 7044F: Pleiotropic effect of *de novo* variants in structural congenital anomalies and neurodevelopment disorder..... 2777

Board 7045F: Pre-implantation genetic testing by polygenic risk scores selects different embryos across score construction methods with randomness 2778

Board 7046F: Reassessment of Variants of Unknown Significance (VUS) in cases of undiagnosed fetal anomalies 2779

Board 7047F: Towards better reproductive health in Qatar: a custom genotyping array for expanded genetic carrier screening..... 2780

Board 7048F: Unraveling the mutational landscape of Placenta Accreta Spectrum (PAS) disorder 2781

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session 2782

Board 4002F: A Bayesian Rare Variant Association Analysis Method Incorporating Functional Variant Information ⭐ 2782

Board 4003F: A multi-ancestry reference panel from 10,721 individuals to improve the genotype imputation for individuals of East Asian- and Polynesian-Ancestry 2783

Board 4004F: A New Approach for Discovery of Functional Links Connecting Non-Coding Regulatory Variants to Gene Targets 2785

Board 4005F: A novel GWAS single-cell eQTL colocalization test improves power and resolves potentially disease-mediating cell types 2785

Board 4006F: A robust and efficient framework in transcriptome-wide association studies using cross-tissue learner 2786

Board 4007F: A scalable workflow for submicron-resolution spatial transcriptomics data from Seq-Scope 2787

Board 4008F: A Statistical Framework for Partitioning the Effects of Alleles Mediated by Parental Genetic Nurture and Direct Genetic Effects while Controlling for Assortative Mating 2788

Board 4009F: Activating Transcription Factor 3 Associated with Nausea in Gastroparesis Patients - a Whole Genome Sequencing Study 2789

Board 4010F: Advances in genetic relatedness software and data simulation for scalable, high throughput analysis 2790

Board 4011F: Allele stacking for cohort-scale analysis of long-read sequencing to identify functional non-coding variants 2791

Board 4012F: An accurate genetic colocalization method for the HLA locus 2792

Board 4013F: An Angle-based Bayesian Transfer Learning Approach for Enhanced Estimation and Inference 2793

Board 4014F: An efficient network Mendelian randomization method for network structure discovery and effect estimation 2794

Board 4015F: An omnibus test for multivariable Mendelian randomization with applications to FunGen-xQTL data 2795

Board 4017F: Ancestry-aware genome-wide association study of African American multiple sclerosis patients in All of Us ⭐ 2796

Board 4018F: Approximation of the polygenic risk score effects in a target population from summary statistics 2797

Board 4019F: Artificial Intelligence for Unsupervised Population Phenotyping to Enhance Disease Genotype to Phenotype Association Discovery 2799

Board 4021F: Association of orofacial cleft risk loci with elevated risk of prostate cancer 2800

Board 4022F: Augmenting Therapeutic Target Discovery with Polygenic Risk Scores . 2801

Board 4023F: Bayesian fine-mapping in two ancestry populations to identify putatively causal variants for cross-ancestry prediction of CYP2A6 activity 2802

Board 4024F: Benchmarking of single-cell eQTL methods with experimental data from DICE database 2803

Board 4025F: Beyond Bias: Advancing PCA for Genetic Research 2804

Board 4026F: BinBVS: Bayesian variable selection for classification of binary features, applied to multi-genic rare variant data from Phenome-wide association studies in UK Biobank 2805

Board 4028F: Characterisation of diverse global ancestries within participants of the UK Biobank 2806

Board 4029F: Characterizing disease-critical contrastive gene programs in multi-context omics assays ⭐ 2807

Board 4030F: Co-expression-wide association studies implicate protein-protein interactions in complex disease risk ⭐ 2808

Board 4031F: Combining common and rare variants for disease risk prediction ⭐ .. 2809

Board 4032F: Common Variant Gene Score Accurately Aggregates Genetic Effects for Alzheimer's Disease 2810

Board 4033F: Computationally efficient approaches for whole-genome quantile inference with related samples 2811

Board 4034F: Construction of skin aging score using machine learning model for discovery of skin aging genes 2812

Board 4035F: Convergent functional impact of common polygenic risk for neuropsychiatric disorders (NPDs) in genes associated with NPDs through rare, disruptive variation 2813

Board 4036F: DAESC-GPU: A GPU-powered scalable software for single-cell allele-specific expression analysis 2814

Board 4037F: Data-driven phenotyping to optimize human facial traits in GWAS and rare variant analysis 2815

Board 4038F: Deep learning predicts TF binding variation, illuminating breast and ovarian cancer biology 2816

Board 4039F: Detecting meiotic non-crossovers recombination in small pedigrees .. 2817

Board 4040F: Deviations from additivity in *APOE4*-associated Alzheimer's disease ⭐ 2818

Board 4041F: Discovery of rare coding variants with loss of function effect using a sliding window technique on clinicogenomic data 2819

Board 4042F: Disentangling chronic pain genetic heterogeneity from UK Biobank self-reports using graph-embedded topic model 2820

Board 4043F: Efficient large scale PheWAS analysis with PheTK 2821

Board 4044F: Efficient Penalized Regression Approach for Admixture-Aware Polygenic Risk Scores 2823

Board 4045F: EIR-auto-GP: A toolkit for deep learning based genomic prediction and analysis 2824

Board 4046F: Empowering Researchers: A Cost-Effective and Versatile PRS Calculator for the All of Us Program 2825

Board 4047F: Enhancing gene burden analyses in the UK Biobank with proteomic data 2826

Board 4048F: Enhancing polygenic scores prediction across multiple ancestries with S4-Multi model 2827

Board 4050F: Evaluating effectiveness of polygenic risk scores in African ancestry breast cancer patients from Penn Medicine BioBank 2828

Board 4051F: Evaluating long-read phasing methods in clinically-relevant genes using data from the 1000 Genomes ONT Sequencing Consortium 2829

Board 4052F: Evaluating Methods for Genome-Wide Association Studies in Diverse Ancestral Populations 2830

Board 4053F: Evaluating multi- and cross-ancestry approaches to Alzheimer's disease polygenic risk scores in diverse cohorts 2831

Board 4054F: Evaluating the impact of race, ethnicity, and ancestry (REA) as population labels in genomics research ⭐ 2832

Board 4055F: Evaluation of rare variant association methods when incorporating data from multiple sources 2833

Board 4056F: Exome-wide copy number variation in 106,744 individuals from autism spectrum families in the Simons SPARK cohort 2834

Board 4057F: Explore the use of graph neural networks to study the connection between the gut microbiome and bone health 2835

Board 4059F: Exploring the Effects of Genetic Variants on Gene Regulatory Networks 2836

Board 4060F: Exploring the mediating role of inflammation and immune-related gene expression in the chronic stress-cognitive function pathway in European and African ancestry populations 2838

Board 4061F: Fast and efficient toolkit for rare variants association testing 2839

Board 4062F: Fast multi-cohort approximate fine-mapping 2840

Board 4063F: Functional components of heritability in admixed African-ancestry All of Us whole-genome sequencing data 2841

Board 4064F: Gene Dosage Architecture in Complex Traits and Common Diseases with Mediation of the Brain and Body ⭐ 2841

Board 4065F: Gene-environment interactions are critical for future improvement in translational genetics 2843

Board 4066F: Gene-level Analyses of Novel Olfactory-related Signal from Severe SARS-CoV-2 GWAS Reveal Robust Association with Disease Mortality 2844

Board 4067F: Genetic determinants of sleep architecture inferred from accelerometer-based wearable devices in the UK Biobank ⭐ 2845

Board 4068F: Genetic evidence for a dose-dependent response to a *TYK2* inhibitor in psoriasis 2846

Board 4069F: Genetic Variants Associated with Muscle Strength in the United Arab Emirates: Insights from the Emirati Genome Project 2847

Board 4070F: Genetically regulated enhancer RNA expression predicts enhancer-promoter contact frequency and explains a substantial proportion of GWAS signals not attributable to canonical gene eQTLs 2848

Board 4071F: Genome-wide association study identified novel structural variants associated with Alzheimer's disease pathology in Non-Hispanic Whites ⭐ 2849

Board 4072F: Genome-wide association study of protein-altering and regulatory variants with resistance to *Mycobacterium tuberculosis* infection 2851

Board 4073F: Genome-wide comparison of allele specific gene expression in Non-Hispanic White, African American, Caribbean Hispanic, and Peruvian populations ⭐ 2853

Board 4074F: Genome-wide investigation of genetic variants that function as both eQTLs and pQTLs independently and via meta analysis 2854

Board 4075F: Genome-Wide Linkage and Association of Novel Genes and Pathways with Type 2 Diabetes in Italian Families 2855

Board 4076F: Genome-wide meta-analysis derived polygenic risk score predicts chronic back pain 2856

Board 4077F: gmmcoda: Graphical model for the mixture of compositional data and absolute abundance data with applications to microbiome studies 2857

Board 4078F: GWAS of 59,102 glaucoma cases reveals 230 independent loci and highlights pathways involved in glaucoma 2858

Board 4079F: High Prevalence of *ALPK3* Premature Terminating Variants in Korean Hypertrophic Cardiomyopathy Patients 2859

Board 4080F: HLA region local heritability and genetic correlation reveals novel genetic pattern underlying complex traits across multiple ancestries 2860

Board 4081F: Human protein-small molecule interaction networks reveal the cross-talks between genomic components and metabolome across complex metabolic traits .. 2861

Board 4082F: Identification of microRNA-related SNPs associated with Lung, Kidney, Gastric, and Thyroid Cancer in Korean Cancer Prevention Study-II Biobank 2862

Board 4083F: Identification of new variants associated with urinary incontinence and its subtypes: a meta-GWAS analysis 2863

Board 4084F: Identifying Causal Biomarkers via Knockoffs in the Presence of Unmeasured Confounding 2864

Board 4085F: Identifying Disease-Relevant Cell States via Integration of GWAS and Single-Cell Omics Data Using Context-Specific Annotated Polygenic Risk Scores 2865

Board 4086F: Identifying single-cell eQTLs by leveraging cell type specificity and similarity and incorporating linkage disequilibrium 2866

Board 4087F: Identity-by-descent mapping using multi-individual IBD with empirical genome-wide multiple testing adjustment 2867

Board 4088F: Improved Genetic Association analysis identifies 29 novel loci influencing preclinical to disease Progression in lupus ⭐ 2868

Board 4089F: Improved heritability partitioning and enrichment analyses using summary statistics with graphREML ⭐ 2869

Board 4090F: Incorporating functional annotations to improve polygenic risk prediction in admixed individuals 2870

Board 4091F: Indiscriminate usage of surrogate variables leads to overfitting in differential gene expression analyses 2871

Board 4092F: Inferring causal direction between two traits using $R^2$with application to Transcriptome-Wide Association Studies (TWAS) 2873

Board 4093F: Inferring relative gene-level effect sizes across multiple complex traits using GWAS summary statistics 2874

Board 4094F: Integrating large-scale GWAS and single-cell disease atlas data to identify disease-critical cell populations corroborated by genetic associations 2875

Board 4095F: Integrating Neuroimaging with Genetics to Predict Risk of Hypertension 2876

Board 4096F: Investigating the Phenomic and Genomic Cognitive Architecture in a Multiethnic Asian population 2877

Board 4097F: Investigating the role of neighborhood socio-economic status and germline genetic risk on prostate cancer risk 2878

Board 4098F: Is the association between eye disease and cognitive function modified by a *PDE7A/MTFR1* genetic variant?An analysis of hospital data with replication in the Canadian Longitudinal Study on Aging 2879

Board 4099F: LAWS-HiC: A locally adaptive weighting and screening (LAWS) approach to improve detection of long-range chromatin interactions from Hi-C data 2880

Board 4100F: Leveraging genetic similarity to investigate understudied genetic variation associated with EHR-derived phenotypes in diverse patient biobanks 2881

Board 4101F: Leveraging single cell transcriptomics to identify pathogenic cell lineages 2882

Board 4102F: Leveraging transfer learning to bridge data gaps: Identifying unique sepsis risk factors in chronic disease patients 2883

Board 4103F: MACHINE: a robust multi-ancestry fine-mapping method using a continuous global-local shrinkage prior with applications in psychiatric diseases 2884

Board 4104F: Measuring causal effect heterogeneity across the human phenome by linking biobank-scale genomic data and electronic health records 2884

Board 4105F: Meta-analysis approach for copy number variation aggregation association tests in multi-ethnic studies 2885

Board 4107F: Modernizing crlmm for Robust Genotyping and Efficient Genomic Analysis 2886

Board 4108F: MR.RGM: An R Package for Fitting Bayesian Multivariate Bidirectional Mendelian Randomization Networks 2887

Board 4109F: Multi-level gene-environment interaction analysis of fish oil supplementation and 14 circulating polyunsaturated fatty acid-related traits 2888

Board 4110F: On the use of Random Forest for obtaining Genetic Risk Score 2889

Board 4111F: Optimizing multitrait approach in Brain imaging GWAS ⭐ 2890

Board 4112F: Performance of two-phase designs for the time-to-event outcome of genomic sequencing studies and a case study assessing B-ALL primary mutation associated relapse risk 2891

Board 4113F: Pharmacogenomic genome-wide association study of LDL response to statins using high-throughput electronic health records analysis 2892

Board 4114F: PhenoGPT2: adapting domain-specific Retrieval Augmented Fine-Tuning and In-context learning to enhance recognition of phenotypic abnormality from clinical notes 2893

Board 4115F: Polygenic Modeling with Sparse LD Representation for Enhanced Efficiency and Performance 2894

Board 4116F: Population-based biobank analysis reveals elevated frequencies of G6PD deficiency Variants in Quebec. 2895

Board 4117F: Population-enriched germline variants in Finns pinpoint shared mechanisms shaping hematopoietic mosaic chromosomal alterations and diverse solid tumors ⭐ 2896

Board 4118F: Powerful estimation of the genetic component of gene expression with a Poisson mixed effect model of single cell RNA-sequencing reads..... 2897

Board 4119F: Predict Disease Risks and Trait Values by Screening Thousands of Polygenic Scores 2899

Board 4120F: Prediction at scale of lumping and splitting decisions to define monogenic gene-disease entities..... 2900

Board 4121F: Proteome-wide adjusted pQTLs: A tale about mediation, collider, increase in power and how to distinguish them 2901

Board 4122F: PRS-Guided Sequencing and Large-Cohort Imputation: Unveiling Rare Variant Associations in Complex Diseases ⭐ 2902

Board 4123F: Python-based pipeline for measuring imputation accuracy and assessing the impact of reference panel selection on genetic variant imputation in different study populations and across genomic regions 2903

Board 4124F: Rapid Long-Range Linkage Disequilibrium Calculations at Biobank Scale using GPU Acceleration 2904

Board 4125F: RareEffect: Rare variant effect size estimation and polygenic risk prediction 2905

Board 4126F: Regulatory context stratifies genotype-phenotype maps into biological pathways 2906

Board 4127F: Relatedness and rotator cuff tears: Leveraging biobank data to understand source of associations 2907

Board 4128F: scTWAS: A powerful statistical framework for single-cell transcriptome-wide association studies 2908

Board 4129F: Sex-differences in deviation from Hardy-Weinberg Equilibrium 2909

Board 4130F: Single-nucleus transcriptome-wide association study of human brain disorders 2910

Board 4131F: snputils: A Python library for processing diverse genomes 2911

Board 4132F: Statistical identification of cell type-specific spatially variable genes in spatial transcriptomics 2912

Board 4133F: Streamlined pipeline for Tractor GWAS Unveils Ancestry-Specific Variants in Admixed Populations 2913

Board 4134F: Successful Implementation of a Method for the Reconstruction of Parent Genomes from Two or More Children 2914

Board 4135F: The Bias of Using Cross-Validation in Genomic Predictions and Its Correction 2915

Board 4136F: The effects of rule-based phenotyping on cohort composition and genome-wide association study findings in the UK Biobank 2916

Board 4137F: The genetic atlas of nurture and parent-of-origin effects in ~220,000 UK Biobank individuals ⭐ 2917

Board 4139F: The genetic epidemiology of Usher syndrome in a South Florida population 2919

Board 4140F: The pivotal role of the X-chromosome in the genetic architecture of the human brain 2920

Board 4141F: The role of sleep traits in diabetes onset: Evidence from Mendelian Randomization using self-reported data 2920

Board 4142F: Tissue-specific epigenomic profiles inform pleiotropic partitioning of disease loci 2921

Board 4143F: Trans-ancestry meta-regression analysis MR-MEGA replicates European IBD loci using multiple ancestries 2922

Board 4144F: Transcriptome wide association study identifies novel candidate susceptibility genes for inguinal hernia 2923

Board 4145F: Transferability of single- and cross-tissue transcriptome imputation models across ancestry groups 2925

Board 4147F: Two novel, unbiased, sibling-based genetic methods greatly increase precision of heritability estimates and quantify gene-environment interactions 2926

Board 4148F: Uncovering causal gene-tissue pairs and variants: A multivariable TWAS method that controls for infinitesimal effects ⭐ 2927

Board 4150F: Unveiling the power of allelic series: enhancements and applications of COAST 2928

Board 4151F: Valid inference for machine learning-assisted GWAS 2928

Board 4152F: Viewing the role of aging through the prism of changing genetic effects2929

Board 4153F: Vitamin D receptor-binding variants confer risk of Sjögren's disease in Europeans 2930

Wednesday, November 6, 2024, 2:30 pm - 4:30 pm

Session Title: Cancer Poster Wednesday Session

Board 8001W: A breast cancer risk prediction model integrating both genetic and non-genetic factors in women of Nigeria, Cameroon, and Uganda

Authors:

W. Guo[1], J. Li[1], J. McClellan[1], A. Ntekim[2], O. Ojengbede[2], J. Obafunwa[3], P. Ndom[4], A. Gakwaya[5], N. Chatterjee[6], O. I. Olopade[1], D. Huo[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Ibadan, Ibadan,

Nigeria, [3]Lagos State Univ. Teaching Hosp., Lagos, Nigeria, [4]Yaounde Gen. Hosp., Yaounde, Cameroon, [5]King Ceasar Univ., Kampala, Uganda, [6]Johns Hopkins Univ., Baltimore, MD

Abstract:

Background: It is unknown if adding genetic information to risk prediction models based on epidemiologic factors can improve risk assessment for African women. We aim to build a comprehensive breast cancer risk prediction model, by integrating a polygenic risk score (PRS), pathogenic variants (PVs) of high/moderate penetrance genes, and a questionnaire-based risk calculator.

Methods: This study is based on the Nigerian Breast Cancer Study, which enrolled breast cancer cases and controls in Nigeria and extended to Cameroon and Uganda between 1998 and 2018. A joint PRS was utilized, which combines a PRS previously developed in the African ancestry training dataset (using data from 5 genome-wide association studies) and a 313-variant PRS developed previously in women of European ancestry. We excluded individuals whose samples were utilized in the training of the PRS for breast cancer. We integrated the PRS (consisting of 54,016 common genetic variants), the effects of PVs in *BRCA1*, *BRCA2*,*PALB2, ATM, CHEK2, TP53, BARD1, RAD51C*, and *RAD51D*, with epidemiological risk factors ascertained from questionnaires (the NBCS calculator). We estimated lifetime absolute risk of breast cancer by age 80 years using the R package iCARE, while accounting for competing risk and correlation between family history and genetic factors.

Results: A total of 866 cases and 666 controls were included in the analysis. The age-adjusted area under the ROC curve (AUC) was 0.579 (95% confidence interval [CI] 0.549-0.610) for the PRS only model and 0.609 (95% CI 0.578-0.640) for the PRS plus PV model. In the full model containing both genetic and non-genetic risk factors, the age-adjusted AUC increased to 0.729 (95% CI 0.704-0.755). Using a threshold of lifetime absolute risk of 0.05, the full model classified 27.8% of study participants as high risk, compared with 19.2% in the model containing epidemiological factors only and 9.5% in the genetic model that contains both PRS and information on the PVs.

Conclusion: There is no universal breast cancer screening program in Africa, in part due to limited resources and relatively lower incidence rate. However, the majority of breast cancers in Africa are diagnosed at advanced stages; personalized risk assessment coupled with risk stratified screening for early detection and prompt treatment have the potential to reduce breast cancer mortality. We have developed a breast cancer risk prediction model specific to Nigerian women, combining genetic and non-genetic factors. Further development and validation in other African countries will be necessary to identify women at high-risk for breast cancer for targeted screening and prevention interventions.

Session Title: Cancer Poster Wednesday Session

Board 8002W: A comprehensive quantitative catalog of differentially expressed mice tissues whole transcriptomefor understanding early cancer-causing events

Authors:

M. Tariq[1], O. Jejelowo[2]; [1]Shenzhen Inst. of Advanced Technology (SIAT), Shenzhen, China, [2]Ctr. for Bionanotechnology and Environmental Res., Texas Southern Univ., Houston, TX

Abstract:

Radiation exposure presents major challenge during human space exploration missions. Astronauts travelling in space during long duration missions are exposed to new radiation sources, which their bodies had not encountered on earth. They are also exposed to mixed and higher levels of radiation from the space environments, including high-energy proton radiations that could have severe health implications on human health during interplanetary missions. Radiation is genotoxic and induces DNA lesions in human cells that could induce development of cancer and tumor formation. An understanding of how cells respond to proton radiation would shed light on how space radiation might affect humans during exploration missions and also on the etiology of various cancers. This study examined alterations in global gene expression response to whole-body acute dose proton irradiated (2.0Gy) brain, liver and testes tissues of mice by whole transcriptome analysis using next-generation sequencing. There was significant deregulation ($p < 0.05$) of a total of 622 genes in brain tissues (534 genes were upregulated, 88 genes were downregulated), 968 genes in liver tissues (456 genes were upregulated, 512 genes were downregulated) and 763 genes in testis tissues (349 genes upregulated, 414 genes downregulated). Many of the genes that were dysregulated in response to radiation in the brain, liver and testes tissues are those involved in significant pathways of tumorgenesis, including, cell proliferation, cell growth, invasiveness and metastasis. These mice tissue gene expression response to irradiation suggests that radiation exposure impacts early cancer-causing events. Further, we identified functional gene ontology categories associated with proton irradiation. Many of the over represented biological processes in response to proton irradiation include induction and regulation of cellular processes, apoptosis, cell death, cell cycle, caspase activity, cell signaling, calcium mediated cell signaling, exocytosis, response to stimulus and endoplasmic reticulum (ER) unfolded protein response to stress. This study provides a comprehensive quantitative catalogue of differentially expressed mice whole transcriptome in response to proton radiation. Information obtained may enhance accurate estimation of the radiation safety standards in space during exploration missions and on earth during “proton radiation therapy”, for cancer patients. Results are discussed in light of potential

of radiation to trigger early cancer-causing events and the natural response to these events by suppression pathways.

Session Title: Cancer Poster Wednesday Session

Board 8003W: Acute endothelial stresses identify microRNA let-7b and non-coding *SLC11A2 (NRAMP2/ DMT1)* exon as biomarkers that overlap with those detected in malignant and non-malignant diseases

Authors:

C. Shovlin[1], A. Bielowka[1], D. Li[1], M. Bernabeu-Herrero[1], L. Game[2], M. Aldred[3], I. Mollet[4], D. Patel[1]; [1]Imperial Coll. London, London, United Kingdom, [2]UKRI MRC Lab. of Med. Sci., London, United Kingdom, [3]Indiana Univ., Indianapolis, IN, [4]Univ.e NOVA de Lisboa, Caparica, Portugal

Abstract:

Disease biomarkers are often identified many years after the initiating pathologies, hampering mechanistic deductions and disease prevention. Our objectives were to model underexplored acute stresses that blood-exposed cells need to withstand, and cross-reference biomarkers identified for disease states. Normal primary human microvascular endothelial cells (ECs) were treated for 1-6h with cycloheximide 100µM to inhibit protein translation and nonsense mediated decay (modelling the integrated stress response), or 10µM ferric citrate (modelling iron level changes after treatments prescribed 8 million times/year in England). Directional whole transcriptome RNA-seq was performed, differentially expressed genes clustered in pathway analyses, and 1h micro(mi)RNA changes validated by 6h reciprocal changes in target mRNA targets. Customized novel scripts examined expression of 517,225 exons to predict 1h cycloheximide-stabilized exons. Validations were by cel-miR-39-spiked qRT-PCR and RNA-seq in other endothelial cell types, peripheral blood mononuclear cells (PBMCs), and plasma. Pathway analyses indicated appropriate endothelial responses to the treatments. miRNAs fell transiently at 1h after 10µM ferric citrate ($p<0.01$), specifically in let-7 family member pre-miRNAs ('let-7', $p<0.05$), where there was an accompanying differential 6h increased expression of 570 let-7-target mRNAs identified through TargetScan ($p<0.0001$). qRT-PCR and RNA-seq validations in other normal endothelial cells, plasma and PBMCs confirmed pre-let-7b/let-7b-5p as biomarkers for 1h iron responses, and exon 3B of the *SLC11A2* (*NRAMP2/DMT1*)-encoded iron transporter as a novel exon most consistently stabilized following 1h treatment with cycloheximide. Notably, *SLCA11A2* overexpression is a new biomarker/therapeutic target for ovarian cancer while let7s/let7-b/let-7b-5p are now identified as disease biomarkers for multiple cancer types,

organ-specific diseases, and growth retardation. We conclude that biomarkers for diverse disease states overlap with normal cellular responses, suggesting potential for maladaptation.

Session Title: Cancer Poster Wednesday Session

Board 8004W: AlphaMissense pathogenicity scores predict response to immunotherapy and enhances the predictive capability of tumor mutational burden ⭐

Authors:

D. Adeleke[1], R. Gomes[2], R. Jansen[3]; [1]North Dakota State Univ., Fargo, ND, [2]Univ. of Wisconsin, Eau Claire, WI, [3]Univ. of Minnesota, Minneapolis, MN

Abstract:

The US Food and Drug Administration (FDA) approved Immune checkpoint inhibitor (ICI) for treating cancer patients with tumor mutation burden (TMB) < 10 mutations/megabase. However, the variation in what defines a high TMB raised serious concerns on suitability of TMB as a predictive biomarker. Recently, Cheng et al. developed AlphaMissense (AM) - a deep learning model to predict the pathogenicity score (PS) for 71 million missense variants. In this study, we evaluate AM PS as a possible biomarker for ICI treatment response, and its association with TMB. Using the publicly available dataset from Samstein et al. Nat Genet 2019, we analyzed the clinical and genomic data of 840 advanced cancer (Bladder, Breast, Cancer of Unknown Primary, Colorectal, Esophagogastric, Glioma, Head & Neck, Melanoma, Non-Small Cell Lung, and Renal Cell) patients treated with ICI. We computed PS for each patient as a sum of AM PS > 0.4 and assessed correlation with TMB. Furthermore, we derived tumor mutational pathogenicity rate (TMPR) for each patient - as the ratio of PS and TMB . Lastly, we evaluated the association of the PS and TMPR with survival after treatment with ICI within cancer types and overall. Survival and Survminer R packages were used for survival analysis, and R Stats package for correlations. There are 9,725 distinct nonsynonymous mutations in the cohort, 36% are pathogenic and 35% are benign based on AM PS cut-off of 0.4. Also, we observed a strong positive correlation between the TMB and PS, r(840) = 0.8341 p < 0.001. In all cancers combined, patients with higher somatic TMB or PS (highest 30% in each histology) had better overall survival (OS) HR_TMB=0.49(95% CI, 0.39 - 0.62); p < 0.001, HR_PS=0.67(95% CI, 0.54 - 0.83); p < 0.001. On the contrary, patients with low TMPR had better OS, HR_TMPR=0.74(95% CI, 0.6 - 0.9); p = 0.003. For cancer type-specific analysis, patients with higher TMB or PS showed a significant association with OS in colorectal and Non-Small Cell Lung cancers only (p < 0.001) while patients with low TMPR showed better OS in

melanoma and Head and Neck cancers only (p < 0.001). This study indicates that there is a strong correlation between PS and TMB, and both are associated with improved survival in patients receiving ICI across multiple cancer types. While there may not be one universal definition for either high TMB or high PS, this study showed that pathogenic variants which are about 36% of all TMB offer as much efficacy prediction as overall TMB. It also appears that the ratio of PS to TMB could offer additional information important for certain types of cancers when predicting survival following ICI treatment, rather than a single measure score.

Session Title: Cancer Poster Wednesday Session

Board 8005W: An ex vivo screening for growth factor combinations reveals estrogen-responsive cellular signatures contributing to immunosuppressive subtypes of ovarian cancer

Authors:

K. Lee[1,2], M. Kim[3], H. Kang[1,2], S. Lim[3], J-K. Lee[3,2,1]; [1]Med. Res. Ctr., Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of, [2]Dept. of Anatomy and Cell Biology, Seoul Natl. Univ., Coll. of Med., Seoul, Korea, Republic of, [3]Dept. of BioMed. Sci., Seoul Natl. Univ., Coll. of Med., Seoul, Korea, Republic of

Abstract:

Epithelial ovarian cancer (EOC) is the most lethal gynecological malignancy, often presenting at an advanced stage with a 5-year survival rate below 40%. Large-scale genomic studies, such as TCGA, have elucidated the molecular characteristics of high-grade serous ovarian cancer and identified four distinct consensus molecular subtypes. However, the underlying biological mechanisms driving this heterogeneity remain poorly understood. To better reflect the diverse growth environments of ovarian cancer, we created 128 combinations of 7 different growth factors and tested their effects on tumor samples from 18 ovarian cancer patients. We found that estrogen response is a key clustering factor characterizing growth phenotypes, and an estrogen-responsive gene signature is associated with the immunosuppressive subtype of ovarian cancer. Additionally, tumors with high GPER activity exhibited elevated expression of mesenchymal subtype-related signatures. Transcriptomic profiles of single-cell resolution showed that the estrogen-responsive cell fraction had strong interactions with TGF-β-type cancer-associated fibroblasts (CAFs), which are known to promote an immunosuppressive tumor microenvironment. Importantly, this estrogen-responsive fibroblast proportion correlates with worse overall survival in EOC patients. In conclusion, our study demonstrates that the estrogen-responsive cell fraction and gene signature provide a fundamental biological basis for

the consensus molecular subtypes of ovarian cancer. These findings suggest that estrogen signaling plays a crucial role in driving the molecular heterogeneity and distinct tumor microenvironments observed in ovarian cancer.

Session Title: Cancer Poster Wednesday Session

Board 8006W: An integrated platform for cancer precision medicine: Genetic testing coupled with patient-derived organoid (PDO) and patient-derived xenograft (PDX) models ⭐

Authors:

Y. Da, J. Ma, H. Peng, Z. Zhou, Y. Wan, Y. Yu, J. Ma, P. Zhang, T. Gong, C. Zhang; The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China

Abstract:

Genetic testing by next generation sequencing (NGS) plays critical roles in cancer precision medicine. However, most cancers do not carry genetic mutations that may guide the target therapy and immunotherapy. Moreover, tumor heterogeneity and drug resistance pose additional challenges to effective treatment of cancers. Therefore, developing a feasible and effective preclinical platform is urgently needed. The patient-derived organoids (PDOs) that capture tumor heterogeneity and maintain genetic stability have emerged as a promising preclinical model for human cancers. The patient-derived xenografts (PDXs) that retain the biological behavior and therapeutic response of the parental tumors have been widely used as in vivo models for preclinical studies of cancers, but the long testing time, low successful rate and high cost make them unsuitable for high-throughput drug screening and guiding individualized therapy. The mini-PDX models, which were established by injecting patient-derived primary tumor cells into hollow fiber capsules and implanting them under the skin of mice, are more rapid and accurate in vivo drug screening models that not only shortens the cycle of animal experiments in vivo to one week, but also may effectively and accurately predict patients' clinical responses to targeted therapy and chemotherapy. In this study, we tested a variety of cancer types, including lung cancers, gastric cancers, pancreatic cancers, ovarian cancers, breast cancers and kidney cancers, and developed a feasible platform by integrating the genetic testing with the PDO and PDX models for a rapid drug screening in vitro and in vivo, which holds a great promise for personalized cancer therapy.

Session Title: Cancer Poster Wednesday Session

Board 8007W: An MDS patient with deletion 20q and a t(9;22)(q34;q11.2): A case report and review of the literature

Authors:

S. Roush[1], **C. Tirado**[1,2,3], R. Hurtado[2], A. Zavala[1], E. Lee[1], T. Ahmed[1,3]; [1]Stony Brook Univ. Hosp., Stony Brook, NY, [2]The Intl. Circle of Genetics Studies, Stony Brook Chapter, Stony Brook, NY, [3]The Cytogenetics Lab at Dept. of Pathology, Stony Brook Univ. Hosp., Stony Brook, NY

Abstract:

We report a 76 year old male patient with myelodysplastic syndrome (MDS) with a concurrent deletion 20q and t(9:22) translocation. Past medical history is significant for prostate cancer status post radiation therapy and a 28 pack/year smoking history. In 2016 patient underwent a coagulation study following the discovery of a DVT. Molecular genetic BCR-ABL by PCR analysis was positive for BCR-ABL (p210) mutation at level of 0.54% (international scale). Subsequent bone marrow aspiration revealed a hypercellular bone marrow with a small monoclonal B-cell population observed on flow cytometry, consistent with chronic myelogenous leukemia (CML). FISH analysis demonstrated t(9;22) translocation in 1.5% of nuclei, in addition to a loss of 20q12 in 5% of nuclei. Patient was started on Nilotinib therapy. Follow up BCR::ABL1 testing six months later did not detect BCR::ABL1, however subsequent FISH analysis on bone marrow aspirates performed one and seven years after initial diagnosis continue to show deletion 20q (1-3% of nuclei). Morphology features of bone marrow aspirates have described a hypercellular bone marrow with myeloid and megakaryocytic hyperplasia. Megakaryocytes were described as micromegakaryocytic and hyperlobulated in appearance. Absolute neutrophil count have persistently remained low (<2.4K/uL) and peripheral blasts <5%, supporting the diagnosis of MDS with concurrent deletion 20q.

Session Title: Cancer Poster Wednesday Session

Board 8008W: Assessing the impact of genetic variations on metformin's efficacy in treating colorectal cancer

Authors:

A. Saeed, B-A. Bengoechea-Alonso, J. Ericsson, B. Mifsud, G. Nemer; Hamad Bin Khalifa Univ., Doha, Qatar

Abstract:

Colorectal cancer (CRC) ranks as the world's third most common cancer. Early-stage CRC does not present any symptoms, often leading to advanced-stage diagnosis. CRC treatment efficacy varies due to high heterogeneity, prompting the search for alternatives. Metformin, a widely used anti-diabetic drug, has potential anti-cancer properties. This study evaluated the impact of genetic variations on metformin's efficacy in treating CRC. The HCT116 colorectal cancer cell line harboring *KRAS* mutation was utilized, with its tumor suppressor genes (p53 and p21) knockouts. The methodology involved a dose-response (0.5, 1, 2, 4, and 8 mM) and time-course (24 and 48 hours) treatment to evaluate cell growth inhibition through microscopic visualization, trypan blue-based live cell counting, and flow cytometry-based cell cycle analysis, followed by RNA sequencing analysis. Results indicated varying sensitivity to metformin-induced cell growth inhibition, with HCT116-wild type showing the highest sensitivity to metformin, followed by p53 and p21 knockouts of HCT116. Differential RNA expression of the cell lines identified *FAM111A*, *DUSP5*, *FGD6*, *PLAUR,* and *TNFAIP3* as the only genes commonly up- or down-regulated among the three cell lines following treatment with metformin. Comparison of p53 and p21 knockouts of HCT116 using pathway enrichment analysis revealed that the p21 knockout cell line resisted metformin's effect. The p53 knockout cell line displayed stunted cell growth despite the absence of the p53 protein. In conclusion, CRC with *KRAS* mutations is more sensitive to metformin than the ones with tumor suppressor gene mutations, highlighting the need to treat CRC based on the genetic mutation profile of the patients. Furthermore, these findings suggested that metformin's anti-cancer properties may depend on multiple molecular pathways to inhibit cell growth, requiring further investigation for validation.

Session Title: Cancer Poster Wednesday Session

Board 8009W: Association of*HOXB13* G84E with any and lethal prostate cancer among 648,421 men

Authors:

T. Crawford[1,2], T. Nelson[3,4], R. Karunamuni[4], H. Desai[1,2], C. Teerlink[5,6], H. Carter[7], M. Pagadala[8,4], P. R. Alba[5,6], S. L. DuVall[5,6], M. E. Danowski[9], C. A. Brunette[9,10], D. Ratner[9], I. P. Garraway[11,12], B. S. Rose[8,4], K. A. Cooney[13], J. L. Vassy[9,10], R. L. Hauger[8,4], J. A. Lynch[5,6], T. M. Seibert[8,4], K. N. Maxwell[1,2]; [1]Corporal Michael J Crescenz VA Med. Ctr., Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]VA San Diego Hlth.care System, San Diego, CA, USA, San Diego, CA, [4]Univ. of California San Diego, La Jolla, CA, [5]VA Salt Lake City Hlth.Care System, Salt Lake City, UT, [6]Univ. of Utah, Salt Lake City, UT, [7]UCSD, La Jolla, CA, [8]VA San Diego Hlth.care System, San Diego, CA, [9]VA Boston Hlth.care System, Boston, MA, [10]Harvard Med. Sch., Cambridge, MA, [11]VA Greater Los

Angeles Hlth.care System, Los Angeles, CA, [12]Univ. of California, Los Angeles, Los Angeles, CA, [13]Duke Univ., Durham, NC

Abstract:

Background: The *HOXB13* c.G251G>A:p.G84E variant is a moderate prostate cancer (PCa) risk allele in individuals of European descent. However, there are minimal data on population-based risk estimates, clinical phenotypes, and patient outcomes associated with this variant. We report age-specific PCa risks and clinical pathological associations in a population-based cohort of *HOXB13* p.G84E carriers identified from 648,421 men that were genotyped as a part of the Million Veteran Program.
Methods: *HOXB13* p.G84E carriers were identified by single nucleotide polymorphism (SNP) genotyping in Million Veteran Program (MVP) participants. Cox proportional hazard models were used to determine the association of the variant with a lifetime risk of developing PCa, including grade, severity, and death. Multivariable logistic regression models were used to determine the association of the variant with risk of PCa diagnosis on first biopsy in a sub-cohort of Veterans diagnosed with PCa in VA healthcare.
Results: This MVP study included a total of 648,421 males with 1885 (0.3%) carriers of the *HOXB13* G84E variant. Overall incidence of PCa in *HOXB13* G84E carriers was 31.2% compared to 12.7% in non-carriers (p=2.2e-16). *HOXB13* G84E carriers had an increased lifetime risk of any, metastatic, and fatal prostate cancer, with hazard ratios of 2.76 (95% CI 2.53-3.01, p=6.1e-117), 2.71 (95% CI 2.08-3.53, p=1.9e-13), and 2.56 (95% CI 1.49-4.42, p=7.2e-4). In contrast to previous studies, *HOXB13* G84E carriers and non-carriers had a similar age of PCa diagnosis. The distribution of Gleason scores was marginally lower in carriers vs non-carriers (p=0.02), and the rates of metastatic disease at diagnosis were similar. In the subset of MVP participants diagnosed with PCa at first biopsy (n=24,101), *HOXB13* G84E carriers (n=212, 0.88%) had higher PSA at first biopsy (p<.001) and higher risk of PCa and aggressive PCa on first biopsy (OR 2.61, 95% CI 1.96-3.51, p<0.001; OR:1.74, 95% CI 1.25-2.39, p<0.001).
Conclusions: In the largest cohort of men with the *HOXB13* G84E variant studied to date, this variant confers increased lifetime risk of any PCa, though lower than prior estimates based largely on high-risk clinically ascertained cohorts. Further work is needed to determine if initiation of early PCa screening in HOXB13 p.G84E carriers would improve patient outcomes.

Session Title: Cancer Poster Wednesday Session

Board 8010W: Building an Ecosystem for Genetic Testing of Hereditary Cancer Syndromes in Algerian Population: The Experience of our Research Laboratory (2008-2024)

Authors:

F. Cherbal[1], C. Mehemmai[1], M. Saadi[1], Y. Santoudji[1], D-E. Seddik[1], F. Khider[1], A-L. Boumehdi[1], H. Gaceb[1], F. Gouaref[2], K. Bentabak[3], H. Mahfouf[4], F. Zebboudj[5], M. Oukkal[6], M. Maaoui[7], W. Benbrahim[8], K. Boualga[9]; [1]Molecular Genetics Team, LBCM-Tamayouz, FBS, USTHB, Algiers, Algeria, [2]Med. Clinic of Debussy, EHS Pierre and Marie Curie, Univ. of Algiers 1, Algiers, Algeria, [3]Oncological Surgery and Liver Transplantation services, EHS Pierre and Marie Curie, Univ. of Algiers 1, Algiers, Algeria, [4]Mohamed El Kolli Publ. Hosp., Academic Med. Oncology Services, Rouiba, Sch. of Med., Univ. of Algiers1, Algiers, Algeria, [5]Mohamed El Kolli Publ. Hosp., Academic Gen. Surgery Services, Rouiba, Sch. of Med., Univ. of Algiers-1, Algiers, Algeria, [6]Clinic of Med. Oncology Amine Zirout, Univ. Hosp. of Beni-Messous, Sch. of Med., Univ. of Algiers-1, Algiers, Algeria, [7], Academic Gen. Surgery Services, Kouba, Sch. of medicine, Univ. of Algiers-1, Algiers, Algeria, [8]Med. Oncology Services, Ctr. Anti Cancer Batna, Univ. Batna2 Chahid Mostefa Ben Boulaid, Batna, Algeria, [9]Radiation Oncology Services and Lab. for Cancer Res., Anticancer Ctr. of Blida, Blida, Algeria

Abstract:

Background: New medical and genomics discoveries regarding human hereditary cancers, the availability of genetic counselling and testing for medical routine diagnosis and precision medicine have increased public attention to cancer genetics**.** To date, in Algeria, like in most of emerging countries, there is a limited access in clinical genetic services, mainly due to the costs, the availability of genetic testing and the lack of public insurance for coverage of genetic services. Here, we report the experience of our research laboratory and the building of an ecosystem for genetic testing of hereditary cancer syndromes in Algerian population (2008-2024). **Methods:** The patients and their families were referred through a network of five university hospitals, two anti-cancer centers and three private medical clinics, which provide oncology services throughout Algeria. Family histories of cancer were obtained from interviews, pedigrees and medical records of patients. In our laboratory, we tested eight (8) cancer predisposition genes in patients diagnosed with HBOC syndrome, FAP syndrome and Lynch syndrome (LS), respectively. *BRCA1*, *BRCA2*, *PALB2*, *APC*, *MLH*1, *MSH2*, *MSH6* and *PMS2* were screened using PCR-Sanger sequencing. In addition, we used NGS with multi-gene hereditary cancer panel (Color Genomics) to screen for these 8 genes. **Results:** Genetic testing of 374 HBOC patients, 66 FAP patients and 87 suspected LS patients, respectively, has been performed. The analysis revealed eleven (11) distinct germline pathogenic variants and one likely germline

pathogenic variant in *BRCA1* gene, five (5) distinct germline pathogenic variants in *BRCA2* gene, one new VUS in *PALB2* gene, 13 distinct germline pathogenic variants in *APC* gene and seven distinct germline pathogenic variants in MMR genes, respectively. Interestingly, six distinct recurrent germline pathogenic variants in *BRCA1* gene, one recurrent germline pathogenic variant in *BRCA2* gene, three distinct recurrent germline pathogenic variants in *APC* gene and three recurrent germline pathogenic variants in MMR genes, respectively, have been identified in our study. We also noticed a geographical difference in the distribution of the spectrum of sequence variants of *BRCA1*, *BRCA2* and MMR genes, respectively. Germline variants detected in the eight predisposition genes in our study have been submitted to ClinVar database (Organization ID: 508461) in order to help the genetic testing of Algerian patients living abroad. **Conclusions:** We built an ecosystem in Algeria to promote genetic testing and cancer screening to help patients to be diagnosed as early as possible and early identification of high risk family members.

Session Title: Cancer Poster Wednesday Session

Board 8011W: Challenges of germline predisposition testing in patients with hematologic malignancies

Authors:

J. Feusier[1,2], S. Brown[1,2], J. Solimine[1], L. Perry[1,2], K. Cone[1,2], M. Williams[1,2]; [1]ARUP Lab., Salt Lake City, UT, [2]Univ. of Utah, Salt Lake City, UT

Abstract:

The field of inherited predisposition to hematologic malignancies is growing due to increasing knowledge of these conditions and availability of genetic testing. It is critical to identify causative germline mutations in patients with hematologic malignancies so that related potential stem cell donors may be screened and transplantation with affected donor cells can be avoided. Peripheral blood is commonly used for germline testing; however, in patients with hematologic malignancies, blood may contain somatic variants that can confound germline analysis. Somatic variants may cause a false positive result if germline testing is performed on blood from patients with active disease. Furthermore, somatic genetic rescue of pathogenic germline variants, such as monosomy 7 in individuals with pathogenic germline SAMD9 or SAMD9L variants, can lead to a false negative result if germline testing is performed on blood. To avoid these issues, the recommended specimen type for germline genetic testing is cultured fibroblasts from a skin biopsy, but this is an invasive collection that adds several weeks to the turn-around-time due to culturing. Treatment decisions or stem cell transplant donor selection

may be contingent upon germline test results; therefore, it is critical to balance timeliness with accuracy. In this study, we describe the challenges of conducting informative germline testing in patients with hematologic malignancies. We examined results of 156 specimens submitted to our clinical laboratory for testing via our massively parallel sequencing (MPS) panels that target genes associated with predisposition to myeloid malignancies and bone marrow failure syndromes. Of these, only 17.9% (n=28) of tests were ordered on cultured fibroblasts, which is the recommended specimen type. One quarter (n=7) of the cultured fibroblast samples had a germline pathogenic predisposition variant identified, compared to 21.9% (n=28) of blood samples with a variant detected. Somatic mosaic findings were suspected in half (n=14) of the positive peripheral blood samples based on the MPS result, which complicated analysis and clinical interpretation and warranted further testing to clarify the result. The suspected mosaicism included: monosomy 7 (n=2), TP53 (n=4), GATA1 (n=2), GATA2 (n=2), RUNX1 (n=2), PTPN11 (n=1), and TERT (n=1) variants. Eleven of the 14 cases were known to have an active hematologic malignancy. Our results illustrate the limitations of peripheral blood testing for patients with active hematologic malignancies and underscore the importance of selecting the appropriate test and sample type for germline testing.

Session Title: Cancer Poster Wednesday Session

Board 8012W: Characterization of chromosome 22q deletion in papillary thyroid carcinoma

Authors:

O. Lee[1], D. Karyadi[1], S. Hartley[1], W. Zhou[2], M. Machiela[1], M. Tronko[3], T. Bogdanova[3], L. Zurnadzhy[3], L. Morton[1], S. Chanock[1]; [1]Natl. Cancer Inst., Bethesda, MD, [2]Leidos BioMed. Res./NCI, Bethesda, MD, [3]V.P. Komisarenko Inst. of Endocrinology and Metabolism of the Natl. Academy of Med. Sci. of Ukraine, Kyiv, Ukraine

Abstract:

Deletion in the long arm of chromosome 22 (22qDEL) is the most prevalent somatic copy number alteration (SCNA) observed in papillary thyroid carcinoma (PTC), the most common type of thyroid cancer. The genomic profile and contributory role of 22qDEL in PTC have not been fully investigated. We conducted a pooled analysis of 22qDEL by integrating patient and clinical information with molecular profiles from 1,094 primary PTC tumors across four large published PTC genomic studies: The Cancer Genome Atlas (TCGA) PTC profile study, the National Cancer Institute post-Chornobyl PTC study (CTB), the Memorial Sloan Kettering Cancer Center Integrated Mutation Profiling of Actionable Cancer Targets (MSKCC-IMPACT) PTC cases, and the

Seoul National University (SNU) thyroid cancer profile study. The majority of PTC with 22qDEL demonstrated arm-level loss of heterozygosity (86%). Even in PTC cases with partial 22qDEL, we observed a loss across 70% of chromosome 22 that included 22q12 and 22q13, which harbor known cancer genes such as *NF2*, *CHEK2*, and *EP300*. More than 90% of 22qDEL co-occurred with established oncogenic drivers of PTC. *RAS* point mutations most frequently co-occurred with 22qDEL (50.4% versus 9.3% for other drivers). A higher fraction of 22qDEL tumors were clonal in *RAS*-driven PTCs (78.3%), suggesting 22qDEL occurs early in *RAS*-driven tumor development. We did not observe any association between 22qDEL and established PTC risk factors such as age, sex, and radiation, or with clinical features including tumor size, multifocality, and metastasis. Additionally, there were no significant differences in tumor mutation burden, burden of copy number alterations, or mutational signatures by the presence of 22qDEL. Differential gene expression analysis of CTB and TCGA based on 22qDEL status revealed downregulation of most genes located on chromosome 22q and significant enrichment of immune-related biological pathways, particularly those involved in T cell and innate immunity, implicating a role of 22qDEL in immune dysregulation in PTC. Our study indicates that while 22qDEL is not a primary driver of PTC, but instead acts as a crucial co-factor with *RAS* point mutations to potentially alter the immune landscape at the transcriptomic level. Our findings underscore the importance of understanding the role of recurrent SCNAs in cancer, highlighting the co-occurrence of 22qDEL with a subset of PTC driver mutations that in turn could guide new therapeutic approaches.

Session Title: Cancer Poster Wednesday Session

Board 8013W: Charting Cancer Evolution Landscape on the Coordinate of Somatic Mutations by Single-cell NanoSeq

Authors:

Y. Niu, M. Niu, Y. Zhang, J. Luo, J. Sinson, A. Thompson, C. Zong; Baylor Coll. of Med., Houston, TX

Abstract:

Accurate detection of somatic mutations in single tumor cells is greatly desired as it allows us to quantify the single-cell mutation burden, which is crucial for understanding the level of intra-tumor heterogeneity (ITH). In addition, in comparison to the CNV-based approaches, the mutation-based evolutionary analysis allows a more straightforward determination of evolution distance. Although the mutation-based approach has been attempted with single-cell whole-genome sequencing, the error rate of 1-2 false positives per million bases in current methods

has constrained their application on mutation-based phylogenetic analysis. Here, we develop scNanoSeq chemistry which can efficiently filter out DNA damage and amplification-induced errors by duplex sequencing of both Watson and Crick strands. By using scNanoSeq, we profile the somatic mutations in 842 single cells from 21 human breast cancer samples. Our analysis accurately determines the mutation burdens in individual tumor cells, revealing a notable degree of inter-tumor variations (ranging from 963 - 24,578 somatic mutations per diploid genome). Next, we employ CellPhy algorithm to construct the mutation-based evolution phylogenetic trees. After adjusting for CNVs, we infer the molecular timing of the emergence of the major tumor population as well as that of each subclone. As a result, our observations unveil the following characteristic features of tumor evolution in breast cancer. Firstly, we observe the existence of a single dominant clone and the associated clonal sweep in almost all samples. Secondly, we observe a stem structure before the onset of the clonal sweep, which we define as the stem evolution structure. Thirdly, following the clonal sweep, we observe a rapid evolution that results in a few major subclones. Finally, we show that selection pressure within the dominant population rapidly drops to a baseline level, thereby leading to the attainment of effectively neutral evolution. In summary, the accurate characterization of somatic mutations in single tumor cells by scNanoSeq allows us to unveil novel insights into the dynamics of cancer evolution.

Session Title: Cancer Poster Wednesday Session

Board 8014W: Circadian Rhythm Genes in Chronic Myeloid Leukemia: An Untold Tale

Authors:

S. Alsaafin[1], D. Velayutham[1], P. Jithesh[2]; [1]Hamad Bun Khalifa Univ., Education City, Qatar, [2]Hamad Bin Khalifa Univ., Doha, Qatar

Abstract:

Background: Circadian rhythm and its regulating genes are essential in various physiological processes through the 24-hour oscillations that happen throughout the day in response to stimuli. Over 20 genes wereidentified as part of the “Clock Genes” (CGs) affecting all the tissues in the body. The main orchestrators of circadian rhythm are BMAL1and CLOCK genes, where they form a complex that binds to E-boxes of some target CGs to increase their expression, and these targets will in turn inhibit BMAL1 expression in a tightly regulated negative feedback loop. The disruption of these loops has been linked to cancer and other chronic conditions since they are involved in cell cycle regulation. Circadian rhythm research has been mainly done on solid

tumors, however, the research in hematological malignancies is not as mature despite making up to 10% of adult neoplasms. Here we hypothesized that CGs have significant involvement in hematological malignancies measuring, their alterations may aid in the diagnosis and prognosis prediction using Chronic Myeloid Leukemia (CML) as a model disease. Methods: Differential expression of CGs was done on CML studies from RNA expression data available in Gene Expression Omnibus (GEO), where only studies with newly diagnosed CML patients with blood or DNA samples were chosen. Analysis was done using R studio and DESeq2. Pathways analysis was done using PANTHER and Ingenuity Pathway Analysis (IPA) software. Results: Five studies were found to fit the criteria from manual search.CGs were separated into 3 tiers according to their proximity to their circadian rhythm: Tier I Circadian Rhythm genes (n=43), Tier IICircadian associated genes (n=184), and Tier III Cell cycle controlling genes (n=99). We found 20/43 Tier I genes diffrentailly expressed,includng the main oscillaters CLOCK, BMAL1 and TIM. Additionally, 67/184 genes from Tier II and 26/99 genes from Teir III were differentially epxpressed. The associated pathways with these dysregulated CGs were p53 pathway, Wnt signalling and Gonadotropin-releasing hormone receptor pathway. Conclusion: CGs that mainly control circadian rhythm are disrupted in CML and they are closely associated with main cancer pathogenesis pathways which suggests their contribution to the disease.

Session Title: Cancer Poster Wednesday Session

Board 8015W: Clinical characteristics and Lynch Syndrome Associated Variants in Taiwan study

Authors:

H-N. Chuang, C. Yi-Peng, K. Pei-Ying, L. Wan-Tzu, L. Yi-Jun, L. Szu-Chia, T. Hsin-Ju, C. Yen-Ju, J. Pei-Pei, H. Tzu-Hung, L. Ying-Cheng; Taichung Veterans Gen. Hosp., Taichung, Taiwan

Abstract:

Background: Lynch syndrome (LS), an autosomal dominant disorder characterized by increased risks of colorectal and endometrial cancers, is defined by pathogenic variants in mismatch repair (MMR) genes, including MLH1, MSH2, and MSH6. Data on the prevalence and associated cancer risks of LS in Han Chinese populations remain scarce. Objectives: This study aims to delineate the clinical impact of MMR gene variants within a cohort at a Taiwanese medical center, in collaboration with the Taiwan Precision Medicine Initiative (TPMI). Methods: We analyzed data from 89 individuals identified with MMR gene variants from 42,828 TPMI Participants: MLH1 (n=22, 25%), MSH2 (n=47, 53%), and MSH6 (n=20, 22%). The overall prevalence of MMR variants was calculated, and cancer incidence rates among carriers were assessed. Results: The

prevalence of MMR variants in the study population was 1 in 481. The distribution of MLH1, MSH2, and MSH6 variants were 24.7%, 52.8%, and 22.5%, respectively. Additionally, cumulative cancer incidences were 40.9% for pathogenic MLH1 carriers, 29.8% for MSH2, and 40% for MSH6. Among the 19 individuals who underwent colonoscopic screening, the prevalence of polyps was similar to that of the control group (adenoma detection rate: 32% vs. 26%, p=0.585). A meticulous analysis of the detected polyps in 7 participants, considering factors such as location, size, morphology, and pathological features, showed no significant differences compared to controls. Conclusion: The data underscore a significant cancer risk associated with LS-related MMR variants in the Taiwanese population. The apparent underdiagnosis of LS highlights the urgent need for enhanced surveillance and genetic counseling in this demographic. Our findings suggest that adjustments in the current screening protocols may be warranted to better identify and manage at-risk individuals

Session Title: Cancer Poster Wednesday Session

Board 8016W: Coding mutations of genitourinary development genes carry high risk for prostate cancer.

Authors:

J. Smith[1,2,3,4], A. L. Jones[5], W. D. Dupont[6,7,8]; [1]Dept. of Med., Div. of Genetic Med., Vanderbilt Univ. Med. Ctr.., Nashville, TN, [2]Med. Res. Service, Tennessee Valley Hlth.care System, Veteran's Admin.., Nashville, TN, [3]Vanderbilt-Ingram Cancer Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr.., Nashville, TN, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Dept. of Biostatistics, Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]Med. Res. Service, Tennessee Valley Hlth.care System, Veteran's Admin., Nashville, TN, [8]Vanderbilt-Ingram Cancer Ctr., Vanderbilt Univ. Med. Ctr.., Nashville, TN

Abstract:

Prostate cancer has the greatest heritable risk among the common cancer types. Segregation analyses of familial clustering best fit models of rare, high-risk alleles of age-dependent penetrance. Roughly five percent of cases meet criteria for hereditary prostate cancer with three or more affected close relatives. Considerable genetic heterogeneity is currently thought to underlie these families. Most remain unexplained by the limited number of established causal alleles. This study sought to discover rare but recurrent causal mutations under a case-control study design, gaining power by study of familial cases. Carrier frequencies of known 8q24 and *HOXB13* high-risk alleles are enriched eight- to sixteen-fold among familial cases. We investigated two independent familial case-control populations to discover high-risk causal

mutations, using genome-wide data of the Nashville Familial Prostate Cancer Study and of the International Consortium for Prostate Cancer Genetics. We sought further replication of observations within four independent biobanks: the Million Veteran Program, All of Us, the UK Biobank, and FINNGen. We detected coding mutations that were each associated with high risk of prostate cancer within three genes sharing an unexpected commonality. They are each required for genitourinary development and have previously been established to cause related Mendelian disorders, although none were known to predispose to prostate cancer. One of the coding mutations reached genome-wide significance (meta-analysis OR = 2.59, *P* = 3E-09); carriers were observed among discovery population cases from the US, UK, Finland, Australia, and Germany, and the association was also well-supported by each of the biobanks. Although *HOXB13* is not currently established to cause a Mendelian urogenital disorder, mouse models of it too result in a defect of genitourinary development.

Session Title: Cancer Poster Wednesday Session

Board 8018W: Comprehensive Clinico-genetic Profiling Reveals Prognostic Significance of Cytogenetic Aberrations in Multiple Myeloma: Implications for Risk Stratification and Treatment Outcomes in an Indian Patient Cohort

Authors:

B. Kar, A. Thakur, S. Tulpule, N. Shukla; Kokilaben Dhirubhai Ambani Hosp. & Med. Res. Inst., Mumbai, India

Abstract:

Multiple myeloma (MM), the second most prevalent hematologic malignancy, accounts for 1% of cancer cases and 2% of cancer-related deaths globally. This study integrates cytogenetic data with comprehensive clinical profiling in an Indian patient cohort to enhance risk stratification and treatment optimization in MM. We analyzed 263 suspected plasma cell dyscrasia patients using karyotyping and FISH, combined with immunological, hematological, biochemical, and imaging data to generate detailed clinico-genetic profiles. Among these patients, 73.8% had active MM, with cytogenetic aberrations detected in 40.3%, while 33.5% were cytogenetically negative. FISH identified key aberrations including del(13q14.2) (35.57%), gain(1q21-22) (28.87%), t(4;14) (8.76%), and del(17p13.1) (5.15%). Patients were stratified into high-risk (54.64%), intermediate-risk (35.57%), and standard-risk (9.79%) groups. IgG Kappa (32.4%) was the most common subtype, while IgA Kappa presented the highest cytogenetic risk (2.19%).

Treatment regimens varied, with VRd being most prevalent (18.87%, mean risk score: 1.7), and lower risk scores observed in transplant recipients (1.47) versus non-recipients (1.48). VCd without transplant had a mean risk score of 2.17, and VTD was used in lower-risk scores for both transplanted (1.89) and non-transplanted patients (4.4). Treatment response analysis revealed relapse in 8.25% (mean risk score: 2.45), remission in 14.95%, and death in 7.22% (mean risk score: 1.85). Cytogenetically negative patients showed longer mean overall survival (OS) (122±4 months) compared to positive patients (94±9 months), with similar progression-free survival (PFS) between negative (68±4 months) and positive (71±5 months) groups. Second-line treatment with ASCT significantly extended survival (mean: 77±8 months), while first-line treatment offered longer PFS (62 months). The hazard ratio (HR) for cytogenetic-negative status was 0.79, indicating a 21% lower risk of death or progression compared to the cytogenetically positive patients. Significant associations were found between cytogenetic risk and myeloma subtype (χ^2=28.796, $p<0.05$), R-ISS stratification (χ^2=55.371, $p<0.001$), CRAB features (χ^2=30.362, $p=0.002$), and treatment response (χ^2=36.571, $p=0.006$). This study highlights the critical role of integrating cytogenetic data into clinical practice for improving risk stratification and treatment strategies, leading to better patient outcomes. Our findings from an Indian cohort underscore the prognostic value of cytogenetic aberrations, advancing personalized medicine in MM management.

Session Title: Cancer Poster Wednesday Session

Board 8019W: Conquering SPP1-driven resistance of solid tumors to CAR T cell therapy

Authors:

H. Natri[1], S. Gholamin[2,3], Y. Zhao[4], S. Xu[2], M. Aftabizadeh[2], B. Comin-Anduix[5], S. Saravanakumar[2], C. Masia[2], R. Wong[2], L. Peter[1], M-i. Chung[1], E. D. Mee[1], B. Aguilar[2], R. Starr[2], D. Y. Torrejon[6], A. Kalbasi[7], D. Alizadeh[2], X. Wu[4], A. Ribas[5], S. Forman[2], B. Badie[8], N. Banovich, E[1], C. Brown, E[2]; [1]Translational Genomics Res. Inst., Phoenix, AZ, [2]City of Hope Beckman Res. Inst. and Med. Ctr., Duarte, CA, [3]California Inst. of Technology, Div. of Biology and Biological Engineering, Pasadena, CA, [4]Dept. of Computational and Quantitative Med. within Beckman Res. Inst., City of Hope, Duarte, CA, [5]Div. of Surgical Oncology, Dept. of Surgery, UCLA, Los Angeles, CA, [6]Div. of Hematology-Oncology, Dept. of Med., UCLA, Los Angeles, CA, [7]Div. of Radiation Oncology, Dept. of Med., Stanford, Palo Alto, CA, [8]Div. of Neurosurgery, City of Hope, Duarte, CA

Abstract:

Glioblastoma multiforme (GBM) is an aggressive and lethal form of brain cancer that is highly resistant to traditional therapies. Chimeric antigen receptor (CAR) T cell therapy has emerged as a promising alternative for the treatment of GBM and other cancers. Still, CAR T therapy faces challenges and limitations in the treatment of solid tumors. The suppressive tumor microenvironment (TME), characterized by complex interactions among immune and stromal cells, has emerged as a key contributor to resistance to CAR T cell therapy and other immunotherapies. Despite their abundance and diversity in the TME, the involvement of macrophages in modulating responses to CAR T cell therapies remains poorly understood.
We have previously identified and described SPP1+ macrophages as potential drivers of resistance to an IL13Rα2-targeted CAR T cell therapy in patients with recurrent or refractory high-grade glioma (ASHG 2023). Here, we have further investigated and targeted the cellular pathways producing the immunosuppressive TME responsible for tumor resistance. Through an integrative analysis of multimodal single-cell RNA-seq data from GBM and an interferon-signaling deficient syngeneic mouse model, we demonstrate the role of congruent suppressive pathways in mediating resistance to CAR T therapy across solid tumors. Specific ligand-receptor interactions and immunometabolism pathways were upregulated in SPP1+ macrophages, elucidating the mechanisms underlying the suppressive effects.
To evaluate its functional impact in promoting resistance to CAR T therapy, we targeted SPP1 with an antibody alongside CAR T therapy in syngeneic mouse tumors. Tumors primed with anti-SPP1 prior to CAR T administration showed a significant reduction in size and were enriched for infiltrating immune cells, demonstrating that SPP1 blockade shapes the TME and reverses the suppressive effects, prolonging survival in syngeneic resistance models. These findings illuminate the central role of SPP1+ macrophages in fueling the immunosuppressive TME and driving solid tumor resistance to CAR T cell therapies. Targeting SPP1 may serve as a universal strategy for altering immune dynamics in solid tumors, mitigating resistance to CAR T and other immunotherapies.

Session Title: Cancer Poster Wednesday Session

Board 8020W: Consideration about Japanese variants in *TP53* based on the ToMMo-54KJPN: enhanced prediction accuracy with AlphaMissense

Authors:

T. Mariya, M. Idogawa, H. Nakase, T. Tokino, T. Saito, A. Sakurai; Sapporo Med. Univ., Sapporo, Japan

Abstract:

Our study aimed to assess the frequency of *TP53* variants in the general Japanese population using the ToMMo-54KJPN dataset and to develop a comprehensive method for identifying pathogenic variants through a combination of *in silico* and *in vitro* analyses. We annotated *TP53* variants in the 54KJPN using ClinVar dataset to confirm the prevalence of pathogenic variants. Additionally, we examined the concordance between *in vivo* and *in vitro* functional analyses and ClinVar classification for 2,308 *TP53* missense variants. Functional analyses were conducted by the recommended methods in the ClinGen *TP53* Expert Panel, as well as a novel *in silico* analysis method of AlphaMissense (AM). The allele frequency of *TP53* pathogenic variants registered in ClinVar within 54KJPN was 1.10×10^-4. Remarkably, all these variants were also classified as pathogenic by AM analysis. Even in the whole *TP53* pathogenic variants of ClinVar dataset, AM showed the highest diagnostic accuracy among all *in silico/in vitro* analyses with a sensitivity of 95.1% and a specificity of 91.3%. Furthermore, the combination of AM with specific *in vitro* analysis score resulted in even higher accuracy, with a sensitivity of 97.8% and specificity of 99.0%. Application of this combination method to evaluate 30 VUS in 54KJPN yielded positive results for 43.3%, with an allele frequency of 1.75×10^-3. There is a population among carriers of *TP53* variants of VUS for whom careful management is desired, and it is anticipated that further accumulation of clinical data will be beneficial in the future.

Session Title: Cancer Poster Wednesday Session

Board 8021W: Data Science Approaches to Understanding Hereditary Cancer Syndromes in Colombia

Authors:

D. Cardona[1], J. P. Valencia-Arango[1], J. P. Gallo[1], M. Naranjo[2], N. Galleo[3], H. VELASCO[4]; [1]Data Sci. Dept., BioSci. Ctr. – Ayudas Diagnósticas SURA. Medellín, Colombia., Medellín, Colombia, [2]Ayudas Diagnósticas SURA, Medellin, Colombia, [3]Omic Sci. Ctr. - Ayudas Diagnósticas Sura. Medellín, Colombia, Medellín, Colombia, [4]Personalized medicine Group - BioSci. Ctr., Ayudas Diagnósticas Sura. Medellín, Colombia, MEDELLIN, Colombia

Abstract:

The global burden of cancer, influenced by both acquired and hereditary mutations, necessitates comprehensive research efforts. However, the underrepresentation of Latino or mixed populations in existing studies underscores a gap in data science research. Initiatives such as the Omic Science Center's program aim to address this gap by integrating genetic and clinical data, advancing our understanding of hereditary cancer patterns. This study, focusing on 1377

Colombian patients, highlights the pivotal role of data science in driving precision medicine approaches for cancer management.The study utilized observational healthcare databases covering four regions and 11 cities in Colombia. Genetic data, extracted from PDF reports, focused on SURA affiliates referred for hereditary cancer testing between October 2019 and November 2021. Next-Generation Sequencing was used to analyze variants in 30 genes aligned with NCCN guidelines. Data processing involved text extraction with Python and R, and deployment of a data science tool on Azure infrastructure for cross-referencing genetic, clinical, and sociodemographic data.Among the 1377 patients included, predominantly women (92.81%), a significant portion (63%) were from the northwestern region. Most patients (40.37%) were aged between 31 and 44 years, with 95.35% having personal histories of cancer, primarily breast cancer (75.86%), followed by colorectal (12.8%) and ovarian cancer (5.33%). Hereditary cancer testing revealed 145 positive results and 587 uncertain outcomes, with positive patients averaging 46 years old. Positive rates varied across age groups and geographical regions, with higher positivity observed in patients aged 31-44 and over 50, particularly in the northeast and central regions. Additionally, among the 145 patients with positive results, 42.6% had at least one variant of uncertain significance, with a significant proportion (95.9%) having a personal history of cancer.Our study provides valuable insights into the genetic characteristics of hereditary cancer patients in Latin America, addressing gaps in scientific evidence. Findings on the prevalence of breast, colorectal, and ovarian cancers align with known hereditary contributions to these cancer types. Furthermore, our data science implementation shows promise for variant reclassification and tailored interventions, as demonstrated by regional dynamics observed, particularly in the Northern region.

Session Title: Cancer Poster Wednesday Session

Board 8022W: Deciphering the genetic mechanism of sex differences within multiple myeloma

Authors:

A. Opakunle, R. S. Huang; Univ. of Minnesota-Twin Cities, Minneapolis, MN

Abstract:

Multiple Myeloma (MM), a blood cancer, exhibits notable sex differences in both incidence and mortality rates, with men suffering rates approximately 1.5 times higher than women. It has been established that genetics plays a role in MM disease susceptibility and survival outcomes. Yet whether and if so, how genetics may affect observed sex differences in MM is unknown. To address this knowledge gap, we set out to investigate and define the sex-biased genetic risk

factors and sex-differentiated regulatory mechanisms contributing to observed sex differences in MM incidence. By leveraging the Multiple Myeloma Research Foundation's CoMMpass Study (MMRF-COMMPASS), one of the most extensive and comprehensive data repositories containing genetic and sex-related clinical data on MM, we performed a series of sex-aware analyses to interrogate the sex-specific genetic effects on MM incidences and outcomes. Specifically, a sex-stratified genome-wide association study (ssGWAS) using germline DNA, and somatic analysis focused on sex differences in somatic mutation patterns. We identified 255 single nucleotide polymorphisms (SNPs) that contribute uniquely to MM risk in men and women. Our sex-stratified somatic mutation analysis revealed 15 genes that show significant sex differences in somatic mutation patterns, as reflected by tumor gene mutation frequency. Additionally, 80 out of 255 SNPs showed sex-biased association with gene expression changes through sex-stratified expression quantitative trait loci (eQTL) analysis within GTEx whole blood samples. Notably, we identified a female bias in MM risk allele (rs114916116), located in the chr5_178143635_G_A_b38 region of the G-protein coupled receptor kinase 6 (GRK6) gene within GTEx whole blood samples. We observed a corresponding female bias in this gene's expression. These findings suggest a potential sex-specific influence on the genetic and gene expression patterns. GRK6 is a potential therapeutic target in MM due to its significant role in mediating the chemotactic responses of T and B-lymphocytes. Our findings shed light on the sex-differentiated genetic mechanisms underlying MM development. By integrating various analyses and datasets, our research seeks to understand the genetic factors contributing to MM risk, focusing on sex differences. In addition, sex-aware analyses fill a critical gap in personalized medicine for reducing cancer disparities and may uncover new sex-specific targets for developing personalized therapies.

Session Title: Cancer Poster Wednesday Session

Board 8023W: Differential gene expression at diagnosis in high hyperdiploid pediatric acute lymphoblastic leukemia with negative and positive MRD at the end of induction treatment phase

Authors:

S. Lobo-ALves[1,2,3], L. Oliveira[1,3], S. Bispo[1,3], G. Prezia[1,2,3], R. Rosati[1,2,3]; [1]Inst. de Pesquisa Pelé Pequeno Príncipe, Curitiba, Brazil, [2]Faculdades Pequeno Príncipe, Curitiba, Brazil, [3]Natl. Sci. and Technology Inst. for Children's Cancer Biology and Pediatric Oncology - INCT BioOncoPed, Porto Alegre, Brazil

Abstract:

Introduction: The high hyperdiploid (HHD) is a molecular subtype of acute lymphoblastic leukemia (ALL) with favorable initial prognosis. However, a portion of HHD ALL patients respond worse to treatment than others, once they remain with detectable minimal residual disease (MRD) in the induction treatment phase. It highlights the need for additional investigation of the existing heterogeneity in treatment response. Here, we focus on detecting differentially expressed genes (DEGs) in order to characterize the complete and incomplete responders of the induction treatment, also longing to identify a treatment response biomarker that can be detected at the diagnosis and perhaps become an additional risk stratification component. **Methods**: In this pilot study we compared gene expression of 9 HHD ALL patients splitted in two subgroups; 4 patients with undetectable MRD at the end of induction (complete responders) versus 5 patients with detectable MRD at the same treatment point (incomplete responders). We used Illumina NovaSeq to sequence poly-A RNA from leukocytes isolated from diagnosis bone marrow or peripheral blood (both with at least 70% of blasts). We conducted differential gene expression analysis and additional enrichment analysis in order to categorize the DEGs and uncover their common functions. **Results**: We detected 226 DEGs, 118 upregulated and 108 downregulated in the diagnosis samples of incomplete responders in relation to complete responders at the end of the induction phase. The downregulated genes were mainly categorized in immune response and cytokines pathways. Among the upregulated genes there were few shared functions, possibly because a large part of the genes are poorly explored. Moreover, GSEA analysis showed the enrichment of important pathways in the development of cancer, such as the *Notch signaling pathway* in the complete responders group. **Conclusion**: We detected relevant DEGS between complete and incomplete treatment responders HHD ALL patients. These genes can be further evaluated as early biomarkers for treatment response and better stratification of HHD pediatric ALL. Furthermore, they can also be investigated as key players in the disease maintenance and progression, being candidates to become therapeutic targets.

Session Title: Cancer Poster Wednesday Session

Board 8024W: DNA Methylation and Glucose Metabolism Gene Variants as Predictive Markers for Pancreatic Cancer Risk

Authors:

V. Aaren[1], R. Uppala[2]; [1]Andhra Univ., Visakhapatnam, India, [2]Green Cross Molecular Lab., Ahmedabad, India

Abstract:

Background. Pancreatic cancer (PC) is among the most aggressive and difficult malignancies to manage, marked by poor survival rates and scarce treatment options.
PC lack of early detection tools or markers often leads to misdiagnosis or missed diagnosis, resulting in advanced-stage detection and poorer prognoses due to late-emerging symptoms. Glucose metabolism is pivotal in PC, profoundly impacting tumor growth, progression, and treatment responses. Aberrant glucose utilization is central to the disease's development. **Objective:** This study aims to identify PC-related genes using genome-wide methylation analysis and develop an early prediction model based on potential biomarkers. M**ethods.** We conducted genome-wide methylation analysis on blood DNA samples from 12 PC cases and an equal control using the Illumina HumanMethylation 850 BeadChip array. This array quantitatively measures over 850,000 methylation sites across the genome at single-nucleotide resolution. Our analysis involved detailed bioinformatic and statistical methods, including false discovery rate (FDR) and Pathway analysis. **Results:** The study identified 8 CpG sites with significant methylation variations (FDR $P < 0.001$). Four glucose metabolism genes showed hypomethylation, while four were hypermethylated in PC patients. The most significant associations were found in *SDHA, P4HA1, GALK1,* and *BRCA1,* followed by *LRP6, PGM1, PRKAG2,* and *ESR1* genes, implicated in PC due to their dysregulation. These genes play various roles, such as tumor suppression (BRCA1), hormone-related mechanisms (ESR1), and metabolism pathways (GALK1 and PGM1). Additionally, they contribute to abnormal cell growth, invasion, and metastasis (P4HA1), as well as cell survival and growth under stress conditions (PRKAG2). SDHA is involved in the citric acid cycle and electron transport chain. These genes are pivotal in PC development, potentially linked to pathways including breast cancer, galactose metabolism, biosynthesis of nucleotide sugars, and amino sugar and nucleotide sugar metabolism. **Conclusions:** In summary, the crucial role of glucose metabolism genes in PC development underscores their potential as novel therapeutic targets. Moreover, this study's findings significantly advance our understanding of the glucose metabolism in PC, suggesting that peripheral blood methylation changes could serve as promising diagnostic biomarkers for the disease in the future.

Session Title: Cancer Poster Wednesday Session

Board 8025W: Effective batch effect correction in large-scale TruSeq and capture-based RNA sequencing

Authors:

C. McCabe[1], N. Fadra[1], D. O'Brien[1], H. Yan[1], E. Braggio[2], S. Slager[1]; [1]Mayo Clinic, Rochester, MN, [2]Mayo Clinic, Phoenix, AZ

Abstract:

Large omics datasets, analyzed using robust bioinformatics tools, increase the reproducibility of results and enable the discovery of gene expression patterns that may be undetected in smaller datasets. However, in these large datasets, samples are often collected over many years and sequenced at different centers. This introduces significant technical variation that could lead to bias and confound results. Careful data pre-processing, such as effective normalization and batch effect correction techniques, is necessary to mitigate these factors. We developed a set of recommendations for pre-processing of RNA-seq data to correct for these batch effects. When combining samples from multiple batches, particularly when different sequencing platforms are used, it is crucial to account for sequencing depth, batch effects, and technical variability. We addressed this variability by determining a threshold for low expression based on variation between batches. Genes not meeting the threshold were removed from each batch individually. While this approach does reduce the number of genes in the original dataset, it effectively removes noise and improves the reliability of the results. To adjust for batch effects, we utilized ComBat-seq, including phenotypes (such as race) that were unequally distributed between batches as covariates in the model. We applied our approach to a cohort of 676 Chronic Lymphocytic Leukemia (CLL) patients with RNA-seq data obtained in three batches using both TruSeq poly(A) and custom-capture from total RNA sequencing. The resulting corrected gene counts were used for downstream analyses. The batch-corrected gene counts were validated using multiple differential gene expression analyses, as well as using 26 individuals sequenced with two RNA sequencing protocol to confirm our pre-processing approach. These results and corresponding visual representations of the batch corrected data confirmed that we had reduced batch effects while preserving biological differences. Further, among the 26 duplicate individuals, we compared gene expression values of batch-corrected genes across the two RNA-seq protocols and observed high correlation (R=0.97). Our work outlines a tailored, effective approach that enables us to integrate RNA-seq samples sequenced using different chemistries, confounded by multiple variables, into one large, cohesive, and robust dataset. This integrated dataset can be utilized for accurate gene expression-based subtyping and biomarker discovery.

Session Title: Cancer Poster Wednesday Session

Board 8026W: Evaluating CRLF2 rearrangement status along with relevant diagnostic impact in Ph-Like acute lymphoblastic leukemia

Authors:

R. Lewis, B. Hong; Univ. of Utah & ARUP Lab., Salt Lake City, UT

Abstract:

Philadelphia chromosome-like acute lymphoblastic leukemia (Ph-like ALL) is a high-risk subtype with molecular features similar to Ph-positive ALL but without the BCR::ABL1 fusion. It involves rearrangements of ABL class genes (ABL1, ABL2, CSF1R, PDGFRB) and JAK-STAT pathway genes (CRLF2, JAK2, EPOR). Identifying these rearrangements is crucial for diagnosis, prognosis, and treatment. CRLF2 rearrangements, the most common genetic alterations in Ph-like ALL, occur in 5-8% of pediatric and 10-15% of adult ALL cases. Clinically, CRLF2 rearrangements are assessed using break-apart (BAP) fluorescent in-situ hybridization (FISH) probes targeting the proximal and distal regions of CRLF2. This method effectively identifies CRLF2 rearrangements but does not specify the rearrangement partner. Common CRLF2 rearrangement partners include P2RY8 and IGH. Some studies indicate that P2RY8 is five times more prevalent than IGH, while others suggest it is half as prevalent. Despite the assumption that detected CRLF2 rearrangements are P2RY8::CRLF2, accurately identifying the CRLF2 rearrangement partner is crucial due to distinct prognostic outcomes associated with different partners. This study evaluated internal testing data from a national reference laboratory to determine the prevalence of IGH::CRLF2 and P2RY8::CRLF2 rearrangements, aiming to refine testing strategies and follow-up testing for better characterization. We conducted a batch search of FISH cases using CRLF2 BAP probe in a seven year period, identifying cases positive for a CRLF2 rearrangement. Among these, we determined which had confirmatory testing via FISH (for IGH) or microarray (for P2RY8) and evaluated the age and sex of patients positive for these rearrangements. Out of ~1,700 cases, ~110 (6.5%) cases were positive for a CRLF2 rearrangement. Of the ~110 CRLF2-positive cases, only one-quarter had confirmatory testing. FISH confirmed IGH::CRLF2 in 17% and microarray confirmed P2RY8::CRLF2 in 9%. There was no significant gender difference in CRLF2 rearrangements or their specific types. IGH::CRLF2 rearrangements were more prevalent in younger males (median 24.3 years vs. 39.25 years for females). Additionally, IGH::CRLF2 rearrangements were more common in older individuals compared to P2RY8::CRLF2 rearrangements (median 25 years vs. 8 years). Comprehensive confirmatory testing for CRLF2 rearrangements using FISH and microarray is essential for accurate diagnosis and effective treatment in Ph-like ALL. Differences in age and sex prevalence of CRLF2 rearrangements with IGH and P2RY8 should guide testing and monitoring strategies.

Session Title: Cancer Poster Wednesday Session

Board 8027W: Expanding our understanding of the genetics of mole count

Authors:

M. Law[1,2,3], G. Jayasinghe[1,2], G. Zhu[1], N. Pandeya[1,2], C. M. Olsen[1,2], S. D. Gordon[1], T. Nijsten[4], P. A. Lind[5,2,3], V. Bataille[6], M. Kayser[7], D. A. Mackey[8], G. W. Montgomery[9], N. G. Martin[1], S. E. Medland[1], D. C. Whiteman[1], D. L. Duffy[1,10], S. MacGregor[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Faculty of Med., Univ. of Queensland, Brisbane, Australia, [3]Faculty of Hlth., Queensland Univ. of Technology, Brisbane, Australia, [4]Erasmus MC Univ. Med. Ctr. Rotterdam, Rotterdam, Netherlands, [5]QIMR Berghofer Med. Res. Inst., Herston, Australia, [6]Hemel Hempstead Gen. Hosp., Hemel Hempstead, Hertfordshire, United Kingdom, [7]Erasmus Univ. Med. Ctr. Rotterdam, Rotterdam, Netherlands, [8]Lions Eye Inst., Nedlands, Australia, [9]Univ. of Queensland, St Lucia, Australia, [10]Frazer Inst., Univ. of Queensland, Brisbane, Australia

Abstract:

Melanocytic nevi, also called moles, are clonal expansions of melanocytes, the pigment-producing cells of the skin. Melanoma is a cancer arising from melanocytes, and having a large number of moles is one of its strongest risk factors. Further, melanoma can develop *de novo* or from an existing mole. As a result, as well as increasing our understanding of the etiology of moles, identifying additional genes that influence nevus count will provide new insights into melanoma development.

Here, we report the largest ever GWAS meta-analysis for nevus count, combining previously published data with two new GWAS from QSkin Sun and Health Study, and a third GWAS from the Australian Genetics of Depression Study. The new datasets include ~33,000 individuals with self-reported whole-body nevus count, giving a final sample of over 85,000.

This GWAS meta-analysis identified 28 genome-wide significant nevus count loci, of which 23 were not previously reported for nevus count alone. Post-GWAS gene-mapping of these loci highlights new candidate genes influencing mole count, including support for the *SIKE1* gene, which is involved in immune response. Interestingly, pathway analysis indicates that newly identified risk genes are also associated with cancers that do not have a pigmentation component (e.g., breast, prostate and glioma). This indicates that nevus count analysis can reveal novel pathways influencing melanoma risk that are independent of established factors such as skin pigmentation.

Session Title: Cancer Poster Wednesday Session

Board 8028W: Expanding the central nervous system tumor spectrum of *POT1* germline variants

Authors:

K. Wadt, E. Nielsen, A. Jelsig, J. Foss-Skiftesvik, Z. Lohse, T. Hansen; Univ. Hosp. of Copenhagen, RigsHosp.et, Copenhagen, Denmark

Abstract:

Pathogenic germline variants in the *POT1* gene are associated with increased susceptibility for development of a heterogenous spectrum of neoplasms such as cutaneous melanoma (CM), glioma and hematological malignancies. However, it is only the link to CM that has been consistently observed in previous reports and hitherto leading to clinical surveillance in POT1 variant carriers.The *POT1* tumor predisposition syndrome is rare, but knowledge on the role of germline variants in *POT1* in tumorigenesis have rapidly increased during the past two decades. POT1, Protection of Telomerase 1, located on chromosome 7q31.33, was first described in 2001, as a component of the six-protein Shelterin complex. POT1 consists of four domains; the OB1 and OB2 which interact with ssDNA and the OB3 and HJRL domains, which binds the nearby protein TPP1. Consequently, POT1 and TPP1 affects regulation of telomere length by performing chromosome-end maintenance and modulating telomerase access to 3'ssDNA. While initially only short telomeres were thought to be associated with neoplasms, studies of telomere lengths in families with heterozygous germline variants in *POT1* have established a role of telomeric elongation in tumorigenesis. To our knowledge previously only seventeen patients with a germline variant in *POT1* and CNS tumor have been described in the literature. We describe the phenotype in six unrelated Caucasian families with germline *POT1* variants and highlight the possible association between germline variants in *POT1*, telomere dysregulation and a predisposition to central nervous system (CNS) tumors. Our findings support an association between pathogenic *POT1* germline variants and increased risk of CNS tumors, in particular oligodendroglioma and astrocytoma. Furthermore, our observations highlight the impact of missense variants and telomere elongation in POT1- tumorigenesis. Based on our results combined with those of the existing literature, we find supplementary personalized CNS surveillance highly relevant in certain families based on ACMG classification, family history of glioma, *POT1*-variant type with specific attention to missense variants, and telomere length assessment. Still, further studies on germline variants, telomere length and somatic alterations in CNS tumors are warranted to further elucidate the potential role of *POT1* as a risk gene in gliomagenesis and aid future management and treatment strategies of glioma prone families with germline variants in POT1.

Session Title: Cancer Poster Wednesday Session

Board 8029W: Exploring the yield of germline genetic testing on skin biopsy samples

Authors:

A. M. Murad[1,2], M. Accardo[2], M. F. Jacobs[1,2], E. Koeppe[2], S. C. Quinonez[2,3]; [1]Rogel Cancer Ctr., Univ. of Michigan, Ann Arbor, MI, [2]Dept. of Internal Med., Univ. of Michigan, Ann Arbor, MI, [3]Dept. of Pediatrics, Univ. of Michigan, Ann Arbor, MI

Abstract:

Approximately 5-10% of cancer is due to inherited susceptibility. Guidelines recommend germline genetic testing in individuals with cancer based on factors including cancer type, age of diagnosis and family history of cancer. Guidelines also recommend genetic testing in individuals with a potential germline pathogenic variant detected on tumor sequencing. Sample selection for germline analysis in individuals with a hematologic malignancy requires special consideration, as blood is equivalent to tumor tissue. Additionally, for individuals who have had allogenic hematopoietic stem cell transplant (HSCT), blood is not representative of their germline DNA, making the preferred sample DNA from cultured fibroblasts. Skin biopsies are an easy procedure but are more invasive than blood draw. Recommendation of germline genetic testing in this population may require additional consideration. We performed retrospective chart review of patients who had skin biopsy for germline genetic testing in the University of Michigan Cancer Genetics Clinic to evaluate the yield of testing in this population. Between 1/2017-1/2024, 85 patients had skin biopsy for germline genetic testing. Eleven patients were excluded due to sample failure, when testing via cultured fibroblasts was to determine if a mosaic result detected in blood was germline or when positive findings in cultured fibroblasts were later determined to be somatic in etiology or due to technical interference. All patients had a history of a hematological malignancy/disease or HSCT. Fifty-six (76%) patients had multiple primary cancer diagnoses. Seventy-one (96%) patients had germline genetic testing via a multi-gene panel, one had whole exome sequencing, and two had testing for a known familial variant or known somatic variant only. Forty (54%) patients had negative genetic testing, 10 (14%) had a pathogenic or likely pathogenic (P/LP) variant, and 24 (32%) had a variant of uncertain significance. Of those with a P/LP variant, 9 (90%) met clinical criteria for genetic testing, either due to a potential germline pathogenic variant detected on tumor sequencing (6, 60%) or based on personal and/or family history of cancer (3, 30%). P/LP variants were detected in *DDX41* (3), *BRCA2* (2), *CHEK2* (2), *ATM* (1), *FANCD2* (1), *GATA2* (1), *TSC2* (1). One patient had two pathogenic variants identified. Overall, the diagnostic yield of genetic testing for inherited cancer susceptibility syndromes via cultured fibroblasts is similar to the yield observed in patients tested via blood, buccal swab, or saliva. As anticipated, the yield of germline analysis is highest in individuals meeting clinical criteria for genetic testing.

Session Title: Cancer Poster Wednesday Session

Board 8030W: Generative AI impact on protein stability prediction in breast cancer genes

Authors:

R. Gnanaolivu, S. Hart; Mayo Clinic, Rochester, MN

Abstract:

INTRODUCTION: The loss-of-function (LOF) classification of most missense variants in tumor suppressor breast cancer genes *BRCA1, BRCA2, PALB2*, and *RAD51C* remains unclassified and confounds clinical actionability. Classifying these variants is challenging due to their rarity, leading clinicians to rely on *in silico* predictive methods. Protein stability changes are associated with function, making stability predictors valuable. Stability predictions upon missense variant perturbations require high-resolution protein structures. This study explores using generative AI to predict high-resolution protein structures, which can then be analyzed with *in silico* stability prediction methods to assess LOF activity in ordered regions of the protein. This study also determines the best *in silico* method to predict LOF activity in these genes. Functional classifications from homologous recombination DNA repair assays and variant classifications from the ClinVar database provide a reliable dataset for evaluating the performance of these *in silico* prediction methods. **METHODS:** Using root mean square deviation as a measure of similarity, we evaluated the predictive performance of AlphaFold2 (AF2) and ESMFold in predicting complex and subunit structures of *BRCA1*, *BRCA2, RAD51C*, and *PALB2*. Protein stability upon missense variants was predicted using generative AI produced structures and analyzed with FoldX, Rosetta, and DDGun3D. These predictions were compared with stability predictions from experimentally-derived structures. The protein stability predictions were also compared to those obtained from dedicated *in silico* missense prediction methods. **RESULTS:** AF2 structures of the BRCT domain of *BRCA1* and the DNA binding domain of *BRCA2* analyzed with FoldX predicts LOF significantly better than experimentally-derived structures with an AUC=0.861 (95% CI:0.858-0.863) and AUC=0.836 (95% CI:0.8322-0.841), compared to AUC=0.847 (95% CI:0.844-0.850) and AUC=0.830 (95% CI:0.821-0.8320) respectively. However, protein stability does not predict LOF better than dedicated *in silico* missense predictors. Overall, we find AlphaMissense ranks the best in predicting LOF activity in these breast cancer genes with an average AUC=0.890 (95% CI 0.886-0.895) compared to all other *in silico* missense predictors present in the dbNSFP database. **CONCLUSIONS:** The study reveals that generative AI protein predicted structures can outperform experimentally-derived structures in evaluating function. It also highlights the

predictive performance of AlphaMissense as the premier *in silico* missense prediction method for these breast cancer genes.

Session Title: Cancer Poster Wednesday Session

Board 8031W: Genetic Ancestry and Risk of Thyroid-Related Adverse Events Among Lung Cancer Patients Receiving Immune Checkpoint Inhibitors.

Authors:

E. Cardenas[1], P. Middha[1], Z. Quandt[1], R. Thummalapalli[2], Princess Margaret Lung Group, M. A. Gubens[1], C. M. Lovly[3], G. Liu[4], A. J. Schoenfeld[5], M. Aldrich[6], E. Ziv[7]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Dept. of Med., Mem. Sloan Kettering Cancer Ctr., New York, NY, [3]Dept. of Med., Div. of Hematology and Oncology, Vanderbilt Univ. Med. Ctr. and Vanderbilt Ingram Cancer Ctr., Nashville, TN, [4]Princess Margaret Cancer Ctr., Temerty Sch. of Med., Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto,, Toronto, ON, Canada, [5]Thoracic Oncology Service, Mem. Sloan Kettering Cancer Ctr., New York, NY, [6]Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]UCSF, San Francisco, CA

Abstract:

Immune checkpoint inhibitors (ICIs), have been rapidly adopted into the standard of treatment for many cancers. Due to enhanced immune activation, approximately 10% of patients treated with ICIs develop thyroid immune-related adverse events (irAEs). Prior studies have demonstrated that African American populations have a lower risk of spontaneous autoimmune hypothyroidism. However, there are no studies on the effect of race or genetic ancestry on thyroid irAEs. Using genetic and clinical data from 1090 patients with non-small cell lung cancer (NSCLC) treated with ICIs, we evaluated the role of ancestry in association with thyroid irAEs. Our study comprises 1090 NSCLC patients who received ICI therapy across four institutions between 2009 and 2023. Genotyping was conducted using the Affymetrix Precision Medicine Array, and imputation was performed using the 1000 Genomes Reference. Clinical data were curated by manual review of electronic health records. Genetic ancestry was inferred using principal components (PC) analysis from the full genome-wide data. PC 1-6 results were scaled from 0 to 1, 0 as the lowest value and 1 as the highest value for each PC. We conducted an unconditional logistic regression model to evaluate the association between ancestry (PCs 1-6) and incidence of thyroid irAEs. Analyses were adjusted for age at diagnosis, sex, type of ICI therapy, study site, and with and without polygenic risk score (PRS) for hypothyroidism. We used a previously published (PMID: 34099659) PRS for hypothyroidism. Among 1090

participants, 132 (12.1%) had thyroid irAE. In models adjusting for age, sex, study site, and type of ICI therapy, we found higher values of PC1 (PC1 measures genetic ancestry along the Africa vs out of Africa cline) associated with a lower risk of Thyroid irAEs with an odds ratio (OR) of 0.26 (95% confidence interval [CI]= 0.07-0.71, p=0.02). PRS for hypothyroidism strongly predicted thyroid irAEs with an OR per standard deviation of 1.86 (95% CI: 1.47-2.37, p=4x10-7). However, adding PRS to the model did not substantially change the effect of ancestry with an OR for PC1 of 0.28 (95% confidence intervals [CI]= 0.07-0.77, p=0.03) in the PRS-adjusted model. Of patients treated with ICI, higher is associated with protection from thyroid irAEs. These results are consistent with the lower rate of spontaneous hypothyroidism among African American individuals compared to White individuals. Adjusting for PRS did not change the association with ancestry, suggesting that social and or environmental factors may underlie the lower risk of thyroid irAEs seen in lung cancer patients with higher African ancestry

Session Title: Cancer Poster Wednesday Session

Board 8033W: Genetic testing for solid tumors: A 59-gene panel versus A 1021-gene panel

Authors:

J. Ma, H. Peng, Z. Zhou, Y. Da, Q. Xiao, Y. Wan, Y. Yu, H. Zhang, X. Yuan, Z. Zhong, J. Chen, S. Zou, T. Gong, C. Zhang; The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China

Abstract:

Next generation sequencing (NGS) plays critical roles in guiding target therapy and immunotherapy for solid tumors. Numerous gene panels have been used for solid tumors by different laboratories worldwide. In this study, a total of 2356 and 950 tumor tissue samples were tested for solid tumors with the 59-gene panel and the 1021-gene panel, respectively, at our laboratory inside the hospital since May 2023. We have conducted a comprehensive analysis to compare these two panels, including the cost, turn-around time, number of somatic or germline gene alternations, types and frequencies of genetic alternations including single nucleotide variants (SNVs), copy number variants (CNVs) and structural variants (SVs). While both of the panels cover the genes for targeted drugs approved by FDA, the 1021 gene panel has expanded to provide several important clinical applications, including detection of TMB,MSI,HLA-I and other immune-related genes for assessment of immunotherapy, 38 fusion genes such as *ALK,ROS1,NTRK,NRG* and *FGFR2*, as well as 36 genes associated with the homologous recombination repair (HRR). For the 1021-gene panel, approximately 88% of the

950 tumor samples harbor genetic alternations that may guide target therapy and immunotherapy, ~7% carry genetic alternations without indications for target therapy and immunotherapy, and ~5% did not have any gene alternations. For the 950 tumor samples tested with the 1021-gene panel, the top 10 single nucleotide variants (SNVs) include Tp53 (62.21%), KRAS (22.63%), EGFR (21.37%), APC (16.84%), LRP1B (11.89%), ARID1A (11.05%), SMAD4 (9.37%), PIK3CA (9.16%), MLL2 (8.32%), MLL3 (8.11%); The top 10 copy number variants (CNVs) include MYC (17.68%), CDKN2B (11.79%), CDKN2A (8.63%), WRN (8.42%), EGFR (6.95%), JAK2 (5.05%), RB1 (4.84%), BCL2L1 (4.74%), ERBB2 (4.74%), MDM2 (4.74%); The top 10 structural variants (SVs) include ROS1:HHAT (1.79%), EML4:ALK (1.47%), CBLN2:EGFR (0.74%), CAD:ALK (0.42%), CCDC6:RET (0.32%), ETV6:DUSP16 (0.32%), EZR:ROS1 (0.32%), KLC1:ALK (0.32%), ALK:CAD (0.21%), CD74:ROS1 (0.21%). Moreover, the 1021 gene panel has been used for selection of candidate genes for studies of minimal residual disease (MRD). Our findings will help to design and optimize novel gene panels for solid tumors.

Session Title: Cancer Poster Wednesday Session

Board 8034W: Genome-wide association analysis of tumor markers in WHALE cohort with East Asian ancestries

Authors:

N. Xiang[1,2], T. Fu[1,2,3], S. SU[1,2], Y. Yang[4], L. Wang[1,2,3], G. Liao[5], l. du[6], J. Huang[1], Y. Lin[1,7]; [1]Dept. of Urology, Innovation Inst. for Integration of Med. and Engineering, West China Hosp., Sichuan Univ., Chengdu, China, [2]Frontiers Sci. Ctr. for Disease-related Molecular Network, West China Hosp., Sichuan Univ., Chengdu, China, [3]Key Lab. of Bio-Resource and Eco-Environment of Ministry of Ed., Coll. of Life Sci., Sichuan Univ., Chengdu, China, [4]Hlth.Management Ctr., Gen. Practice Med. Ctr., Innovation Inst. for Integration of Med. and Engineering, West China Hosp., Sichuan Univ., Chengdu, China, [5]State Key Lab. of Oral Diseases, Natl. Clinical Res. Ctr. for Oral Diseases, West China Hosp. of Stomatology, Sichuan Univ., Chengdu, China, [6]Chinese Evidence-Based Med. Ctr., West China Hosp., Sichuan Univ., Chengdu, China, [7]Program in Genetic Epidemiology and Statistical Genetics, Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Background: Tumor markers are valuable for predicting cancers but have limitations, particularly variability due to non-cancerous conditions. Recent studies have shown that tumor markers possess heritability, indicating a portion of their variability stems from genetic components. However, genetic studies about tumor markers have often obtained small sample

sizes and failed to capture East Asian ancestries. To combine clinical phenotype and genotype to enhance the precision of cancer prediction, we conducted genome-wide association studies (GWAS) on common tumor markers. **Method and Result:** We obtained genome-wide genotype data from a total of 7212 participants of the WHALE cohort with East Asian ancestries. SNPs failing to meet the quality control standards were removed using PLINK. GWAS was performed with REGENIE software package for five tumor markers: alpha fetoprotein (AFP), anti-thyroglobulin antibodies (TGAb), alkaline phosphatase (ALP), carcinoembryonic antigen (CEA) and carbohydrate antigen 15-3 (CA15-3). Lead SNPs were identified for each marker, with counts of 8, 2, 10, 16, and 24, respectively. Linkage disequilibrium score (LDSC) regression was applied to estimate the SNP heritability of these markers, which ranged from 0.028 (ALP) to 0.567 (AFP). Notably, top significant genes associated with each marker were pinpointed: *AFP, GABBR1, OBP2B, MAMSTR,* and *TRIM46*, respectively. These genes possess varied functions, with *AFP* and *TRIM46* linked to multiple cancers, *OBP2B* and *MAMSTR* implicated in intracellular and extracellular processes, and *GABBR1* associated with nervous system development. Functional annotation was subsequently performed with $FUMA_{GWAS}$, unveiling associations with pathways: glycerophospholipid metabolic, sphingolipid metabolism, and breast cancer 1q21 amplicon ($P_{bon}<0.05$). Moreover, MAGMA Tissue Expression Analysis highlighted significant tissue expression in whole blood ($P_{bon}<0.05$). **Conclusion:** Our GWAS for five tumor markers identified multiple lead SNPs and genes, which were linked to glycerophospholipid metabolism, sphingolipid metabolism and breast cancer 1q21 amplicon. Additionally, gene expression analysis indicated significant expression in whole blood. These finding generated a valuable resource of genetic associations linking variants to tumor markers in East Asian population, providing insights of adjustments for genetic factors in clinical risk prediction for cancers.

Session Title: Cancer Poster Wednesday Session

Board 8035W: Genome-Wide Mapping of Regulatory Elements and Clusters of Non-Coding Somatic Mutations in Liver Cancer Genomes ⭐

Authors:

H. Nakagawa, R. Son, Y. Kanazashi, **T. A. Johnson**, Y. Murakawa; RIKEN IMS, Yokohama, Japan

Abstract:

Our whole genome sequencing analysis demonstrated somatic mutation profiles of hepato-biliary cancers and several coding genes were preferentially mutated as drivers, such as *CTNNB1*, *TP53*, or *ALB* (Fujimoto et al. *Nature Genet.* 2016). Regarding to non-coding regions, more than 50% of liver cancers have hotspot mutations in the *TERT* promoter, however most of

somatic mutations in non-coding regions are not well annotated or interpreted yet. We here collected nearly 5 million of somatic SNVs/indels from whole genome sequencing data of more than 400 primary hepato-biliary tumors in Japan and searched for mutational clusters in non-coding regions by delimiting the whole genome into 500b windows. In addition, to search for regulatory elements associated with liver, we performed ReapTEC (read-level pre-filtering and transcribed enhancer call) that simultaneously profiles gene expression and enhancer activity (eRNA expression) at single-nucleotide resolution (Oguchi et al. *Science*. 2024) and generated genome-wide maps for regulatory elements of 328 liver tumors and non-tumor liver/biliary tissues. These regulatory element profiles of liver tissues show clear discrimination between tumor and non-tumors and the origins of hepato-biliary tumors. Then, we tested overlaps between hundreds of non-coding mutational clusters in liver tumors and 120 thousand of potential regulatory elements associated with liver. This analysis found several non-coding driver mutations such as *TERT* and liver-specific *MIR122* and regulatory elements related with liver cancer and embryonic switching. These results provide insight into the functions of non-coding regions in liver tissue composition and the driver events for liver carcinogenesis and differentiation.

Session Title: Cancer Poster Wednesday Session

Board 8036W: Genomic data analysis of large lung cancer cohort with novel Native Hawaiian data reveals racial disparity in the landscape of somatic mutations

Authors:

Y. Deng, I. Mohd-Ibrahim, M. Jijiwa; Univ. of Hawaii, Honolulu, HI

Abstract:

In this study, we attempt to add further clarity and definition on the distribution of driver genes across the somatic mutation landscape of non-small cell lung cancer (NSCLC) as well as their association with patient features using more than 2500 secondary DNA-sequencing data from NSCLC patients, including 29 from our own cohort of Native Hawaiians (NH), a population that has not previously been represented in NSCLC studies. By our estimates, this is the largest meta-cohort of NSCLC patients assembled from secondary somatic mutation data. We processed and aggregated the genomic data into highly granular somatic mutations features and our latest findings identified in excess of 100 significantly mutated genes (SMG), more than half of which are putatively novel driver genes. Three racially exclusive driver genes were found; RPS3A and VBP1 in White patients, and OR4F16 in East Asians. Although the SMGs were mostly mutated in

all the three racial backgrounds analyzed (Black, White, and East Asian) at the population level, analysis at the individual level revealed a clear disparity. Patients of East Asian ancestry expressed significantly fewer mutations in SMGs besides having much lower TMB. This remained true even after stratifying the populations studied by smoking status and cancer subtype. Apart from EGFR, none of the SMG identified was strongly associated with patients of Asian ancestry. Cox regression models identified better survivability in East Asians and patients with EGFR, whereas patients with mutations in TP53, KEAP1, and BID had significantly poorer survival. The most recent results from this study identify a clear disparity in somatic mutation of lung cancer patients of different racial backgrounds, which should be taken into consideration in future studies. This was particularly prominent in the case of NH, which current findings showed significantly different mutation profiles than the other races investigated. The results also highlight the benefits of utilizing large, heterogeneous datasets of a specific cancer type in driver gene discovery besides identifying gaps in sequenced demographics. This project was supported by grants from the National Institutes of Health (NIH), Award numbers: P20GM103466, U54MD007601, R01CA223490, 1OT2OD032581-02-PP90Y, 1U54HG013243.

Session Title: Cancer Poster Wednesday Session

Board 8037W: Germline and somatic human leukocyte antigen (HLA) variations across racial/ethnic and ancestral populations in 10,479 patients with cancer

Authors:

X. Wang[1], Z. Zhang[2], X. Lin[3], D. Christiani[4]; [1]Harvard Univ., Boston, MA, [2]Dana-Farber Cancer Inst., Boston, MA, [3]Harvard T.H. Chan Sch Pub Hlth, Boston, MA, [4]Harvard Med. Sch., Boston, MA

Abstract:

Background: The human leukocyte antigen (HLA), a highly polymorphic genomic region where both inherited and acquired variations can impact peptide repertoire diversity, has been associated with mixed results for cancer immunotherapy efficacy and immune-related adverse events (irAEs). However, the prevalence of germline and somatic HLA variations across different racial/ethnic and ancestral populations in cancer patients remains largely unexplored.**Methods:** The study included a total of 10479 patients from TCGA with tumor-normal paired whole exome sequencing data and baseline covariates of clinical stage, sex, cancer histology (including solid tumors and blood cancers), self-reported race/ethnicity (SIRE) and genetically inferred ancestry (GIA) from germline data. HLA variations including 1) germline

HLA genotypes and 2) heterozygosity across class I (HLA-A, B, and C) and class II (HLA-DR, DP, and DQ) genes, 3) supertypes, 4) HLA evolutionary divergence (HED), and 5) somatic HLA loss of heterozygosity (LOH) in class I genes, were compared using multinomial logistic regression.**Results:** DPB1*13:01, associated with Killer-cell Immunoglobulin-like Receptors thus regulating immune response, had a higher prevalence in SIRE-Black (OR=2.31, P<0.001) and SIRE-Asian (OR=3.89, P<0.001) compared to SIRE-White, with larger differences in GIA-AFR (OR=2.55, P<0.001) and GIA-EAS (OR=4.04, P<0.001). Supertype A03, associated with increased risks of irAEs, was more prevalent in SIRE-Asian (39%) and SIRE-Black (32%) than SIRE-White (28%). SIRE-Black had higher class I (OR=1.49, P=0.02) and class II (OR=1.46, P=0.01) heterozygosity than SIRE-White, while GIA-EAS had lower heterozygosity (OR=0.71, P=0.03) than GIA-EUR, not captured by SIRE. Conversely, average HLA class I HED was highest in SIRE-White (6.80), significantly higher than SIRE-Asian (6.62) and SIRE-Black (6.49), with differences driven by HLA-A and -B specific HED. HLA class I LOH had a significantly lower prevalence in SIRE-Black (OR=0.73, P=0.03) compared to SIRE-White, which was consistent across individual HLA class I genes. **Conclusions:** Germline and somatic HLA variations differ significantly across racial/ethnic and ancestral populations in patients with cancer, potentially contributing to heterogeneous immunotherapy responses. Characterizing these differences by genetic ancestry can identify additional associations not captured by SIRE, facilitating the development of more accurate, integrative biomarkers for personalized risk stratification and treatment strategies in a diverse, equitable precision medicine approach.

Session Title: Cancer Poster Wednesday Session

Board 8038W: GWAS and Functional Genomics Reveal Candidate Susceptibility Genes, Including Known Drivers *MECOM* and *KRAS*, in Endometrial Cancer

Authors:

T. O'Mara, M. Barker, D. Glubb; QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Abstract:

Endometrial cancer is the fourth most common cancer diagnosed in women in the United States, with ~60,000 new cases estimated in 2022 and incidence projected to increase by 40-50% over the next decade. Despite this, only four drugs have received FDA approval for endometrial cancer in the last four decades. Genetic studies of disease offer a promising approach to identify new effective drug targets. Indeed, it has been observed that there is a four-fold increase in the proportion of drug targets that are supported by relevant genetic

evidence between pre-clinical development and regulatory drug approval. Furthermore, genome-wide association studies (GWAS) have discovered associations with many known drug targets. Thus, we propose that susceptibility genes at endometrial cancer GWAS risk loci will provide novel therapeutic targets underlying disease causality. Through GWAS we have identified 21 genetic loci associated with endometrial cancer risk. To establish candidate endometrial cancer susceptibility genes at GWAS risk loci, we performed functional genomic assays to map regulatory features in endometrial models, including ATAC-seq and H3K27Ac-HiChIP chromatin looping analysis. We intersected 497 fine-mapped candidate causal risk variants with these features to prioritize potentially functional variants and their regulatory targets. This analysis established 214 candidate susceptibility genes, including known endometrial cancer drivers *KRAS* and *MECOM*. Our findings suggest that these prioritized genes will provide valuable insights into endometrial cancer biology as well as potential therapeutic targets.

Session Title: Cancer Poster Wednesday Session

Board 8039W: High-throughput assay for evaluation of unclassified sequence variants in the breast and pancreatic cancer susceptibility gene *ATM*

Authors:

I. Thapa[1,2], K. Clark[1], S. Tavtigian[3]; [1]Huntsman Cancer Inst., Salt Lake City, UT, [2]Univ. of Utah Sch. of Med., Salt Lake City, UT, [3]Huntsman Cancer Inst, Salt Lake City, UT

Abstract:

Variants of unknown significance (VUS) pose a significant problem in the management of breast and other cancers. Development and calibration of functional assays to evaluate the repertoire of VUS is an invaluable way to provide information on clinically actionable variants.
After *BRCA1* and *BRCA2*, *ATM* is the next most important breast cancer susceptibility gene and can confer risk equivalent to mutations in *BRCA2*, but high-throughput functional assays for this gene are lacking. There are around 7,741 reported VUS in ClinVar, and thus the implementation of VUS in clinical decision-making is challenging due to the complexity of interpreting its clinical relevance. We proposed to develop a cell-based multiplex high-throughput functional assay for *ATM* to aid in *ATM* variant classification. This type of functional screen relies on two components: 1) Variant library generation and 2) Functional output measurement for the variant in question. Saturation Genome Editing (SGE)-based library generation has become one of the preferred methods for variant library generation, but there are caveats associated with

this approach. Usually, these screens are conducted in haploid cells in order to assess the contribution of a single allele of the gene of interest, which is a huge-drawback for performing a biologically-relevant cell-based functional assay. Our lab has a novel strategy whereby one of the two *ATM* alleles has been tagged with an FKBP12^{F36V} degradation-epitope (dTAG). After addition of the dTAGv1 drug, the dTAG-ATM protein is quickly degraded (<1.5 hrs), allowing for the functional assessment of a single allele, within the context of a diploid cell. Many SGE-based functional assays simply utilize cell viability over time as a functional output for the variant library. We propose to use a more biologically-relevant assay for ATM, focusing quantitative assessment of ATM target phosphorylation in response to irradiation. These targets include CHEK2, p53, and KAP1 which are activated upon DNA damage.Taken together our novel design consists of the following format: a) CRISPR-Cas9 technology to introduce *ATM* variants into our dTAG cell line; b) degrading one of the two ATM gene products using the dTAG drug (dTAGv1); c) irradiation to initiate phosphorylation of common ATM substrates; and d) flow sorting cells based on quantitative phosphorylation of targets; d) demultiplexing and variant identification. Our work aims in aiding *ATM* VUS classification, enhancing genetic testing, identifying high-risk individuals, and informing personalized treatment plans.

Session Title: Cancer Poster Wednesday Session

Board 8040W: Homologous repair deficiency characterization from RNA-seq using allelic imbalance

Authors:

J. Gascoyne[1,2,3], J. Wong[3], F. San Lucas[4], S. Shukla[2,5], Y. Jakubek[6], P. Scheet[3]; [1]UTHlth.Houston Graduate Sch. of BioMed. Sci., Houston, TX, [2]Univ. of Texas MD Anderson Cancer Ctr. Dept. of Hamtopoietic Biology and Malignancy, Houston, TX, [3]Univ. of Texas MD Anderson Cancer Ctr. Dept. of Epidemiology, Houston, TX, [4]Univ. of Texas MD Anderson Cancer Ctr. Div. of Pathology and Lab. Med., Houston, TX, [5]Univ. of Texas MD Anderson Cancer Ctr. Dept. of Immunology, Houston, TX, [6]Univ. of Kentucky Coll. of Med., Lexington, KY

Abstract:

Damage to, or otherwise inactivation of, genes of the Homologous Recombination Repair (HRR) pathway leads to inadequate repair of double-stranded DNA breaks. Such defects in the HRR mechanism can manifest along the genome as allelic imbalance in and out of the telomeric regions or as genomic breakpoints of structural variants, including copy number alterations, translocations, inversions, and complex rearrangements. By cataloging the number of these events across the genome, malfunctioning of the HRR pathway, termed Homologous

Recombination Deficiency (HRD), is scored. Higher values of this biomarker positively correlate with sensitivity to platinum-based chemotherapies and PARP inhibition through synthetic lethality, and thus HRD is highly clinically relevant. Traditionally, HRD and its constituent subscores - loss-of-heterozygosity (LOH), large-scale state transitions (LST), and telomeric allelic imbalance (TAI) - were called from SNP array data, though as the technologies have become more available, whole-exome, whole-genome, and even RNAseq-based approaches have been developed. As most RNAseq-based approaches use gene expression patterns rather than the individual HRD metrics (LOH, LST, & TAI) to look for evidence of a damaged HRR pathway, we present an additional avenue through which to explore the somatic HRD landscape by utilizing hapLOHseq, an haplotype-aware tool to detect subtle allelic imbalance, with RNAseq data to identify these HRD-associated events. We augment the inference of HRD signatures with superFreq, a depth-based tool for detection of copy number alterations in RNAseq data to better characterize LOH and LST events. We applied this approach to eight malignancies in TCGA - Breast (BRCA), Colorectal (COAD), Glioblastoma (GBM), Lung Adenocarcinoma and Squamous Cell Carcinoma (LUAD & LUSC), Pancreatic (PAAD), Prostate (PRAD), and Ovarian (OV) cancers -- computing HRD profiles. We compare our results to Shi et al. (Nature Scientific Reports, 2023), who calculated HRD scores from SNP-array data, and obtain a nearly identical ranking of HRD distributions across cancers. Moreover, we show that using this allelic imbalance-informed model yields correlation coefficients that varied across cancer types with some HRD-associated cancers exhibiting higher correlation (HNSC, 0.74; BRCA, 0.73). By using these tools, we demonstrate the feasibility of deriving comprehensive HRD genomic signatures from RNAseq data which may be further refined by incorporation of HRD-associated gene expression to one day expand the utility of existing RNAseq assays to call a valuable clinical biomarker.

Session Title: Cancer Poster Wednesday Session

Board 8041W: Identification of breast cancer risk genes which modulate responses to immune-mediated destruction

Authors:

J. Beesley, Y. Luo, W. Shi, G. Chenevix-Trench; QIMR Berghofer, Brisbane, Australia

Abstract:

Genome-wide association studies have identified more than 230 loci associated with breast cancer susceptibility. A major challenge is now to identify the effector genes that contribute to disease risk. We have previously performed pooled CRISPR screens to identify target genes at risk loci that drive cancer hallmarks, including proliferation or modulating DNA damage

response. We have extended these screens to identify genes that enable tumour cells to avoid immune destruction, specifically in the responses to cytotoxic T lymphocyte (CTL) and Natural Killer (NK) cells. We generated pooled CRISPR libraries and performed knockout and inhibition screens to identify genes that affect the response of breast cells to immune effector killing using an in vitro co-culture system. We identified PRMT7 as a candidate breast cancer risk gene which upon genetic and pharmacological perturbation led to resistance to T cell-mediated killing. The role of *PRMT7* in the context of an intact immune system is being evaluated using murine breast cancer models. Our screens to identify intrinsic factors mediating response to NK cells detected multiple hits across several breast cell line models. For example, we validated the effect of the FH gene (encoding Fumarate Hydratase) in sensitising target breast cells to NK-mediated cytolysis using genetic and pharmacological inhibition in MCF7, MDA-MB-231 and the immortalised HMEC line B80T5. Additional in vitro and in vivo validation of *FH* and other hits is underway. Our results demonstrate that phenotypic pooled CRISPR screens are a useful approach for high throughput functional follow-up of GWAS findings, identifying genes which alter immune responses to breast cancer which offer opportunities to enhance T and NK cell-mediated immunotherapy.

Session Title: Cancer Poster Wednesday Session

Board 8042W: Identification of cross-cancer risk loci among groups of cancers defined by biological and etiological similarities

Authors:

A. Hammermeister Suger[1], T. A. Harrison[1], M. C. Wu[2], B. F. Darst[2], S. Lindstroem[1]; [1]Univ. of Washington, Seattle, WA, [2]Fred Hutchinson Cancer Ctr., Seattle, WA

Abstract:

Purpose: Studying the genetic basis of cancer improves our understanding of cancer biology. While individual cancer types have many unique germline genetic risk factors, prior studies of common genetic variation have revealed modest genetic correlations across cancer types. Examining the shared genetic etiology of multiple cancer types could yield novel insights into biological processes that underly carcinogenesis in multiple tissues. **Methods:** We conducted multi-cancer analyses with >480,000 multi-ancestry participants from the UK Biobank cohort, including 114,790 individuals diagnosed with at least one cancer (invasive or *in situ*) defined by International Classification of Disease 9/10 codes. Cancer type was assigned as the first primary cancer, and 367,691 cancer-free controls were used for all analyses. We performed single variant tests for >8.4 million TOPMed imputed autosomal variants (INFO score >0.8, minor allele

frequency [MAF] >1%) and six cancer groups defined by shared biology or etiology. The biologically related cancer groups were mutually exclusive and defined by the germ layer of origin for the tissue (ectoderm, mesoderm, endoderm). The etiologically related cancer groups were non-mutually exclusive and defined by shared associations with exogenous and endogenous hormones, smoking, and infectious agents. We performed all statistical analyses using regenie with a whole-genome regression model based on 610,793 genotyped autosomal variants (MAF >1%) and adjustment for age, genotyping array, and the first 10 genetic principal components. **Results:** We identified 86 genomic loci where germline genetic variation showed a significant association ($p < 5 \times 10^{-8}$) with at least one cancer group. Three of these loci had no previously reported associations with any cancer type. One locus near *CYB5B* and *NFAT5* (chr 16; lead variant rs4783719; $p = 1.4 \times 10^{-8}$; effect allele frequency [EAF] = 0.33; OR = 1.06, 95% CI = 1.04-1.08) was associated with cancers originating from endoderm-derived tissues. Two loci near *SNX32* (chr 11; lead variant rs61714114; $p = 8.5 \times 10^{-11}$; EAF = 0.53; OR = 1.04, 95% CI = 1.03-1.05) and 17q21.2 (lead variant rs112401631; $p = 4.8 \times 10^{-8}$; EAF = 0.98; OR = 1.11, 95% CI = 1.07-1.15) were associated with cancers originating from ectoderm-derived tissues. **Conclusions:** Our results underscore the potential to identify new cancer loci and expand our understanding of the shared genetic architecture of cancer by examining groups of cancers with biological or etiological similarities.

Session Title: Cancer Poster Wednesday Session

Board 8043W: Interim GWAS in bladder cancer: Towards a multi-omics approach

Authors:

T. Lim[1,2], A. Allione[2], C. Viberti[2], C. Debernardi[2], C. Di Primio[2], P. Gontero[3], C. Sacerdote[4], S. Sanna-Cherchi[1], G. Matullo[2]; [1]Columbia Univ., New York, NY, [2]Univ. of Turin, Turin, Italy, [3]Molinette Hosp., Turin, Italy, [4]Piedmont Ctr. for Epidemiology and Cancer Prevention, Turin, Italy

Abstract:

Background: Bladder cancer (BCa) is one of the most common cancers, its heterogeneous clinical outcomes call for further investigation into germline susceptibility factors that can help in risk stratification. We present the interim data of a GWAS (N = 594), and provide an initial framework for multi-omics analysis in 120 individuals who have paired genotyping, mRNA profiling and methylation data in our study. **Methods:** Genotyping was conducted with OmniExpressExome-8 kit in 594 individuals. PLINK2 was used for QC. The resulting genotypes of

581 individuals (148 cases and 433 controls) were imputed with ToPMed (R2) reference panel. Logistic regression was performed using the additive model adjusting for age, sex, and principal components. Polygenic Risk Score (PRS) calculation was performed with PRSice-2 v2.3.5. mRNA expression profiling of post-diagnosis blood samples was performed with Illumina HT-12 v4 chip (67 cases and 72 controls). Differential expression analysis were published and re-analyzed in R (v4.0.5) using limma and illuminaHumanv4.db. Methylation profiling (64 cases and 68 controls) was performed with the HumanMethylation450K chip and analyzed with ChAMP. Multi-omics Factorial Analysis (MOFA) will be performed using features selected from each omic subsequently. **Results:** GWAS QC resulted in a final cohort of 581 individuals (148 cases and 433 matched controls) and 594,108 SNPs. Post-imputation SNP filtering QC yielded a final set of 7,660,056 SNPs for GWAS. We identified three loci with suggestive association with no inflation ($\lambda = 0.969$): chr10p15.3 (rs2605905, OR=3.06, $P = 1.53 \times 10^{-6}$; *DIP2C*), chr5p13.3 (rs5026245; OR = 2.14; $P = 3.83 \times 10^{-6}$; *ADAMTS12*), and chr15q21.3 (rs17543144; OR = 2.03; $P = 4.46 \times 10^{-6}$; nearest gene *FAM214A*). An assessment of the portability of existing PRS(PGS000782) on this cohort showed significant difference in mean scores between cases and controls ($P = 2.17 \times 10^{-4}$), but were not significant when scores were stratified based on BCa subtypes. 42 candidate genes were differentially expressed (adj. P < 0.1) in the mRNA analysis, whereas there was no differential methylated region in the methylation analysis. **Conclusion:** An initial comparison of our cohort against larger bladder cancer cohorts in the GWAS Catalog showed consistent directionality in effect sizes at previously reported loci. Individually, these data are currently underpowered for making new discoveries but, by leveraging on paired epigenomic and transcriptomic data in our cohorts as building blocks, we aim to construct multi-omics genetic correlations that might help uncover novel candidate molecular markers.

Session Title: Cancer Poster Wednesday Session

Board 8044W: Investigating the relationship between breast cancer risk factors and mammographic AI-generated texture feature

Authors:

X. Wu[1], S. Jiang[2], A. Ge[1], C. Turman[3], G. Colditz[2], R. Tamimi[4], P. Kraft[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Washington Univ. Sch. of Med. in St. Louis, St. Louis, MO, [3]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Weill Cornell Med. Sch., New York, NY

Abstract:

Background: Breast cancer is the most prevalent cancer among women globally. The mammographic risk score (MRS), a novel AI-driven texture feature derived from whole

mammogram images, has exhibited a strong association with breast cancer risk independent of breast density. However, despite its promising potential in risk assessment, the mechanisms linking MRS to breast cancer risk remain unclear. We investigated the relationship between established breast cancer risk factors and MRS.
Methods: 383 women from the Nurses' Health Study II were included. MRS was calculated using a model previously trained on 10,126 cancer-free women from the Joanne Knight Breast Health Cohort. Established breast cancer risk factors including childhood and adult body size, fat distribution, reproductive characteristics (age at menarche, age at first birth, age at natural menopause), and mammographic density measures (dense area, non-dense area, percent density) were evaluated. We performed the following analyses: (i) Linear regressions of MRS on each risk factor. (ii) Linear regressions of MRS on genetic components (variants and polygenic scores) previously associated with each risk factor. (iii) Two-stage least squares (2SLS) regression and two-sample Mendelian randomization (TSMR) of MRS on each genetically predicted risk factor. Significantly associated genetic variants and weights were obtained from the largest genome-wide association studies in women of European ancestry for each phenotype. Analyses were adjusted for age at mammogram, genotyping platform, and the top ten genetic principal components. Associations were highlighted if their *P* value < 0.005 (0.05/10, number of risk factors tested).
Results: Higher polygenic scores for dense-area (β = 0.07 SD difference in MRS per SD change in polygenic score; 95%CIs: 0.03-0.11) and percent density (β = 0.06; 95%CIs: 0.02-0.10) were significantly associated with increased MRS. TSMR also identified significant associations between genetically predicted dense-area and MRS (β = 0.82 SD difference in MRS per SD change in dense-area; 95%CIs: 0.39-1.25), as well as between genetically predicted percent density and MRS (β = 1.11; 95%CIs: 0.48-1.74). No evidence was found in support of a significant phenotypic or genetic association of MRS with the other risk factors.
Conclusion: Apart from mammographic density measures, the current study did not identify significant associations or causal effects of other established breast cancer risk factors on MRS, possibly due to limited sample size/power. Further research is needed to validate the findings.

Session Title: Cancer Poster Wednesday Session

Board 8045W: Large-scale population-based study highlights increased cancer risk among individuals with elevated genetic predisposition to autoimmune diseases

Authors:

J. Sun[1], M. Zhang[1], N. Yang[1], X. Li[1], G. Watt[2], X. Guo[3], M. F. Buas[2], X. Li[1], **X. Shu**[2]; [1]Zhejiang Univ., Hangzhou, China, [2]Mem. Sloan Kettering Cancer Ctr., New York, NY, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Background Immune (dys)regulation plays a vital role in carcinogenesis. We investigated the associations between genetical liability to autoimmune diseases and cancer risk. **Methods** This large-scale population-based study included 356339 participants from UK Biobank. The disease-specific and combined polygenic risk scores (PRSs) (both sex and sex-specific, P <5x10^{-8} and r^2 <0.1) for 6 autoimmune diseases (systemic lupus erythematosus [SLE], rheumatoid arthritis, multiple sclerosis, Crohn's disease, ulcerative colitis [UC], and Type 1 diabetes mellitus [T1D]) were constructed using published effect estimates. 23 site-specific cancers were defined using ICD codes. Multivariable logistic regression and Cox regression models were fitted to calculate odds ratios (ORs) and hazard ratios for associations of autoimmune-PRSs with cancer. We further constructed and included cancer-specific PRSs as covariates. Sex-stratified analysis and interaction analysis between autoimmune-PRS and cancer-PRS on cancer were also performed. Finally, LD score regression was used to assess genetic correlations between diseases. **Results** We found 15 significant associations covering 6 PRSs and 8 cancers and 14 additional associations in sex-stratified populations (FDR<0.05). Specifically, SLE-, UC-, and T1D-PRS showed complex cross-cancer effect: SLE-PRS was positively associated with chronic lymphoid leukemia (CLL) (OR: 1.13 per SD increment, 95% CI: 1.07-1.19), Hodgkin's disease, non-Hodgkin's lymphoma (NHL), and lung cancer, and was inversely associated with anal cancer; UC-PRS was positively associated with NHL and inversely associated with anal cancer and oral cancer; T1D-PRS was associated with a higher risk of CLL and oral cancer and a lower risk of kidney cancer. Similar results were observed in Cox regression models. The combined autoimmune PRS was positively associated with NHL. After additionally adjusting for the corresponding cancer-PRS, majority of the associations remained consistent and significant except for associations of the combined autoimmune-PRS and SLE-PRS with NHL. A significant interaction of SLE-PRS with prostate cancer-PRS on prostate cancer was also observed ($P_{interaction}$ <0.05). No significant genetic correlations between autoimmune diseases and cancers were observed. **Conclusion** Individuals with higher genetic predisposition to autoimmune diseases may also have elevated risks of several hematological malignancies and solid cancers. Our study provides new insights into understanding the role of autoimmune diseases in carcinogenesis, which may inform opportunities for target cancer screening and prevention.

Session Title: Cancer Poster Wednesday Session

Board 8046W: Leveraging Custom CRISPR/Cas9 Screens to Identify AraC-Daunorubicin-Etoposide (ADE), Clofarabine, Gemtuzumab Ozogamicin (GO), and Bortezomib Response Modulators Association with Clinical Outcomes in Pediatric AML

Authors:

N. Nguyen[1], P. Parcha[1], A. Tagmount[1], B. Gregory[1], J. Rubnitz[2], R. Ribeiro[2], X. Cao[3], T. Alonzo[4], R. Aplenc[5], A. Gamis[6], T. Cooper[7], E. Kolb[8], S. Meshinchi[9], S. Pounds[10], C. Vulpe[11], J. Lamba[1]; [1]Univ. of Florida Coll. of Pharmacy, Gainesville, FL, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [4]Keck Sch. of Med. of USC, Los Angeles, CA, [5]CHOP, Philadelphia, PA, [6]Univ. of Missouri-Kansas City Sch. of Med., Kansas City, MO, [7]Seattle Children's Hosp., Seattle, WA, [8]Nemours Children's Hosp., Wilmington, DE, [9]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [10]St. Jude Children's Res. Hosp., Memphis, TN, [11]UF Coll. of Vet. Med., Gainesville, FL

Abstract:

Background and Methods: ADE remains the established standard chemotherapy regimen for pediatric acute myeloid leukemia (AML). In conjunction with ADE, GO and bortezomib have been assessed in a clinical trial. However, the results have demonstrated minimal or no additional benefits in outcomes compared to ADE alone. Additionally, the araC and clofarabine were investigated compared to ADE, and no differences in patient outcomes were observed. To improve treatment outcomes, it is crucial to comprehend the underlying molecular mechanisms of these drugs. In this study, we conducted custom CRISPR screens (2442 genes targeting AML relevant, ADE pharmacology and druggable genes) in nine AML cell lines with exposure to ADE, clofarabine, GO or bortezomib followed by an integration analysis with clinical outcome data from multiple trials with respective drug arms. Nine AML-transduced cell populations were treated with IC30 of each drug or DMSO for 7-doubling times. The abundance of sgRNA was analyzed using MAGeCK-RRA pipeline. Genes that met the significant cut-off at 1.5 SD across most cell lines in the negative selection are considered enriched for resistance, and positive selection is considered enriched for sensitivity. Genes that are enriched resistant to one drug but enriched sensitive to the other are antagonistic pleiotropic genes. All significant genes were further evaluated with pediatric AML clinical cohorts with available diagnosis gene expression to determine their prognostic factors for EFS, OS, and MRD1.Results: 2,206 genes met the significance as resistant, sensitive, or antagonistic pleiotropic genes to ADE combination as ADE genes; 2,315 genes were significant as ADE+GO genes; 2,301 genes were significant as ADE+bortezomib genes; and 1,971 genes were significant as araC+clofarabine genes. ADE genes were further evaluated in four independent clinical cohort patients treated with ADE, while

other drug combinations only had one available cohort for clinical integration analysis. Our preliminary results showed that 37 genes met the criteria as clinically and biologically important drug-resistant, drug-sensitive, and drug- antagonistic pleiotropic genes across four patient cohorts treated with ADE. Overall, our study coupled the results of a CRISPR/Cas9 screen with leukemic cell gene expression and clinical outcomes identified clinically and functionally significant ADE, clofarabine, GO, and bortezomib response genes. Current research focuses on developing a pipeline to predict different drug combination outcomes for prioritizing drug treatments and identifying novel drug targets to overcome drug resistance.

Session Title: Cancer Poster Wednesday Session

Board 8049W: Multimodal analysis of rare *BARD1* missense variant found in a patient with multiple primary tumors

Authors:

F. Chan-Pak-Choon[1,2], Y. Gao[1,3], J. Camacho Valenzuela[1,4], B. Rivera[5,6], R. Cuella-Martin[1,3], W. Foulkes[1,2,4,6]; [1]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [2]Lady Davis Inst. and Segal Cancer Ctr., Jewish Gen. Hosp., Montreal, QC, Canada, [3]Victor Phillip Dahdaleh Inst. of Genomic Med., McGill Univ., Montreal, QC, Canada, [4]Cancer Res. Program, Ctr. for Translational Biology, Res. Inst. of the McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [5]Molecular Mechanisms and Experimental Therapy in Oncology Program, Bellvitge BioMed. Res. Inst. (IDIBELL), L'Hosp.et de Llobregat, Barcelona, Spain, [6]Gerald Bronfman Dept. of Oncology, McGill Univ., Montreal, QC, Canada

Abstract:

An estimated 20-30% of hereditary breast cancers are caused by germline variants in *BRCA1/2*, whereas 10% of cases can be explained by other breast cancer-associated genes including *BARD1*. Truncating variants in *BARD1*, such as those occurring in the BRCA1-binding RING domain, have been associated with low-moderate breast cancer risk. However, most *BARD1* missense variants are classified as variants of unknown significance (VUS), likely due to the low penetrance of the gene and lack of functional data.
We report a father-daughter pair, where the proband (daughter) presented with breast carcinomas at age 50y (ER+/PR+/HER2-), 60y (triple-negative breast cancer (TNBC)) and endometrial carcinoma at 62y; and her father with colon polyps at 67y and colon adenocarcinoma at 82y. Clinical germline testing revealed a *BARD1* VUS, c.2258 G>T; p.(G753V) (ClinVar ID: 64955), occurring within the BRCT domain responsible for the recruitment of the BRCA1/BARD1 complex to DNA repair sites, in both patients. In-silico tools predicted the VUS as

damaging based on amino acid conservation and protein structure. Exome sequencing of blood and tumor DNA confirmed the presence of the germline VUS in all samples and detected loss of heterozygosity at the *BARD1* locus in the proband's TNBC. No second hits were identified in the other tumors. Mutational signature 3, linked to homologous recombination repair (HR) deficiency, was detected in the TNBC.
A functional assay was performed to help evaluate the pathogenicity of the *BARD1* VUS. Taking advantage of the endogenous protein depletion system via auxin-inducible degron (AID) tagging, HCT116 *BARD1*$^{AID/AID}$ cells were stably complemented with GFP (positive control), BARD1 wildtype (negative control), BARD1 C53W (known pathogenic missense variant), or BARD1 G753V (VUS). Functional characterization of *BARD1* G753V was then carried out in the above cell lines by cell survival in the presence or absence of olaparib. BARD1 G753V cells appeared sensitive to Olaparib treatment, resembling the phenotypes observed in GFP or BARD1 C53W cells.
According to the ACMG guidelines and ClinGen recommendations, *BARD1* G753V remains a VUS using PS3_supporting (damaging functional impact), PM2 (absent from control population databases) and PP3 (deleterious in-silico predictions).
Nevertheless, our results suggest that *BARD1* G753V causes HR-deficiency, indicating that carriers may benefit from genetic counselling and that affected carriers, including our proband, might be eligible for olaparib treatment.

Session Title: Cancer Poster Wednesday Session

Board 8050W: Multiomic 6-base data from cell-free DNA enhances the performance of liquid biopsy classifiers ⭐

Authors:

T. Charlesworth[1], F. Puddu[1], R. Crawford[1], A. Johansson[1], E. Lleshi[1], A. Modat[1], J. Scotcher[1], N. Harding[1], J. Teyssandier[1], P. Creed[2]; [1]biomodal Limited, Cambridge, United Kingdom, [2]biomodal, Cambridge, United Kingdom

Abstract:

Liquid biopsy for profiling of cell free DNA (cfDNA) in blood holds huge promise to transform how we experience and manage cancer by early detection and identification of residual disease and subtype. While early work in liquid biopsy focused on the identification of actionable somatic variations at specific loci, the past decade has seen an expansion into epigenetic features, notably methylation. 5-methylcytosine (5-mC) profiles of cancer are differential from non-cancer at many more loci and so provide a stronger signal. Moreover, recent research has

suggested that 5-hydroxymethylcytosine (5-hmC) profiles in cfDNA can be a marker for early cancer. However, a standard blood draw yields an average of only 10ng of cfDNA, presenting the dilemma of how to use limited sample to obtain maximum information. We will present the application of a technology which sequences at single base resolution the complete genetic sequence of input DNA fragments integrated with the modification status (unmodified, 5-mC or 5-hmC) for each CpG from low nanogram input quantities of cell-free DNA (cfDNA). Using this technology, we generated whole genome 6-base data (A, C, G, T, 5-mC, and 5-hmC) on cfDNA extracted from plasma of healthy volunteers and patients with colorectal cancer at stages I-IV. We demonstrate how the technology can be used to extract as complete information as possible from a cfDNA sample. We compare differential 5-mC and 5-hmC regions, identify copy number variation and clinically relevant germline and somatic single nucleotide polymorphisms, and multiple fragmentomic features including size, end motif and nucleosome position, across different stages of colorectal cancer. We demonstrate that the AUC for detection of Stage I CRC from healthy volunteers is 0.69 or 0.72 when training a classifier using only 5mC or 5hmC features, respectively, and that this increases to 0.93 when training a classifier using the combination of 5mC and 5hmC features. We highlight individual CRC-relevant features where the resolution of 5mC and 5hmC distinguishes between healthy, Stage I and Stage IV samples, and that these differences would be invisible when just using 5mC information. We propose that the ability to measure 5-mC and 5-hmC at high accuracy and single base resolution, alongside genomic and fragmentomic profiles, from a limited quantity of DNA will enable greater insight into the ctDNA in plasma and enable the development of more accurate liquid biopsy based disease detection.

Session Title: Cancer Poster Wednesday Session

Board 8051W: Multiple newly identified Ewing sarcoma susceptibility loci expand etiologic knowledge and nominate mechanisms of risk

Authors:

A. Hubbard[1], W. Zhou[1], O. W. Lee[1], E. W. Stratton[1], C. Boyce[2], J. Khan[3], M. M. Hudson[4], K. K. Ness[4], Z. Wang[4], K. A. Janeway[5], P. Lupo, J.[6], L. M. Morton[1], S. J. Chanock[1], T. G. Grünewald[7], O. Delattre[8], M. Machiela[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Westat, Mount Airy, MD, [3]NCI - Ctr. for Cancer Res., Bethesda, MD, [4]St. Jude Children's Res. Hosp., Memphis, TN, [5]Dana-Farber/Boston Children's Cancer and Blood Disorders Ctr., Boston, MA, [6]Baylor Coll. of Med., Houston, TX, [7]German Cancer Res. Ctr./Hopp-Children's Cancer Ctr./Heidelberg Univ. Hosp., Heidelberg, Germany, [8]SIREDO Oncology Ctr., Inst. Curie Res. Ctr., Paris, France

Abstract:

Ewing sarcoma (EwS) is a rare, aggressive pediatric cancer common in children of European ancestry, suggesting a genetic component to susceptibility. The pathognomonic *EWSR1::FLI1* fusion is present in > 85% of cases and binds to GGAA microsatellites resulting in altered regional transcriptional activity. Prior investigations also identified six common susceptibility loci associated with EwS. To expand knowledge of the genetic etiology of EwS, we performed the largest genome-wide association study (GWAS) to date consisting of eight international studies combining 1,730 EwS cases and 9,957 cancer-free controls of European ancestry. EwS cases were more commonly male (57%) with a mean diagnosis age of 15.1 years for males and 13.7 years for females. Metastases were present in 28% of cases and were more common in males, older patients, and those with pelvic tumors. Our analysis strengthened evidence for previously reported EwS loci at 1p36.22, 6p25.1, 10q21.3, 15q15.1, 20p11.22 and 20p11.23, and odds ratios (OR) remained high for GWAS (OR=1.53-2.11). We identified an additional five novel loci reaching genome-wide significance ($p\text{-value}_{meta} < 5\times10^{-8}$) at 1p36.13, 7q32.3, 8q24.1, 14q13.3 and 18q21.2. Risk allele frequencies of newly identified loci ranged from 0.01 to 0.9 with estimated ORs ranging from 1.25 to 2.49. We combined published data from the Ewing sarcoma cell line atlas (ESCLA) with the GWAS results for integrated analyses of susceptibility loci. Lead variants in all established loci and the newly identified 8q24.1 demonstrated significantly less distance to GGAA microsatellites with ≥4 repeats (minimum required for binding) and most showed evidence for open chromatin and binding of the EwS *EWSR1::FLI1* fusion. Multiple genes within ±1 Mb of these novel loci exhibited altered expression upon *EWSR1::FLI1* knockdown in 15 EwS cell lines and were nominated for further functional investigation. Our results support a strong inherited genetic component to EwS risk and detail a genetic architecture in which common germline variants with moderately high effect predispose to a rare malignancy. By integrating gene expression changes with *EWSR1::FLI1* knockdown, chromatin structure, and *EWSR1::FLI1* binding, our analyses identify potential germline-somatic interactions at EwS susceptibility loci and nominate target genes for functional investigation. Ongoing analyses are examining potential difference in genetic architecture by sex, tumor location, age at diagnosis, and metastases.

Session Title: Cancer Poster Wednesday Session

Board 8052W: Mutational analysis of *FANCA* gene deletion in Tunisian Fanconi anemia patients

Authors:

R. Frikha, H. Kamoun; Univ. Hosp., Sfax, Tunisia

Abstract:

Objective: Fanconi anemia (FA) is the most common inherited bone marrow failure syndrome characterized by extensive genetic heterogeneity, with at least 11 FA complementation groups. While chromosomal breakage analysis is this disease's gold standard diagnostic test, molecular confirmation remains challenging. The mutation spectrum in the FANCA gene is very heterogeneous. Recent reports show a high frequency of large intragenic deletions in the Fanconi anemia group A gene in North African FA patients. The purpose of this study is to determine the prevalence of the deletion of exon 15 of the *FANCA* gene in Tunisian FA patients.**Methods:** A retrospective cross-sectional study included 29 patients diagnosed with FA. Sanger sequencing was performed using a high-fidelity DNA polymerase. **Results**: Seven FA patients (24.1%) carried a large deletion of exon 15 of the *FANCA* gene.**Conclusion:** Despite the limitations of our study, our finding highlights the high frequency of this mutation; as an important marker for the diagnostic confirmation and genetic counseling processes in Tunisian FA patients.

Session Title: Cancer Poster Wednesday Session

Board 8053W: Mutational Penetrance of the *CYP17A1* Gene in the Evolution of Uterine Fibroids Coexisting with Pregnancy in Senegalese Women ⭐

Authors:

B. Keneme, C. Diop, M. Sembene; Cheikh Anta Diop Univ., Dakar, Senegal

Abstract:

Uterine fibroids, also known as myomas or leiomyomas, are the most common solid tumors of the gynecological tract. Symptoms associated with fibroids include excessive uterine bleeding, pelvic pain or pressure, infertility and pregnancy complications. It is not uncommon for fibroids to be diagnosed in pregnant women. They occur in between 3% and 12% of pregnant women, with some data indicating a prevalence of up to 20%. Our aim is to evaluate the evolution of uterine fibroids and to understand the genetic link between fibroids and pregnancy in Senegalese women. We used PCR-sequencing to analyze the impact of non-synonymous variants on the protein and variability of the *CYP17A1* gene in 56 patients, including 20 pregnancies women with fibroids and 36 patients with fibroids only. First of all mutations were detected using Mutation Surveyor, then the pathogenicity of non-synonymous variants using *in silico* tools and the functional impact of non-synonymous variants were also analysed. The secondary and three-dimensional structure of the protein, gene/gene and protein/protein interactions were determined, and finally the sequences obtained were verified, cleaned and

aligned. Molecular analyses of the gene showed that the c.-34 T>C polymorphism may be involved in the incidence of the disease and its progression during pregnancy, with non-synonymous variants being pathogenic (p.Val5Gly, p.Ser30Asn, p.Gln57His, p.Val66Asp and p.Arg67His), others (p.Val5Gly, p.Leu8Val, p. Tyr14Asn, p.Phe16Ile, p.Lys26Thr) and one specific to fibroid tissue and pregnancy (p.Ser30Asn) create genetic disorders, and only 3 variants (p.Tyr14Asn, p.Phe16Ile and p.Arg45Ile) stabilize the CYP17A1 protein. Co-expression with other genes and functional associations with other proteins are observed. In conclusion, the c.-34T>C polymorphism is involved in fibroid cell growth during pregnancy through the effect of hormone overexpression, while non-synonymous variants lead to dysfunction in protein synthesis.

Session Title: Cancer Poster Wednesday Session

Board 8054W: One patient with acute promyelocytic leukemia (M3) bearing tetraploid plus double t(15;17) and PML/RARA rearrangements achieved a complete remission and remained heathy during the six-years follow-up period

Authors:

L. Zhu, F. Wen, X. Luo, S. Peng, L. Fan, S. Luan, T. Gong, C. Zhang; The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China

Abstract:

Acute promyelocytic leukemia (APL, M3) accounts for approximately 8-15 % of the acute myeloid leukemia (AML). Previous studies have reported a number of common and variant chromosomal rearrangements associated with APL, including a unique case with a tetraploid clone plus two t(15;17) without other chromosomal changes. Despite all-trans-retinoic acid (ATRA) treatment and chemotherapy, leukemic cells persisted in the blood, and the patient died of an intracranial hemorrhage on the 16th day after admission. Here we reported a 32 years-old male patient, who was admitted to the hospital due to fever and limb fatigue on June 25, 2018. Bone marrow smear showed that the morphology of the blast cells was consistent with M3. PML/RARA rearrangements was confirmed by fluorescence in situ hybridization (FISH). In addition, the chromosomal study of bone marrow revealed a 92,XXYY,t(15;17)(q22;q21)x2 in all 20 metaphases. After a combinational therapy with retinoic acid, arsenic trioxide, and Idarubicin (IDA), this patient has achieved a complete remission (CR). The chromosomal study of the bone marrow and peripheral blood samples on August 17, 2018 showed a normal male karyotype of 46,XY in all 20 metaphases. The patient has remained healthy and the karyotype of the bone

marrow and peripheral blood samples has remained normal through the entire follow-up period in the past six years.

Session Title: Cancer Poster Wednesday Session

Board 8056W: Preclinical Model Atlas to Enable Clinical Translatability for Precision Oncology

Authors:

A. Chen[1], T. Bui[1], S. Qin[1], V. Prindle[2], K. R. Sher[2], A. E. Richardson[2], J. P. Plotnik[2], J. A. Samayoa[1], S. Kongpachith[1], X. Zhao[1]; [1]AbbVie, SOUTH SAN FRANCISCO, CA, [2]AbbVie, North Chicago, IL

Abstract:

Preclinical models, such as 2D cultures, 3D organoids, and patient-derived xenograft (PDX) models, play a crucial role in drug development pipelines. However, their respective suitability for various research questions remains unclear. For example, some compounds targeting KRAS pathways are potent in 2D but not effective in vivo. Comprehensive understanding of the strengths and limitations of these models is paramount for drawing meaningful scientific conclusions. Identification of the representative models to assemble clinical patient segments is essential in enabling our human-centric reverse translational research for hypothesis testing and target validation.

Building upon our previous work focusing on 133 3D culture models from DEPMAP, we have expanded our efforts to encompass a comprehensive array of preclinical models. This includes DEPMAP models (207 3D and 2099 2D), 48 pairs of 2D-3D models from AbbVie, as well as around 800 internally generated and 4582 PDX models with transcriptomic data acquired externally from Charles River Laboratories, CrownBio, XenoStart, Wuxi, and Champion.

In this study, we employed cellTinder, a model-free and likelihood-based method developed in house, to quantify the similarity and gaps better understand the relationship between preclinical models and the corresponding patient populations at the transcriptome level. Comparing to alternative methods such as Celligner, our approach preserves the transcriptomic signals from immune and stroma components of patient tumors, so that model comparisons consider both intrinsic and extrinsic aspects of the diseases. In general, we observe agreement for most of human cancer types in TCGA between cell line models and their corresponding patient diseases. Furthermore, preclinical models (2D, 3D, PDX) are compared across each other by their similarity of towards human cancers. Our approach demonstrates strong correlations with designated indications within The Cancer Genome Atlas (TCGA), with notable alignment between the best-correlated preclinical models and corresponding patient cohorts. This

methodology maintains the crucial immune component of the models while mitigating issues of abnormal alignment observed in alternative methods like Celligner. These findings offer valuable guidance for selecting preclinical models that faithfully represent the patient population, for compound efficacy testing or MOA/target validation, thereby enhancing the translational relevance of our drug development efforts.

Session Title: Cancer Poster Wednesday Session

Board 8057W: Prevalence and cancer phenotype of biallelic pathogenic variants in *CHEK2*: A genome-first approach in the Geisinger and UK Biobank cohorts.

Authors:

S. Kim[1,2], J. Kim[2], M. Ramos[3], J. Haley[4], H. Rao[4], U. Mirshahi[5], DiscoverEHR collaboration, B. Graubard[2], H. Katki[6], D. Carey[4], D. Stewart[2]; [1]Cincinnati Childrens Hosp. Med. Ctr., Cincinnati, OH, [2]Natl. Cancer Inst., Rockville, MD, [3]Penn State Univ., University Park, PA, [4]Geisinger, Danville, PA, [5]Geisinger Clinic, Danville, PA, [6]Natl. Cancer Inst., Rockville, PA

Abstract:

Background: *CHEK2* is a cancer predisposition gene that is associated with an increased risk of cancer. Cancer-focused multigene panel testing often includes *CHEK2*, leading to increased recognition of individuals harboring pathogenic/likely pathogenic (P/LP) variants in *CHEK2*. However, despite extensive research on germline variation in *CHEK2*, there remains a lack of comprehensive information regarding the prevalence of cancer and detailed cancer phenotypes related to biallelic P/LP variants in this gene. To address this gap, our study employed a genome-first approach, analyzing data from two large population databases—the UK Biobank (UKBB) and Geisinger—to investigate the incidence of biallelic P/LP variants in *CHEK2* and explore cancer prevalence and phenotypes in affected individuals. **Methods:** We examined exome data from two population genomics databases: UKBB and Geisinger. The UKBB cohort included 469,681 individuals, while the Geisinger cohort comprised 155,815 individuals. Individuals who carry biallelic P/LP variants in *CHEK2* were included in this study. We calculated the prevalence of biallelic P/LP variants in *CHEK2*, cancer phenotypes, and cancer prevalence in these two population-scale biobanks. **Results:** In the UKBB, 3,245 individuals (0.68%), and in the Geisinger cohort, 3,153 individuals (2.0%) carried at least one P/LP variant in *CHEK2*. Among UKBB participants, 8 individuals (0.002%) had biallelic P/LP variants in *CHEK2*. In the Geisinger cohort, 6 individuals (0.004%) harbored these variants. Among those with biallelic P/LP variants in *CHEK2*, in the UKBB, 5 out of 8 individuals (62.5%) developed benign and/or malignant

neoplasms, including skin cancer, lung cancer, breast cancer, colon cancer, renal cell carcinoma, carcinoid tumor, and chronic lymphocytic leukemia. Notably, 4 of them (50,0%) developed more than two neoplasms. In the Geisinger cohort, 4 out of 6 individuals (66.7%) developed similar neoplasms, with 3 of them (50.0%) having more than two neoplasms. Overall, among the 14 biallelic P/LP *CHEK2* carriers, 4 (28.6%) individuals were diagnosed with breast cancer. Remarkably, one individual was diagnosed with breast cancer at age 33 and later experienced brain metastases at age 39. Additionally, another patient had two primary breast cancers—one at age 57 and the other at age 85. **Conclusions:** In two genomically ascertained cohorts, individuals carrying biallelic P/LP variants in *CHEK2* were rare and had a high prevalence of benign/malignant neoplasms. This research sheds light on the impact of these variants on cancer risk and provides valuable insights for clinical care.

Session Title: Cancer Poster Wednesday Session

Board 8058W: Quantifying the impact of coverage and switch accuracy on detection of somatic copy number alterations in cancer using whole exome sequencing.

Authors:

M. Lastrapes, J. Wong, Z. Ozcan, P. Scheet; MD Anderson Cancer Ctr., Houston, TX

Abstract:

Somatic copy number alterations (SCNAs) are important genomic characteristics of many types of cancer. While detection of SCNAs has relied on tissue microarrays, advances in computational methods have allowed for SCNA detection from next-generation sequencing (NGS) technologies. However, many of these methods rely on estimation of allelic imbalance (AI), and thus require sufficient profiling of heterozygous sites across the genome. NGS techniques like whole exome sequencing (WES) are restricted to profiling only sites within the exome and yield more variation in coverage across these sites, leading to potential challenges in AI-based SCNA detection. We have shown previously that detection of SCNAs from WES and RNA sequencing is still robust using AI-based methods, but we have yet to quantify how these parameters affect SCNA event calling accuracy in NGS data. To quantify the impact of sample coverage and phasing accuracy on SCNA detection from NGS data, we first down-sampled genotype calls from 87 WES samples from the glioblastoma (GBM) TCGA project to coverages ranging from 5X to 50X. In parallel, we assessed the impact of haplotype phasing accuracy in these WES samples by systematically introducing switch errors to the existing phasing at rates ranging from 1% to 70%. We then performed SCNA detection using hapLOHseq. The resulting set of SCNA events was

then compared against events obtained by running hapLOH on matched microarray data, which we applied here as a relative gold standard. Not surprisingly, increased coverage and increased phasing accuracy each resulted in higher gene-level F1 scores. Interestingly, F1 scores remained above 0.5 until switch error rates larger than 26% or an average coverage of 5X or lower, suggesting that AI-based SCNA detection methods are highly robust to limitations in coverage and phasing accuracy. We then aimed to determine whether the addition of orthogonal NGS data could improve SCNA detection accuracy in WES samples. Using the same GBM samples as earlier, we combined the set of WES genotype calls with those from matched RNA sequencing data. We performed SCNA detection as described earlier, this time using the combination of allelic read counts across matched WES and RNA samples. Preliminary analysis of SCNA calls from combined data showed little improvement in F1 scores compared to WES alone, although we are extending this analysis to explore the potential impact on samples with extremely low coverage or poor phasing accuracy. These results reinforce our confidence in SCNA calling from WES data and offer helpful considerations for the use of SCNA detection methods in NGS studies.

Session Title: Cancer Poster Wednesday Session

Board 8060W: Sex chromosome gene expression does not match reported sex in a large proportion of cancer cell lines

Authors:

M. Wilson[1], M. Lee[1], A. Mangelsdorf[1], M. Farmwald[1], I. Ramirez[1], J. Denning[1], T. Allen[1], M. Swamy[1], M. Wells[1], N. Paneuf[1], J. Del Rosario[1], K. Buetow[2], S. Plaisier[1]; [1]Arizona State Univ., Tempe, AZ, [2]Arizona State Univ., Tempe, AR

Abstract:

Although there is a known sex bias in cancer, sex chromosomes are typically excluded from genomic analyses because of unique analytical challenges associated with their sequence and regulation. Sex chromosome genes include tumor suppressors, epigenetic regulators, immune genes, and more which can impact cancer progression and drug response. A common resource used in cancer research is the Cancer Cell Line Encyclopedia (CCLE), which contains 1019 human cancer cell lines from many primary tumors. In this study, we inferred the sex chromosome complement of each of these cell lines using gene expression from the sex chromosomes. We expected the expression of X and Y chromosome genes to match the expected sex chromosome complement given the reported sex of the patient (e.g., typically XX genotype in cell lines from female patients and XY for male patients). After inferring the sex chromosome complement of

the cell lines in the CCLE, we found mismatches between the reported sex of the patient and the inferred sex chromosome complement in approximately half of the samples in each sex. Using the same method for inferring sex chromosome complement, we further inferred the sex chromosome complement from primary human tissues from the Genotype-Tissue Expression (GTEx) consortium and primary human tumor samples from The Cancer Genome Atlas (TCGA) consortium. In normal, non-cancerous tissues, sex chromosome gene expression followed the expected expression pattern given the sex of the person who donated the sample in all cases (i.e. high chrY genes in all males, high XIST in all females). In primary tumor tissues, varying levels of mismatched sex chromosome gene expression was observed across tumor types. In summary, our findings demonstrate that sex chromosome gene expression is often lost in cancer cell lines and, to a lesser extent, this is replicated in primary tumors. Further, our pipeline for inferring sex chromosome complement is a resource for researchers who aim to more fully incorporate sex chromosome genetics into their research.

Session Title: Cancer Poster Wednesday Session

Board 8061W: Single-cell analysis of gastric premalignant lesions and gastric cancer reveals cell lineage diversity and micro-environmentheterogeneity

Authors:

C. Hu[1,2], L. Bian[1], H. Ma[1], C. Yan[1], G. Jin[1]; [1]Nanjing Med. Univ., Nanjing, China, [2]Southeast Univ., Nan jing, China

Abstract:

Background: According to the Lauren classification, gastric cancer (GC) is divided into the intestinal and diffuse types (IGC and DGC). A precancerous cascade for IGC driven by multiple events has been proposed by Correa's hypothesis. However, the mechanism underlying DGC progression is poorly understood. **Methods:** We collected GC tumor (5 IGC and 5 DGC), premalignant lesion (10 intestinal metaplasia) and normal (7 non-atrophic gastritis) from the same individuals with GC underwent surgical resection and performed single cell RNA sequencing (scRNA-seq). Cell differentiation trajectory was reconstructed cell lineage development. For identifying cell subpopulations associated with DGC driver gene mutations, we used Scissor to analyze phenotype (DGC significantly mutation status) -guided single-cell subpopulation. Reclustering of the fibroblasts revealed different cell types based on known marker genes. We isolated primary fibroblasts from surgical removed DGC or the adjacent normal samples. FASC was used to sort extracellular matrix related CAFs (eCAFs). RNA

interference, plasmid overexpression, as well as proliferation, tube formation abilities of HUVECs, were used to evaluate the biological functions of *FAP*. ELISA kit was used to detect the levels of PLAU and CXCLs in supernatants. The binding of transcription factors was confirmed by CUT&RUN. **Results:** After quality control, we obtained and profiled 116,733 cells which mainly assigned to 7 major cells. For IGC, we found that a proportion of enterocytes with high expression level of proliferation-related genes progressed to IGC by reconstructing pseudotemporal trajectory. Combined with whole-exome sequencing in DGC, we confirmed cancer cells with *ARID1A* deficient interact strongly with fibroblasts. Particularly, eCAFs were enriched in the DGC with high expression of *FAP*. The CM of cancer cells with *ARID1A* deficient contained more PLAU, which can promote the expression of ERK phosphorylation and ELK3 in normal fibroblasts by interacting with PLAUR. Moreover, we further found that *ELK3* can bind to the promoter region of *FAP* to induce eCAFs transformation which expressed higher levels of CXCLs to promote angiogenesis rather than affecting the phenotype of cancer cells directly. **Conclusions:** Our results provided an evidance for Correa's hypothesis in the development of IGC. Besides, we found that cancer cells in DGC with *ARID1A* inactivation can secrete PLAU, which binds to PLAUR on the surface of normal fibroblasts and activates the RAS/ERK/ELK3 pathway. *ELK3* directing conversion of normal fibroblasts to eCAFs which can secret CXCLs to promote the development of DGC by inducing angiogenesis.

Session Title: Cancer Poster Wednesday Session

Board 8062W: Single-cell landscape deciphering cancer cell-of-origin and cellular heterogeneity in malignant transformation of 13 major tissues

Authors:

R. Luo; UTHlth., Houston, TX

Abstract:

Deciphering disease progression and the sophisticated tumor ecosystems is imperative for exploring tumorigenesis mechanisms and developing novel prevention strategies. Here, we dissected heterogeneous tissue microenvironments during malignant transitions by leveraging data from 4,972,145 single cells in 1396 samples from 62 datasets spanning 13 major tissue types. Within transitional stem-like subpopulations highly enriched in precancerous lesions and cancers, we identified 30 recurring cellular states, including hypoxia and epithelial senescence, revealing a high degree of plasticity in epithelial stem cells. By characterizing the dynamics of stem-cell crosstalk with the microenvironment along the pseudotime axis, we uncovered

distinct roles of ANXA1 at different stages of tumor development. ANXA1 expression levels in stem cells were decreased from the healthy to the precancerous stages, which promoted inflammatory responses by recruiting neutrophils and regulating monocyte differentiation towards M1 macrophages. In contrast, during malignant progression, upregulated ANXA1 fostered M2 macrophage polarization and cancer-associated fibroblast transformation. Our spatiotemporal analysis further provided insights into mechanisms responsible for immunosuppression. Collectively, this study provided a systematic view of cancer origins, and suggested that restoring and maintaining the balance of inflammation and their mediators (e.g., AnxA1/FPRs signaling) may represent a novel approach to control the evolution of precancerous lesions and mitigate the risk for cancer development.

Session Title: Cancer Poster Wednesday Session

Board 8063W: Smoking and Pathogenic Germline Variants in Patients with Lung Cancer

Authors:

E. Esplin[1], S. Nielsen[1], E. Russell[1], R. Ellsworth[1], A. Watson[2], D. Aisner[3], W. Korn[1]; [1]Invitae, San Francisco, CA, [2]Univ. of Colorado Sch. of Med., Aurora, CO, [3]Univ. of Colorado Sch. Med., Aurora, CO

Abstract:

Background: Few studies have investigated the correlation of smoking history with pathogenic germline variants (PGVs) in patients (pts) with lung cancer. With emerging opportunities for PGVs in DNA damage-repair (DDR) genes and precision therapies, we investigated the prevalence of PGVs in pts with lung cancer, stratified by smoking status.
Methods: Germline genetic testing (GGT) (Invitae Corp.) and insurance claims (Komodo Healthcare MapTM) data were reviewed for 14,317 pts with an ICD code for lung cancer from 2015-2023 and hereditary cancer GGT of > 10 genes. PGV rates were assessed overall, in DDR genes and in lung cancer related genes. Smoking history was defined by ICD codes. Number of genes tested varied per ordering clinician preference. Clinically actionable PGVs were defined as those associated with clinical management recommendations or trial eligibility per current standard of care guidelines. Chi-square tests were used with significance set to $p < 0.05$.
Results: The cohort was predominantly female (70.9%), white (65.7%), and over half had a history of smoking (57.1%). PGVs in 73 known cancer-risk genes were identified in 1,804/14,317 (12.6%) pts. Prior analyses show PGV rate does not vary with personal or family history. PGV frequencies were 12.6% (1,026/8,175) in pts with smoking history compared to 12.7%

(778/6,142) in those with no reported smoking history. There was no statistically significant difference in DDR gene, TP53, or EGFR PGV rates for pts with or without smoking history (p>0.05). Mean (SD) age of lung cancer diagnosis was significantly younger for pts with a history of smoking and DDR PGVs (61 (11)) compared to those with negative results (62 (12), p = 0.005). There was no significant difference in age of diagnosis between pts with no smoking history and DDR PGVs (60 (12)) and those with negative results (59 (13), p=0.888). PGV rates by clinician-reported ancestry: Black/African-American, 9%; Asian or Pacific Islander, 11%; Hispanic, 12%; White, 13%. Among genes with >1,000 pts tested, PGVs were most common in BRCA2 (2.2%), CHEK2 (2.0%), ATM (1.6%), SPINK1 (1.6%), and BRCA1 (1.2%).
Conclusions: In 14,317 pts with lung cancer, 12.6% had PGVs. There was no significant difference in enrichment of PGVs in DDR genes when pts were stratified by reported smoking history, suggesting the PGVs of these genes associate independently with lung cancer. The age of onset for pts with DDR gene PGVs varied with smoking history, suggesting that risk conferred by these PGVs could be additive to smoking risk. These data reinforce prior studies supporting consideration of GGT for all pts with lung cancer, independent of age or reported smoking history.

Session Title: Cancer Poster Wednesday Session

Board 8064W: Spatial guided contrast learning for predicting spatial transcriptomics from histology images to enhance breast cancer patient stratification and immunotherapy response

Authors:

P. Hu; Western Univ., London, ON, Canada

Abstract:

Spatial transcriptomics (ST) has revolutionized cancer research by providing detailed insights into cellular arrangements and gene expression patterns within tumor microenvironments. This spatial information holds immense potential for refining patient stratification and tailoring treatment strategies. However, the prohibitive costs and expertise required for ST data acquisition hinder its widespread clinical adoption. In contrast, conventional hematoxylin and eosin (H&E) stained histology images are widely used in routine clinical practice, exhibiting predictive capacity for ST data. We hypothesize that a deep learning-based framework will be able to predict ST data from unlabeled breast cancer (BC) patients in The Cancer Genome Atlas (TCGA). Furthermore, the predicted ST data can be used to predict the immune checkpoint inhibitor (ICI) treatment response of BC patients. We developed a digital pathology foundation

model-driven spatial guided contrastive learning to predict spatial transcriptomic profiles from histology images. Specifically, the approach adopts a contrast learning approach for cross-modal prediction and imputation through a joint embedding space between image embedding and expression embedding. Furthermore, it incorporates a graph neural network that serves as a spatial guide. This enhances the model's performance in spatial region detection and increases interpretability. Comprehensive testing on multiple datasets shows that the approach significantly outperforms existing methods in terms of gene expression prediction and spatial region identification, and better retains biological information. By employing the BC subtype-specific cell types identified from the predicted ST data, we generated a list of BC subtype-specific and cell type-specific immune-related gene signatures. These gene signatures were used to build classification models that can accurately predict the response status (responder vs no responder) of BC patients treated with ICI (Durvalumab or Olaparib and Pembrolizumab 2) in I-SPY2 clinical trials. The approach provides a robust solution for generating spatial transcriptomics data from histology images, thereby elucidating the molecular signatures of tissues and offering a promising avenue for precision oncology.

Session Title: Cancer Poster Wednesday Session

Board 8065W: Testing results and clinical applications from 73 lymphoma patients using a 115 gene panel for lymphoma

Authors:

H. Peng, J. Ma, Z. Zhou, X. Yuan, Y. Da, Q. Xiao, X. Li, Y. Wan, Y. Yu, H. Zhang, Z. Zhong, J. Chen, T. Gong, C. Zhang; The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China

Abstract:

Lymphomas include Hodgkin lymphoma (HL) and Non-Hodgkin lymphoma (NHL) that are different in the ways they develop, spread, and are treated. It is important to know the type of lymphoma because the treatments and prognosis vary for different types of lymphoma. In this study, we have developed a 115 gene panel for lymphoma and tested 73 patients with lymphoma, including 63 cases of diffuse large B-cell lymphoma (DLBCL), 3 cases of angioimmunoblastic T-cell lymphoma, 2 cases of follicular lymphoma, one case of Burkitt lymphoma (BL), one case of B-cell lymphoma (small lymphocyte lymphoma), one case of intermediate large cell lymphoma, one case of follicular T-helper cell lymphoma and one case of Mantle cell lymphoma. DLBCL is the most common type of non-Hodgkin lymphoma, and frequently harbor alterations in KMT2D, TP53, CREBBP, PIM1, and BCL2. The top 20 gene alternations detected in the cohort of 63 DLBCL patients are listed

following: *PIM1* (41%), *CD79B* (33%), *TP53* (33%), *MLL2* (30%), *MYD88* (26%), *ETV6* (24%), *PCLO* (20%), *BTG1* (17%), *LRP1B* (17%), *MYC* (17%), *SPEN* (15%), *TET2* (15%), *B2M* (13%), *CARD11*(13%), *IRF4* (13%), *P2RY8* (13%), *ARID1B* (11%), *KLHL6* (11%), *RELN* (11%), and *SGK1* (11%). The testing results have been successfully used for determination of the cell origin and classification of molecular subtypes of DLBCL.

Session Title: Cancer Poster Wednesday Session

Board 8066W: The mutational spectrum of arsenic-related human tumors

Authors:

N. Jain[1], E. Wu[1], M. Kibriya[1], F. Jasmine[1], E. Wilkinson[1], S. E. Haque[2], H. Ahsan[1], Y-Y. He[1], B. Pierce[1]; [1]The Univ. of Chicago, Chicago, IL, [2]UChicago Res. Bangladesh, Dhaka, Bangladesh

Abstract:

Inorganic arsenic, a naturally occurring metalloid found in groundwater, is a well-established human carcinogen. Epidemiological studies from countries with high levels of arsenic contamination of groundwater show clear links between arsenic exposure and increased risk of skin, lung, bladder, and kidney cancer. However, studies investigating the mechanisms of arsenic-related mutagenicity in humans remain limited. To date, there has been one study that examined the mutational spectrum of a single arsenic-exposed lung cancer patient and two in vitro studies that examined the mutagenic impact of short-term arsenic exposure. To improve our understanding of arsenic mutagenesis in humans, we performed whole-exome sequencing (WES) on tumor samples from 19 patients at hospitals in Dhaka, Bangladesh, with known or suspected arsenic exposure based on arsenic concentrations in urine and toenail. Our data includes bladder (n=11), kidney (n=3), non-melanoma skin (n=2), lung (n=1), breast (n=1), and ureter (n=1) cancers. We identified somatic mutations using the tumor-only variant calling method Mutect2 and extracted mutational signatures using SigProfiler. The mutation burden (mutations per megabase) across all exomes ranged from 51.3 to 260.9 (median 82.1) for single-base substitutions (SBS), 1.8 to 4 (median 2.6) for double-base substitutions, and 3.2 to 23.6 (median 16.1) for small insertions and deletions. While the somatic mutation burden is comparable to findings from previous studies, our tumors tend to be on the higher end of the observed range. We identified five mutational signatures with strong similarity to those in the Catalog of Somatic Mutations in Cancer: SBS1 (deamination of 5-methylcytosine), SBS5 (aging-related endogenous mutational process), SBS19 (unknown), SBS30 (defective DNA repair associated with NTHL1 mutations), and SBS32 (azathioprine treatment). The SBS5 signature,

which contributed the highest mutation load across all exomes, has been previously associated with bladder cancer, the most common type in our dataset. The high mutation load and lack of novel signatures suggests that arsenic potentially exacerbates endogenous mutational processes rather than leaving its own distinct signature. To further investigate, we are performing WES in additional kidney, lung, and bladder samples and in HaCaT cells that have undergone malignant transformation after chronic low-level arsenic exposure. We will also compare the patterns of somatic mutations observed to cancers from populations in different geographic regions, as this will further highlight the impact of environmental and lifestyle factors on mutational patterns.

Session Title: Cancer Poster Wednesday Session

Board 8067W: Unveiling Telomere Dynamics: Alterations in Telomeric Nuclear Architecture and Variability in Telomere Lengths in Alternative Lengthening of Telomere

Authors:

O. Samassekou[1], S. Mai[2]; [1]Univ. of Sci., Technics and Technologies, Bamako, Mali, [2]Univ. of Manitoba, Winnipeg, MB, Canada

Abstract:

Introduction: Cancer cells employ two mechanisms, telomerase and alternative lengthening of telomeres (ALT), to maintain their telomeres at a minimum length, enabling continued cell proliferation and exacerbating genomic instability. While some genomic variances have been identified between these pathways, the precise consequences on telomeres and genomic instability remain elusive. Thus, investigating the nuclear organization of telomeres and their individual lengths in cells using either of these mechanisms will yield valuable insights into their impact on genomic instability.**Methodology:**We used cell lines that were positive for telomerase and ALT to compare both mechanisms. Fluorescence in situ hybridization with probes specific to telomeres was used to study telomeric nuclear architecture and the length of individual telomeres in both nuclei and metaphases. Subsequently, three-dimensional imaging and Teloview™ were used to assess telomeric nuclear architecture, including the number of telomeres, their intensities, and telomere aggregates. Finally, the lengths of individual telomeres were evaluated based on fluorescence intensities of each chromosome arm.**Results**: We observed notable distinctions between telomerase-positive and ALT cells, with the latter exhibiting a higher frequency of telomeric aggregates compared to telomerase-positive cells. Telomeres in ALT cells demonstrated greater variability in both length and structure. This

increased variability was apparent not only among different telomeres within the same cell but also among homologous telomeres and those located on the same chromosome arms. For instance, we observed that the length ratio between the short and long arms on the same chromosome did not significantly differ among homologous chromosomes in telomerase-positive cells, whereas in ALT cells, this ratio could vary by ten to twentyfold.Furthermore, the spatial distribution of telomeres within the nuclei of ALT cells was less orderly and more dispersed compared to the more structured and organized distribution observed in telomerase-positive cells. This disorganized rearrangement of telomeres in ALT cells may contribute to their unique mechanisms of telomere maintenance, which rely on recombination-based processes to elongate telomeres. **Conclusion**: The differences in telomere dynamics between ALT cells and telomerase-positive cells underscore the distinct biological pathways these cells use to maintain telomere length and integrity. Understanding these differences is crucial for developing targeted therapies for cancers that utilize the ALT pathway for telomere maintenance.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5001W: A Combination of polygenic risk score and measured blood pressure for risk stratification in cardiovascular disease mortality.

Authors:

H. Okumiyama[1], R. Fujii[1,2], M. Nagayoshi[2], M. Nakatochi[2], Y. Tsuboi[1], K. Suzuki[1], K. Wakai[2], K. Matsuo[3], J-MICC Study Group; [1]Fujita Hlth.Univ. Sch. of Med. Sci., Toyoake, Japan, [2]Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [3]Aichi Cancer Ctr. Res. Inst., Nagoya, Japan

Abstract:

Background: Blood pressure (BP) control is a simple and effective preventive measure for cardiovascular diseases (CVD). Recent studies suggested that BP is affected by individual genetic predisposition. However, clinical goals for BP control do not account for genetic predisposition. Polygenic risk score (PRS) is calculated using the summary statistics of previous GWAS and indicates the individual predisposition for a phenotype. Meanwhile, in clinical settings, it is less evident whether combining BP PRS with routinely measured BP could be beneficial. Therefore, we examined joint associations of BP control status and BP PRS with CVD mortality in a general Japanese population.**Methods:** A total of 14,086 Japanese middle-aged adults who lived in the 14 study-designated areas throughout Japan is included in this study. PRS for systolic blood pressure (SBP) and diastolic blood pressure (DBP) were developed by using summary statistics from the Biobank Japan. PRS-CS was used to calculate both in PRS SBP and PRS DBP. Participants

were followed up from the baseline survey (2005-2014) to the end of 2022. Controlled BP was defined as SBP of 140 mmHg or less or DBP of 90 mmHg or less regardless of antihypertensive medications.**Results:** At the baseline survey, 5,063 (45.0%) were men and the mean age was 55.0 (SD: 9.3) years. Of 11,252 eligible participants, 100 CVD mortality cases were documented during a median 14.4 (IQR: 12.1-15.6)-year follow-up period. Compared with controlled BP, HRs (95% CI) of CVD mortality among high BP group were 1.93(1.28, 2.92) in SBP and 2.57 (1.71, 3.88) in DBP. Compared to controlled BP + low PRS, HRs (95% CI) in the high PRS were 2.27 (1.11, 4.65) and 2.66 (1.30, 5.47) even though controlled SBP and DBP.**Conclusions:** Our results suggested that genetic predisposition for SBP and DBP could add more information to BP control about risk stratification on CVD mortality. Integrating PRS with BP measurements could identify “hidden CVD risks” even though BP is normal or is controlled BP by antihypertensives. Future guidelines for high BP are expected to include genetic predispositions as a new chapter and could promote the reduced CVD risk with the incorporation of BP PRS.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5002W: A genetic analysis of brain cortical curvature and major depressive disorder in Mexican Americans

Authors:

R. Hernandez[1], J. Curran[1], D. Glahn[2], J. Blangero[1], A. M. Winkler[1]; [1]Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX, [2]Boston Children s Hosp. & Harvard, Boston, MA

Abstract:

Major depressive disorder (MDD) is one of the most prevalent mental health conditions in the United States. Its prevalence is particularly high among Hispanic individuals. While extensive research has linked brain structural measurements, such as reduced gray matter volume, to MDD, there is growing interest in other brain morphology indices, such as cortical curvature. Previous studies suggest that there may be a genetic component that influences cortical curvature, and that MDD is associated with lower gyrification. In the present study, we investigate, in a sample of Mexican Americans, the degree to which cortical curvature is heritable, and its association with MDD. This study utilized MRI scans from 1407 participants of the Genetics of Brain Structure study. All participants are members of randomly selected families (extended pedigrees) of Mexican American ascent. MDD diagnostic procedures used a consensus-based approach and the MINI-Plus interview for DSM-IV and ICD-10 diagnoses, supplemented by the Beck Depression Inventory II and Eysenck Personality Questionnaire. The cohort is comprised of 819 females and 588 males, aged between 18 and 85 years (mean =

40.96; SD = 15.48). Images were processed using FreeSurfer 7.4.1, an image processing tool that automatically generates several cortical morphology measurements. For this analysis, we focused on the mean curvature, which represents the average of minimum and maximum curvatures, further smoothed with a Gaussian filter of FWHM = 10mm. The images were downsampled to the resolution of fsaverage4 (2562 vertices per cerebral hemisphere). Heritability and association analyses between cortical curvature and MDD were conducted using SOLAR 8.1.1. Correction for multiple testing was based on the false-discovery rate, with q = 0.05. MDD was found to be heritable, with a heritability estimate of (h^2=0.32, p=3x10^{-7}). The heritability of cortical curvature varied greatly across the cortex, with higher values observed in regions such as the superior and medial temporal gyri, cingulate gyrus, and areas surrounding the lateral cerebral fissure, central sulcus, parietooccipital fissures, lingual gyrus, calcarine sulcus, cuneus, parahippocampal gyrus, orbital frontal gyrus, and callosal sulcus. The heritability in these regions ranged from 0.17 to 0.62. The association between cortical curvature and MDD was found to be weak, sparse, and broadly lacking statistical significance. While both MDD and cortical curvature are both heritable traits, their association was found to be weak.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5003W: A genetic case-control study of three single nucleotide polymorphisms in *ABCB4*, *ABCG8* and *SLC10A2* to explore their contribution towards the onset of cholelithiasis in Pakistani population

Authors:

M. Ajmal, A. M. Butt, A. Zafar, S. Laraib, Q. Malik, N. Ahmed; COMSATS Univ. Islamabad (CUI), Islamabad, Pakistan

Abstract:

Cholelithiasis, also known as gallstone disease (GSD), is a common disorder characterized by the accumulation of biliary materials in the form of stones in gallbladder. The condition remains asymptomatic mostly until later stages when symptoms including abdominal pain, nausea, and vomiting start. Currently, laparoscopic cholecystectomy is the most common surgical procedure to cure and relieve affected individuals. Studies have reported involvement of multiple genetic factors in GSD, but our population is still unexplored. Therefore, a total of 251 individuals including clinically diagnosed cases and controls were selected for this study. The variants rs754287486, rs11887534 and rs56398830 in *ABCB4*, *ABCG8* and *SLC10A2*, respectively, were genotyped using Tetra ARMS-PCR. Amplified PCR products were electrophoresed on 2% agarose

gel to reveal the genotype of every participant. Data analysis was done using Epi Info and Python in VS Code. Missing values in the data were imputed using SimpleImputer module of scikit-learn Python library. Data analysis without imputed values revealed lack of association of rs754287486 (OR=0.78, LB=0.23, UB=2.69, p=1.00), rs11887534 (OR=0.86, LB=0.32, UB=2.30, p=1.00), and rs56398830 (OR=1.53, LB=0.33, UB=7.19, p=0.70) with the risk of developing GSD. Other two independent variable age and gender were not associated with the disease also. Data analysis with inclusion of imputed values revealed significant association of rs56398830 (OR=4.17, LB=1.21, UB=14.32, p=0.03) with GSD. Interestingly, risk alleles of all three single nucleotide polymorphisms (SNPs) were not found in a homozygote state in our data indicating the absence of one genotype for each SNP. Thus, suggesting a crucial function of *ABCB4*, *ABCG8* and *SLC10A2* during development. According to our knowledge, this is the first study in Pakistan, aimed to explore the role of the genetic factors in cholelithiasis. Multifactorial disorders underly complex interplay of several contributing factors either environmental and/or genetic leading to development of disease phenotype. Although the *ABCB4*, *ABCG8* and *SLC10A2* genes have previously been implicated in GSD or related complications, our preliminary findings suggest lack of association of studied SNPs with GSD in Pakistan. In addition, missing genotypes in our data suggest recessive lethal nature of the risk alleles, which might prevent normal fetal development leading to premature death of the developing embryos. Analysis with larger dataset including more cases and controls may refine and clarify the current findings, especially significant association of rs56398830 with GSD using imputed values.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5004W: A loss-of-function *CCR2* variant is associated with lower cardiovascular risk

Authors:

M. Georgakis[1], R. Malik[2], N. Hasbani[3], J. Li[4], J. Huffman[5], S. Damrauer[6], C. Kooperberg[7], J. Rotter[8], P. De Vries[9], M. Dichgans[1]; [1]LMU Munich, Munich, Germany, [2]ISD Munich, Munich, Germany, [3]UTHlth.Houston, Dallas, TX, [4]Geisinger, Lewisburg, PA, [5]Palo Alto Veterans Inst. for Res., Boston, MA, [6]U Penn, Philadephia, PA, [7]Fred Hutchinson Cancer Ctr., Seattle, WA, [8]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [9]UTHlth.at Houston, Houston, TX

Abstract:

Background and Aims: Ample evidence links CCL2, a key chemokine governing monocyte trafficking, with atherosclerosis. However, it remains unknown whether targeting the CCL2

receptor CCR2 could provide protection against cardiovascular disease.
Methods: Computationally predicted damaging (REVEL>0.5) variants within CCR2 were detected in whole-exome-sequencing data from 454,775 UK Biobank participants and tested for association with cardiovascular endpoints in gene-burden tests. Given the key role of CCR2 in monocyte mobilization, variants associated with lower monocyte count were prioritized for experimental validation. The response to CCL2 of human cells transfected with these variants was tested in migration and cAMP assays. Validated loss-of-function variants were tested for association with cardiovascular endpoints, atherosclerosis burden, and vascular risk factors. Significant associations were replicated in six independent datasets (n=1,062,595).
Results: Carriers of 45 predicted damaging CCR2 variants were at lower risk of myocardial infarction and coronary artery disease. One of these variants (M249K) was associated with lower monocyte count and decreased signaling and chemoattraction in response to CCL2. While M249K showed no association with conventional vascular risk factors, it was consistently associated with a lower risk of myocardial infarction (Odds Ratio: 0.66 95% Confidence Interval: 0.54-0.81,p=6.1x10-5) and coronary artery disease (Odds Ratio: 0.74 95% Confidence Interval: 0.62-0.87, p=2.9x10-4) in the UK Biobank and in six replication cohorts. In a phenome-wide association study, there was no evidence of higher infections risk among M249K carriers.
Conclusions: Carriers of an experimentally confirmed loss-of-function CCR2 variant are at a lower lifetime risk of myocardial infarction and coronary artery disease without carrying a higher infection risk. Our findings provide genetic support for the translational potential of CCR2-targeting as an atheroprotective approach.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5005W: A modular network-based approach improves post-omics and post-GWAS hypothesis generation about disease genes and mechanisms ⭐

Authors:

A. McKim[1], A. Krishnan[2], C. A. Mancuso[2]; [1]Michigan State Univ., East Lansing, MI, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

A major post-omics or post-GWAS goal is to take the hundreds (or even a few thousands) of disease- or trait-associated genes identified from these studies and combine them with other biological data to benefit two-folds: i) refine and reprioritize the list genes, removing false positive hits, and discovering novel genes (that are likely false negatives), and ii) gain insights

into the molecular pathways, phenotypes, and cell types these genes converge on to ultimately impact the disease. Genome-scale molecular interaction networks have been invaluable in providing a molecular scaffold for understanding how genes work with each other and, when combined with machine learning (ML), discovering previously-unknown genes associated with pathways and phenotypes. However, directly applying this network-based approach to complex diseases is challenging because, unlike genes related to pathways and phenotypes, complex disease genes are not localized in gene networks within a shared neighborhood. This is confirmed by the fact that building a single network-based ML classifier for the disease as a whole leads to poor gene discovery performance.
We propose a novel network-based ML approach, ModGenePlexus, to address this challenge. ModGenePlexus works in two stages. First, using semi-supervised learning, it decomposes the large disease gene list into coherent 'modules' that each contain a subset of original disease genes and some new candidate genes tightly connected to each other in the underlying network. Second, using supervised learning, ModGenePlexus trains one ML classifier per disease module to learn network patterns unique to that module and predict additional novel genes related to that module. We applied ModGenePlexus to large disease gene lists derived from several transcriptomics studies and GWASs. Using systematic and rigorous evaluations, we demonstrate that ModGenePlexus yields improved performance and facilitates a more nuanced and biologically specific interpretation of the known and novel disease genes. Additionally, we show that, by using networks to provide biological context, ModGenePlexus is capable of using even genes with nominal p-values to improve prediction performance. These findings underscore the advantages of ModGenePlexus over conventional methods that consider the entire disease gene set as a single unit, thereby expanding the applicability of network-based ML to post-GWAS or post-omics hypothesis-generation, especially for answering questions about functional, cellular, and phenotypic convergence of disease genes.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5006W: A Multi-ancestry Transcriptome-Wide Association Study of Alzheimer's Disease

Authors:

X. Sun[1], M. Mews[1], N. Wheeler[1], P. Benchek[1], C. Reitz[2], A. Naj[3], J. Below[4], B. Kunkle[5], A. Griswold[5], W. Bush[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Columbia Univ., New York, NY, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Vanderbilt Univ., Nashville, TN, [5]Univ. of Miami, Miami, FL

Abstract:

Alzheimer's Disease (AD) is a neurodegenerative disorder affecting millions worldwide, with its exact cause still unknown. Transcriptome-wide association studies (TWAS) have advanced the analysis of non-coding variants by integrating gene expression data. However, most TWAS methods predominantly use single-ancestry expression data, mainly from European populations. To address this gap, our study employed multivariate adaptive shrinkage in R (MASHR) and LassoSum to leverage regulatory effects and adjust effect estimates due to Linkage Disequilibrium (LD) across diverse populations. We incorporated African, European, and Hispanic expression and genotype data from MAGENTA Project to identify AD-associated genes. Quality control (QC) on genotype data involved filtering SNPs with a minor allele frequency > 0.05. Post-QC, 224 African, 236 European, and 339 Hispanic samples from MAGENTA, with both whole-blood expression and genotype data, were analyzed. Cis-eQTL analysis was performed for each protein-coding gene within a ±1 Mb region. MASHR utilized weights and standard errors to adjust based on correlation patterns in the data, allowing for both shared and condition-specific effects, thus increasing power and effect size estimation accuracy. We then applied LassoSum to compute Elastic Net estimates of MASHR posterior means, accounting for LD, using the MAGENTA reference panel for each population. Finally, we used S-PrediXcan to apply our models to ancestry-specific AD GWAS summary statistics, focusing initially on the African American AD GWAS from Kunkle et al. (n=8,006). We applied Bacon to control for bias and inflation in the TWAS results. After the correction, the inflation factor was 1.27. 70 significant genes were identified after Bonferroni correction. Gene enrichment analysis on GO biological processes highlighted the negative regulation of platelet activation and G2 DNA Damage Checkpoint Signaling. *ALOX12*, an important lipoxygenase, emerged as a significant gene involved in the negative regulation of platelet activation and its effect size was positively associated with AD. Preliminary results suggest that *ALOX12* may contribute to oxidative stress and inflammation, leading to blood-brain barrier breakdown. Our approach addresses a limitation in typical MASHR models, which only selected top SNPs, resulting in sparse prediction models. In contrast, our method considers genes that are potentially polygenic with multiple eQTLs and small effects. We plan to apply this method to all three ancestries and perform a meta-analysis of the results in the future.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5007W: A multi-pronged strategy for target identification in Polycystic Kidney Disease

Authors:

S. Sarangi[1], C. J. Guerriero[2], M. Peterson[1], K. Sharma[1], S. Moore[1], J. L. Brodsky[2]; [1]Paradigm4, Waltham, MA, [2]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Polycystic Kidney Disease (PKD) is a genetic disorder characterized by growth of fluid-filled cysts in the kidneys. PKD is progressive and leads to reduced kidney function and to kidney failure in 75% of cases by the time patients are in their 60's. The most common form of this disease is Autosomal Dominant Polycystic Kidney Disease (ADPKD). ADPKD, which is typically diagnosed in adulthood, affects more than 600,000 people in the USA and 12.5 million people worldwide. Consequently, ADPKD carries a considerable economic and societal burden, and the estimated total annual costs in 2018 ranged from $7.3 to $9.6 billion in the USA alone. Although multiple clinical studies have been conducted, only one medication has been approved so far to slow disease progression. Thus, there is an urgent need to improve biomarker identification strategies and accelerate the development of therapies for this devastating disease.
In previous studies, we had explored associations between variants of the gene PKD2 and phenotypes related to ADPKD using the UK Biobank (UKBB) dataset (Application ID: 51518). PKD2 encodes the protein polycystin-2 (PC2) and mutations in this gene have been linked with ADPKD. PC2, a ubiquitously expressed protein, is a member of the transient receptor potential channel family of proteins. It functions as a cation permeable channel helping in calcium transport and cell signaling. In this study, we present results of further characterization of PKD2 variants using multiple approaches. We screened for variants across TOPMed, ClinVar, and the Mayo Clinic database, and performed two-sample Mendelian randomization (MR) analysis with PKD2 variants and phenotypes consisting of ADPKD-relevant phecodes and those selected from EMBL's GWAS catalog. The MR analysis was conducted in REVEAL, a bioinformatics platform developed by Paradigm4, and utilized the 470K whole exome sequencing dataset from the UKBB. Following identification of PKD2 mutations, we used Rhapsody, a machine learning (ML) tool for predicting the impact of amino acid substitutions on protein pathogenicity. Finally, we employed a yeast-based assay that reports on PKD2 folding, transport, and activity at the cell surface to test and validate effects of identified mutations.
Developing viable therapies for ADPKD has proved challenging at least in part due to the high phenotypic variability arising from an interplay of genetic, epigenetic, and environmental factors that impact robust target identification and informed patient selection. The multiple and complementary strategies reported in this study will lead to a deeper understanding of the mechanism of action in complex diseases such as ADPKD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5009W: A novel stop gain *SORL1* mutation in a Peruvian family with Alzheimer's Disease of the PeADI Study

Authors:

M. Cornejo-Olivas[1,2], A. Griswold[3], A. Saldarriaga-Mayo[2,1], P. Mena[3], R. S. Rodriguez[1,2], L. Adams[3], P. Whitehead[3], R. Isasi[3], M. Illanes-Manrique[1,2], E. Sarapura-Castro[1,2], F. Rajabli[3], K. McInerney[3], C. Manrique-Enciso[2], G. Beecham[3], S. Castro-Suarez[4,5], P. St George-Hyslop[6], I. Araujo-Aliaga[1,2], M. L. Cuccaro[3], J. Vance[3], M. Pericak-Vance[3]; [1]Neurogenetics Working Group, Univ. Cientifica del Sur, Lima, Peru, [2]Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [3]John P. Hussman Inst. for Human Genomics. Univ. of Miami Miller Sch. of Med., Miami, FL, [4]CBI en Demencias y Enf. Desmielinizantes, Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [5]Global Brain Hlth.Inst., San Francisco, CA, [6]Columbia Univ., New York, NY

Abstract:

Objective:*SORL1* risk variants have been reported in both early-onset and late-onset AD. The aim of this abstract is to report a novel *SORL1* AD-risk variant within a Peruvian AD family.**Background:**Common and rare variants in *SORL1* have been associated with an increased risk of Alzheimer's disease (AD). Since 2019, we have run an international collaborative research initiative to ascertain a Peruvian cohort for Alzheimer’s disease and other related dementias for genetic studies (PeADI).**Design/Methods:**A Peruvian family (4 AD cases and two mild cognitive impairments (MCI) was recruited through the PeADI study. All six family members completed a full cognitive assessment, plasma-based biomarkers pTau181 and Aβ42/40, and whole genome sequencing. Variants within AD risk genes as determined by the ADSP Gene Verification Committee were prioritized and variant interpretation was performed according to ACMG recommendations. Local ancestry analysis was performed within *SORL1* region. SORL1-transfected HEK293 APPsw cells were developed, lysed and the medium harvested to perform immunoblot assays.**Results:**We identified a *SORL1* c.5019G>A (p.Trp1673Ter) variant in the four AD diagnosed siblings within the Amerindian ancestry region of the gene. The two MCI cases did not carry the novel variant. The identified *SORL1* variant corresponded to a heterozygous stop-gain variant in exon 36 replacing tryptophan by a stop codon at position 1673 of the *SORL1* protein. This variant has not been previously reported in databases including gnomAD and ClinVar. The 4 AD cases had on average 2.3X higher plasma pTau181 concentrations than the 2 MCI (2.03 ± 0.28pg/ml vs 0.88 ± 0.7pg/ml). There was no noticeable difference in the Aβ42/40 ratio. This variant is classified as likely pathogenic according to ACMG. Immunoblotting assays show the absence of full-length SORL1 in HEK293 APPsw cells, which produce a relatively stable truncated protein.**Conclusions:**We report the first Peruvian AD family carrying a likely pathogenic stop-gain *SORL1* variant of Amerindian ancestry origin. Preliminary

functional analysis shows that this variant produces a relatively stable truncated SORL1 protein, which is secreted into the medium.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5010W: A polygenic risk score and case finding for chronic obstructive pulmonary disease

Authors:

J. Zhang[1], B. Hobbs[2], E. Silverman[2], H. Xu[1], J. Dupuis[3], G. O'Connor[1], M. Cho[4], M. Moll[2]; [1]Boston Univ., Boston, MA, [2]Brigham & Women s Hosp., Boston, MA, [3]McGill Univ., Montreal, QC, Canada, [4]Brigham & Women s Hosp., Duxbury, MA

Abstract:

Background: Chronic obstructive pulmonary disease (COPD) is a leading cause of morbidity and mortality and remains underdiagnosed worldwide. As universal screening for COPD is not recommended, effective case-finding approaches are clinically relevant. Genetic factors play an important role in the pathogenesis of COPD. We hypothesized that a person's genetic risk for COPD, aggregated into a polygenic risk score (PRS), facilitates the diagnosis of COPD. **Methods:** We included non-Hispanic white participants without a self-reported history of COPD from the Genetic Epidemiology of COPD (COPDGene) of current and former smokers and the community-based Framingham Heart (FHS) studies. We used the polygenic risk score for COPD (PRS_{COPD}) that we previously built based on the UK Biobank and SpiroMeta lung function genome-wide association studies. We calculated a PRS_{COPD} for each COPDGene and FHS participant as a weighted sum of the effect of genetic variants across the genome. The outcome of this study was GOLD 2-4 COPD defined using FEV_1/FVC <0.7 and FEV_1 % predicted <80%. We performed logistic regression analysis using three sets of nested predictive variables: age, sex, smoking status, and pack-years of smoking (model 1), model 1 + respiratory symptoms including chronic cough, wheezing, and dyspnea (model 2), and model 2 + PRS_{COPD} (model 3). Model discrimination, calibration, and net benefit were assessed using area-under-receiver-operating-characteristic-curve (AUC-ROC), calibration plots, and decision curve analysis, respectively. We adjusted for principal components of genetic ancestry in all models. **Results:** Of 3,135 COPDGene and 3,551 FHS participants without a prior diagnosis of COPD, 586 (18.7%) COPDGene and 289 (8.1%) FHS participants met spirometric criteria for GOLD 2-4 COPD. There were significant differences in the AUC-ROC between model 1 and model 2 (COPDGene: 0.71 versus 0.74, p<0.001; FHS: 0.76 versus 0.78, p=0.008) and between model 2 and model 3 (COPDGene: 0.74 versus 0.78, p<0.001; FHS: 0.78 versus 0.85, p<0.001). The model with the

PRS_{COPD} (model 3) was well calibrated in both COPDGene and FHS. Based on the decision curves, clinical factors and PRS_{COPD} (model 3) provided a superior benefit compared to clinical factors (models 1 and 2) in guiding the decision for further diagnostic testing (e.g., spirometry) in a wide range of threshold risk probabilities. **Conclusions:** Among people without previously diagnosed COPD, adding a PRS to clinical factors improves the identification of persons with increased risk for COPD. Further studies of the clinical implementation and benefit of a PRS in the case-finding for COPD are needed.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5011W: Aberrant recursive splicing as a cause of human disease

Authors:

R. Harripaul[1,2,3], P. Boone, M.[1,2,4], R. Yadav[1,2,3], R. Collins[1,2], M. Grzybowski[5], N. Makhortova[6], A. Sanchis-Juan[7,1], M. Larson, O[8], P. Gonzalez Tovar[4], J. Lemanski[1,2], D. Ben-Isvy[1,2,9], M. Handley[10], X. Zhao[1,2], S. Erdin[1,2], C. De Esch[1,2], K. Mohajeri[1,2], M. Moyses Oliveira[1,2], J. Gusella[1], D. Tai[1,2], S. Sanchez-Roige[11], B. Currall[1,2], S. Slaugenhaupt[1], C. McGraw[12], R. Doan[13], A. Palmer[11], A. Geurts[5], M. Talkowski[14,15,3]; [1]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [2]Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Dept. of Neurology, Harvard Med. Sch., Boston, MA, [4]Div. of Genetics and Genomics, Boston Children's Hosp., Boston, MA, [5]Dept. of Physiology, Med. Coll. of Wisconsin, Milwaukee, WI, [6]Human Neuron Core, Boston Children's Hosp., Boston, MA, [7]Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [8]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Cambridge, MA, [9]Div. of Med. Sci., Harvard Med. Sch., Boston, MA, [10]Harvard Stem Cell Inst. Flow Cytometry Core, Massachusetts Gen. Hosp., Cambridge, MA, [11]Dept. of Psychiatry, Univ. of California San Diego, La Jolla, CA, [12]Dept. of Neurology, Boston Children’s Hosp., Boston, Boston, MA, [13]Div. of Genetics and Genomics, Boston Children’s Hosp., Boston, Boston, MA, [14]Massachusetts Gen. Hosp., Boston, MA, [15]Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

Recursive splice sites are rare motifs hypothesized to facilitate splicing across massive introns and to shape isoform diversity of long, brain-expressed genes. Yet, their purpose is unsubstantiated, and the role of recursive splicing (RS) in disease remains unclear. In a cross-phenotype analysis of large and rare copy number variants in ~1M individuals, we identified deletions of ~100kb in the first intron of CADM2, removing a recursive splice site (RS1), that conferred substantial risk for attention deficit hyperactivity disorder (ADHD; odds ratio = 42).

CADM2 encodes a neuronally expressed cell adhesion molecule and has been repeatedly linked to ADHD-related traits, including impulsivity and drug/alcohol use. To uncover the molecular impact of RS, we used CRISPR to model the above ~100kb patient deletions, or to delete or invert a targeted ~500 bp region including RS1, in human induced pluripotent cells (hiPSCs). We differentiated these models to induced glutamatergic neurons (iNs), then performed RNA sequencing (RNA-seq) and cellular phenotyping. Total CADM2 expression is nominally decreased in patient deletion iNs (12.7% reduction, p=7.6 E-3). This is exacerbated in the homozygous targeted deletion iNs (22% reduction, false discovery rate (FDR=6.3 E-4). RNA-seq also provided a census of CADM2 transcripts including novel transcripts retaining RS exons. All edited lines exhibited significant decreases in the most abundant transcript containing RS exons: heterozygous targeted deletion with 80.4% reduction (p=1.9 E-3); heterozygous targeted inversion with 75.3% reduction (p=2.0 E-3); and homozygous targeted deletion with 92.8% reduction (p=2.8 E-6). Thus, deleting CADM2 RS1 shows significant changes in transcript diversity and abundance, with homozygous deletion models intensifying the effects seen in heterozygotes. On the transcriptome-wide level, heterozygous patient and homozygous targeted deletions showed significant effects with 124 and 716 differentially expressed genes (DEGs, FDR <0.1) respectively. There is a significant sharing of DEGs in these two edits, underscoring shared processes and pathways between these two models (hypergeometric p=2.4 E-28) which showed enrichment of terms associated with neuronal processes, namely neurogenesis, synapse, and axon development (FDR < 0.1). In summary, we demonstrate a functional role for RS in the expression of a neurological gene CADM2, show that alterations to RS can contribute to the risk of human diseases, and provide details about the genetic risk for ADHD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5012W: Advanced Genetic Characterization and Pathway Analysis of IgA Nephropathy: Insights from Rare Variant, Exome-Wide Studies, and Genetic Diagnostics

Authors:

N. Elefant[1], S. Krishna Murthy[1], S. Yang[1], H. Milo Rasouly[1], N. Vena[2], K. Bogyo[2], A. Khan[1], L. Liu[1], J. Zang[1], M. Marasa[3], K. Mucha[4], M. Mizerska-Wasiak[4], B. Julian[5], J. Novak[5], F. Scolari[6], S. Sanna-Cherchi[7], K. Kiryluk[8], A. Gharavi[7]; [1]Columbia Univ. Med. Ctr., New york, NY, [2]Columbia Univ. Med. Ctr., New York, NY, [3]Columbia Univ. Med. center, new york, NY, [4]Med. Univ. of Warsaw, Warsaw, Poland, [5]Univ. of Alabama at Birmingham, Birmingham, Alabama, Birmingham, AL, [6]Univ. of

Brescia – ASST Spedali Civili, Brescia, Italy, Brescia, Italy, [7]Columbia Univ, New York, NY, [8]COLUMBIA Univ. HEALTH Sci.S, New York, NY

Abstract:

Background: IgA nephropathy (IgAN) is a common form of glomerulonephritis, characterized by the deposition of IgA-containing immune complexes in the glomeruli, leading to progressive renal damage. The etiology of IgAN is multifactorial, involving genetic predispositions and environmental factors. While common variants have been implicated by GWAS, the role of rare variants and monogenic diseases in pathogenesis remains incompletely understood.**Methods:** In this study, we conducted exome and genome sequencing in 5338 IgAN cases recruited from multiple nephrology centers across diverse geographical regions. We matched 4,999 unrelated high quality IgAN cases to 18,950 controls based on genetic ancestry to perform exome-wide association analysis (ExWAS) and rare gene burden analysis. Cases and controls FASTQ data were all processed, aligned, and called using the same pipeline.**Results:** Diagnostic analysis based on a gene panel of known glomerulonephritis and immune dysregulation genes yielded a 3.6% and 0.9% diagnostic rate, respectively. 21 syndromes were detected, and the most frequently encountered genes were *COL4A3*, *COL4A4*, *COL45*, *CFH*, *NLRP3*, *INF2* and *CASP10*. The ExWAS confirmed several known loci significantly associated with IgAN risk, consistent with previous GWAS studies. Rare variant analysis identified several potentially deleterious variants in genes related to immune dysregulation and the TGFβ signaling pathway, including *SMAD3* and*CXCR4*.
Conclusion: We detected an expectedly high diagnostic yield in genes related to kidney and immune-related diseases, indicating that genetic testing would be valuable in the clinical work up of IgAN. The rare variant analysis suggests involvement of the TGFβ pathway, highlighting potential new targets for therapeutic intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5013W: Alteration of JAK-STAT signaling in a patient with TYK2 variants

Authors:

P. Sharma, F. Navid, N. Watts, S. Brooks, R. Colbert; NIAMS, NIH, USA, Bethesda, MD

Abstract:

Tyrosine kinase 2 (TYK2) is a member of the Janus kinase family that associates with the cytoplasmic domain of type I/II cytokine receptors and STATs to transmit signals that mediate

both innate and adaptive immune responses. Genetic variants in TYK2 gene are associated with a number of inflammatory diseases including axial spondyloarthritis (AxSpA). Exome sequencing of a patient with severe IBD and AxSpA revealed compound heterozygous TYK2 gene variants, p.A53T and p.R703W, with allele frequencies of 0.007301 and 0.006692, respectively, in gnomAD. In Silico predictors such as PolyPhen, SIFT and Mutation Taster revealed these variants to be damaging. Thus, we were interested in whether these variants were of functional significance. To assess this, we used skin fibroblasts to produce induced pluripotent stem cells (iPSCs) and then differentiated the iPSCs into mesenchymal stem cells (MSCs) and osteoblasts. Insilico protein modeling was done to predict the effect of these variants on protein function. Bulk RNA sequencing was performed and comparative profiling of IFN-α response gene set was done in patient and control cells.MSCs from the patient and 3 healthy controls cells were stimulated with IFNα for various times ranging from 5 to 60 minutes. Cell lysates form cultured MSCs were used in Western blotting analysis for measuring phosphorylated TYK2 and downstream phosphorylated STAT1 using specific antibodies. Preliminary results suggest that stimulation with IFNα resulted in higher phospho-STAT1 levels and possibly higher phospho-TYK2 in patient cells compared to healthy controls. Alizarin staining of cultures obtained from MSC-derived osteoblasts revealed increased mineralization by patient-derived cells. Protein modeling data suggests that the A53T variant could change the conformation of the FERM-SH2 domain, which might alter cytokine receptor binding. The R703W variant may affect dimerization, or it might affect the conformation of the pseudo-kinase domain, and thereby the activity of the kinase domain. As compared to controls, RNA sequencing data showed increased expression of IFN-α response gene set in patient cells. Preliminary evidence suggests that these TYK2 gene variants contribute to gain-of-function.IFN responses in patient cells which may partially be responsible for elevated proinflammatory response downstream of TYK2. Further experiments are being performed to determine downstream effects of increased signaling on gene expression and to better understand the possible role of these variants in AxSpA and IBD pathogenesis

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5014W: Alzheimer's Disease risk and somatic loss of the Y chromosome

Authors:

H. Cukier[1], E. Palmer[2], P. Benchek[3], N. Wheeler[4], S. Smeisek[4], A. Naj[5], J. Haines[6], M. Pericak-Vance[7], L. Forsberg[8], Y. Song[4], W. Bush[6]; [1]Univ. of Miami, Miami, FL, [2]Case Western Reserve Univ., South Euclid, OH, [3]Case Western Reserve Univ., Cleveland Heights, OH, [4]Case Western

Reserve Univ., Cleveland, OH, [5]Univ. of Pennsylvania, Philadelphia, PA, [6]Case Western Reserve Univ, Cleveland, OH, [7]Univ. of Miami Miller Sch. of Med, Miami, FL, [8]Uppsala Univ., Uppsala, Sweden

Abstract:

Mosaic loss of the Y chromosome (LOY) is a somatic, age-related event that has been previously associated with a variety of diseases of aging. A prior study of European cohorts demonstrated an association between LOY and Alzheimer's Disease, and more recent molecular studies have shown that LOY can also occur within microglia, suggesting a potential functional role in AD pathogenesis. In this study, we further validate the association between LOY in blood and AD via prospective analyses of 1,447 males. We define a continuous estimate of LOY by computing the median of the log R ratio (mLRRY) from raw intensity genotyping data. We also perform Mendelian Randomization analysis on 10,013 males across 26 cohorts from the Alzheimer's Disease Genetics Consortium (ADGC) dataset. Age at blood draw was stratified into four bins consistent with prior literature (<=65, 66-75, 76-85,>=86), and Wilcoxon rank sum tests were performed to examine the distribution of LOY by age at blood sampling. LOY increased with age and was highly significant over all samples and between all age groups. Evaluating a cohort of 3,117 males, there are significant differences between males 65 years old and younger compared to all other age groups (66-75 p=0.00061, 76-85 p=7.9 x 10^{-12}, >=86 p = 5.5 x 10^{-13}).To evaluate the association between LOY and risk of developing AD in follow-up, we fit a Cox proportional hazards model associating binary LOY to years of disease-free follow-up within 1,447 individuals for whom blood was sampled prior to any AD diagnosis. Of these, there were 42 subsequent AD events with all others remaining controls. Overall, we find a significantly increased hazard (1.98 [1.17-6.91], p = 0.0021) in men with LOY versus those without. To account for potential biases in sample ascertainment with respect to age and case/control status, we performed Mendelian Randomization analyses of LOY. Using the Wright-based instrument, we had a significant two-sample Mendelian randomization effect using the Inverse Variance Weighted approach (p = 0.008). Significant results from these analyses provide further evidence for a role of LOY as a potential biomarker of Alzheimer's Disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5015W: Alzheimer's disease microglia exhibit an enrichment of somatic cancer driver mutations, correlating with disease-associated inflammatory states ⭐

Authors:

Y. Huang[1,2], Z. Zhou[1,2], M. Talukdar[1,2], S. Marro[3], E. Papapetrou[3], E. Lee[1,2], C. Walsh[1,2]; [1]Boston Children's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Alzheimer's disease (AD), an age-associated neurodegenerative disorder, is characterized by progressive neuronal loss and the accumulation of misfolded proteins like amyloid-β and tau. Neuroinflammation, mediated by microglia and brain-resident macrophages, plays a pivotal role in AD pathogenesis, yet the intricate interactions among age, genes, and other risk factors remain elusive. Somatic mutations, known to accumulate with age, instigate clonal expansion across diverse cell types, impacting both cancer and non-cancerous conditions. Utilizing molecular-barcoded deep panel sequencing on 311 prefrontal cortex samples of AD patients and matched controls, our study unveiled an elevated occurrence of somatic mutations within cancer driver genes in AD brains. Recurrent somatic mutations, often multiple, were observed in genes associated with clonal hematopoiesis (CH). Remarkably, these somatic cancer driver mutations were specifically enriched in CSF1R+ microglia of AD brains and exhibited signs of positive selection, suggesting mutation-driven microglial clonal expansion (MiCE). Single-nucleus RNA sequencing of the temporal neocortex samples from additional 62 AD patients and matched controls revealed a nominal increase in subchromosome-level somatic mutations associated with CH in AD microglia, with mutant microglia exhibiting upregulated pro-inflammatory and AD-associated genes. We further created three lines of induced pluripotent stem cell (iPSC)-derived microglia, each carrying driver mutations in one of the most mutated genes in AD brain. Through single-cell RNA sequencing, we verified that these mutant microglia consistently displayed an amplified inflammatory and disease-associated transcriptional profile compared to their wild-type counterparts. Our findings indicate that somatic cancer driver mutations in microglia are prevalent in normal aging but further enriched in AD, driving MiCE and promoting inflammatory, disease-related microglial signatures. This study provides crucial insights into microglial clonal dynamics in AD, potentially paving the way for novel approaches to AD diagnosis and therapy.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5016W: An iPSC-based neural model of sialidosis uncovers glycolytic impairment-causing presynaptic dysfunction and deregulation of Ca^{2+} dynamics

Authors:

H. Furuya[1], H. Odaka[2], N. Tadahiro[3], T. Matsushita[1], T. Era[3]; [1]Kochi Med. Sch., Kochi Univ., Kochi, Japan, [2]Cellular and Molecular Biotechnology Res. Inst., Natl. Inst. of Advanced Industrial Sci. and Technology (AIST), Central 6, Ibaraki, Japan, [3]Dept. of Cell Modulation, Inst. of Molecular Embryology and Genetics, Kumamoto Univ., Kumamoto, Japan

Abstract:

Sialidosis is a neuropathic lysosomal storage disease caused by a deficiency in the NEU1 gene-encoding lysosomal neuraminidase and characterized by abnormal accumulation of undigested sialyl-oligoconjugates in systemic organs including brain. Although patients exhibit neurological symptoms, the underlying neuropathological mechanism remains unclear. Here, we generated induced pluripotent stem cells (iPSCs) from skin fibroblasts with sialidosis and induced the differentiation into neural progenitor cells (NPCs) and neurons. Sialidosis NPCs and neurons mimicked the disease-like phenotypes including reduced neuraminidase activity, accumulation of sialyl-oligoconjugates and lysosomal expansions. Functional analysis also revealed that sialidosis neurons displayed two distinct abnormalities, defective exocytotic glutamate release and augmented α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate receptor (AMPAR)-mediated Ca^{2+} influx. These abnormalities were restored by overexpression of the wild-type NEU1 gene, demonstrating causative role of neuraminidase deficiency in functional impairments of disease neurons. Comprehensive proteomics analysis revealed the significant reduction of SNARE proteins and glycolytic enzymes in synaptosomal fraction, with downregulation of ATP production. Bypassing the glycolysis by treatment of pyruvate, which is final metabolite of glycolysis pathway, improved both the synaptsomal ATP production and the exocytotic function. We also found that upregulation of AMPAR and L-type voltage dependent Ca^{2+} channel (VDCC) subunits in disease neurons, with the restoration of AMPAR-mediated Ca^{2+} over-load by treatment of antagonists for the AMPAR and L-type VDCC. Our present study provides new insights into both the neuronal pathophysiology and potential therapeutic strategy for sialidosis.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5017W: Analysis of the coding regions of 33,763 individuals identifies known and novel gene associations for idiopathic nephrotic syndrome

Authors:

J. Ke[1], P. Gundula[1], S. Shril[2], G. Yask[1], T. Lim[1], A. Mitrotti[3], I. Pisani[4], M. Bodria[5], A. Cristina Simões e Silva[6], W. Morello[7], G. Montini[7], U. Maggiore[4], E. Fiaccadori[4], L. Gesualdo[3], G.

Ghiggeri[5], M. Kretzler[8], K. Kiryluk[1], M. Saleem[9], R. Gbadegesin[10], M. G. Sampson[2], A. G. Gharavi[1], D. S. Paul[11], S. Petrovski[11], M. R. Pollak[12], D. B. Goldstein[13], F. Hildebrandt[2], S. Sanna-Cherchi[1]; [1]Columbia Univ., New York City, NY, [2]Boston Children Hosp., Boston, MA, [3]Univ. of Bari, Bari, Italy, [4]Univ. of Parma, Parma, Italy, [5]IRCCS Istituto Giannina Gaslini, Genoa, Italy, [6]Federal Univ. of Minas Gerais, Belo Horizonte, Brazil, [7]Ospedale Maggiore Policlinico di Milano, Milan, Italy, [8]Univ. of Michigan, Ann Arbor, MI, [9]Univ. of Bristol, Bristol, United Kingdom, [10]Duke Univ., Durham, NC, [11]AstraZeneca, Cambridge, United Kingdom, [12]Beth Israel Deaconess Med. Ctr., Boston, MA, [13]Actio BioSci.s, San Diego, CA

Abstract:

Background: Approximately 120 genes have been identified as associated with either an monogenic form of idiopathic nephrotic syndrome (INS) or a kidney disease that phenocopies focal segmental glomerulosclerosis (FSGS). Systematic rare variant studies across the age of onset, response to therapy, and ancestries are still lacking. Advances in high-throughput genomic sequencing technologies provide an opportunity to assess the whole genetic landscape of INS, leading to more precise diagnosis, therapy and counseling. **Methods:** We conducted an exome sequencing (ES) study on 5,266 cases with INS caused by FSGS or minimal change disease (MCD), including 1,998 steroid sensitive cases (SSNS) and 3,268 cases those were not known to be responsive to steroids (either steroid resistant or untreated/unknown). A comprehensive clinical genomic screening was performed to identify “solved cases” harboring diagnostic/pathogenic Mendelian mutations in known FSGS genes. Per-gene burden of rare coding variants was assessed by exome-wide collapsing analysis comparing abovementioned cases and 28,637 population controls under dominant and recessive models. Analyses were conducted on the entire dataset and then again after removal of solved cases and *APOL1* high-risk genotypes. **Results:** In the clinical diagnostic analysis, we identified rare pathogenic variants in 616 (11.70%) cases, with the most frequent diagnoses attributable to *NPHS2*, *COL4A3*, *COL4A5*, *COL4A4*, *NPHS1*, *WT1*, and *INF2* variants, but with different distribution according to genetic ancestry (AFR having the lowest yield), age of onset (diagnostic rate double in pediatric than adults), and response to therapy (lower but non-zero yield in SSNS). Exome-wide collapsing analysis retrieved association for these known genes: *WT1* (best P= 1.09 × 10^{-23}; OR= 15.64), *COL4A5* (best P= 2.37 × 10^{-17}; OR= 9.83), *INF2* (best P= 1.87 × 10^{-11}; OR= 12.22), and others, under dominant models, and *NPHS1* (best P= 1.56 × 10^{-21}; OR= 25.18), *NPHS2* (best P= 3.93 × 10^{-15}; OR= 30.58), *SMARCAL1* (best P= 9.91 x 10^{-9}; OR= 31.97) and several others, under recessive models. This analysis re-classified *CD2AP* as an autosomal recessive cause of INS, and pointed to mutations in *CLCN5* and *OCRL* as common causes of FSGS phenocopies. Removal of solved cases and re-analysis prioritized 6 novel candidate genes: two exceeded the 5% false discovery rate based on 1,000 permutations (*FAM71B* and *SAV1*); four (*RAPGEF2*, *RALBP1*, *GNB1L* and *HMOX1*)

when mutated in the mouse result in glomerulopathy. **Conclusions:** These findings expand our understanding of the genetic underpinning of INS, identify novel candidate genes, and highlight the genetic heterogeneity of disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5018W: Ancestry-aware GWAS identifies novel common variants associated with susceptibility to disseminated coccidioidomycosis.

Authors:

S. Jensen[1], S. Spendlove[2], A. V. Stephens[1], K. Hou[1], R. Mester[1], G. R. Thompson[3], R. H. Johnson[4,1], A. Heidari[5,1,6], R. Kuran[4,1], B. Pasaniuc[1], H. Pimentel[1], M. Butte[1], V. A. Arboleda[1]; [1]UCLA, Los Angeles, CA, [2]Kaiser Permanente Molecular Genetic Pathology Lab., Los Angeles, CA, [3]UC Davis, Davis, CA, [4]Kern Med., Bakersfield, CA, [5]Dignity Hlth., Bakersfield Mem. Hosp., Bakersfield, CA, [6]Morehouse Sch. of Med., Atlanta, GA

Abstract:

Coccidioidomycosis, or Valley Fever, is a fungal infection affecting approximately 200,000 people in the United States yearly. While a majority of those infected are asymptomatic or have mild respiratory symptoms (Uncomplicated Valley Fever, UVF), some have more severe lung disease (Complicated Pulmonary Coccidioidomycosis, CPC) or extra-pulmonary infection (Disseminated Coccidioidomycosis, DCM). Epidemiologists have identified associations between patient self-identified race or ethnicity (SIRE) and dissemination risk. Meta-analysis of demographic information from 1,509 patients found that those who self-identified as African American or Black were 4.7 times more likely to have DCM than patients of other reported backgrounds (Fisher's Exact Test, p=2.78x10-13). While differences in health outcomes between SIRE groups are most often caused by environmental and socioeconomic factors, differences in sub-population allele frequencies of genetic variants due to shared evolutionary history can also contribute to disease risk, as with sickle cell anemia or Tay-Sachs disease.
We analyzed whole genome sequencing (WGS) or whole exome sequencing (WES) from 635 participants: 361 with UVF, 90 with CPC, and 184 with DCM. We used WGS for 2,782 samples from the 1000G, HGDP, and SGDP datasets as reference populations for genetic ancestry analyses, including principal components analysis (PCA) with smartPCA, unsupervised global ancestry estimation with ADMIXTURE, and local ancestry calculation with RFMIX. We then conducted genome-wide association studies (GWAS) with admix-kit using several ancestry-aware methods and investigated whether alleles associated with DCM were more common in a genetic ancestry.

Admixture mapping (ADM) GWAS identified 230 loci where DCM patients were significantly more likely to have alleles from one genetic ancestry than another, spanning a region of chromosome 2 that contains the lncRNA *EPCAM-DT* and genes *CALM2*, *STPG4*, and *TTC7A* ($p < 1 \times 10^{-5}$). Most significant variants found in this locus had higher allele frequencies in both DCM patients and in gnomAD African reference populations. *CALM2* and *STPG4* have previously been associated with immune functions, including regulation of IL-2 signaling. Variants associated with risk of dissemination were found in other regions, including variants in *PRELID2*, *ZFHX3*, and *ZMYND15* ($p < 5 \times 10^{-8}$). This is the first GWAS of coccidioidomycosis to identify any genome-wide significant associations, due to increased sample size and novel analysis methods. Identifying genetic risk factors for coccidioidomycosis dissemination will enable improved screening, prevention, and treatments.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5020W: *APOE4* stratified Alzheimer's disease polygenic risk score association with circulating proteins but not with circulating metabolites

Authors:

E. Drzymalla; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Background: Mechanisms underlying Alzheimer's disease (AD) are largely unclear, in part reflecting challenges in phenotype classification and biological pathway identification as well as a late disease onset. Multi-omics studies that integrate metabolomics, proteomics, and AD polygenic risk scores (PRS) can help address these challenges by measuring diverse biological mechanisms and identifying individuals at elevated AD risk (via AD PRS) decades before disease onset. Methods: AD PRS was constructed using European-derived weights from prior GWAS for 83 variants, excluding the *APOE4* locus. To validate AD PRS performance (continuous or dichotomized at the 90th %ile) in UK Biobank, we used Cox proportional hazard models with AD diagnosis, defined by ICD-10 G30 or F00 or ICD-9 331, as the outcome, AD PRS as the exposure, and adjusted for age, sex, education, and first ten principal components (PCs). To identify mechanisms by which increased AD genetic risk (continuous AD PRS, dichotomized at the 90th %ile AD PRS, or rs429358 *APOE4* AD risk variant alone) may increase AD incidence, we used linear regression for each 159 metabolites (measured using the Nightingale Health NMR platform) or 1463 proteins (Olink) outcomes on the AD PRS exposure, adjusting for age, sex, education, first ten PCs, and applying a 5% FDR correction. The analyzes were also stratified

by *APOE4* status (0, 1, or 2 variants). Results: With complete data, there were 253,438 UK Biobank participants (mean age = 57, 54.0% female, 1617 AD cases) for metabolites and 51,333 participants (mean age = 57, 54.0% female, 556 AD cases) for proteins. The AD PRS was validated with AD in UK Biobank (continuous PRS: per SD HR = 1.41, 95% CI = 1.35-1.48, p-value = $<2x10^{-16}$, binary PRS: HR = 1.92, 95% CI = 1.69-2.18, p-value = $<2x10^{-16}$). No metabolites were significant in the continuous or binary PRS model overall or upon *APOE4* stratification, but 23 and 8 proteins were associated with the continuous and binary PRS respectively in all participants. Interaction between the PRS and *APOE4* status was not significant for any protein. However, 148 (98%) metabolites and 312 (21%) proteins were associated with *APOE4*. Conclusions: *APOE4* is a strong predictor of AD and has been previously associated with lipid metabolites. The lack of association between the AD PRS and metabolites suggests the *APOE4* locus is the main driver between AD genetic risk and the metabolome, however, we analyzed a targeted metabolite platform. While the AD PRS was associated with proteins outside of the *APOE4* locus, highlighting known and potentially novel proteins, the *APOE4* locus was associated with 10x more proteins and is a strong correlate for proteome measures.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5021W: Application of Evolutionary Theory and Machine Learning to Discover Novel Risk Genes in Coronary Artery Disease

Authors:

A. Bigler, J. Asmussen, L. Danny, K. Wilhelm, O. Lichtarge; Baylor Coll. of Med., Houston, TX

Abstract:

Coronary artery disease (CAD) is the leading cause of death among cardiovascular diseases worldwide and is associated with both environmental and genetic risk factors. Known GWAS loci account for 15.5% of the disease risk, while the estimated disease heritability is between 40-60%. To discover novel CAD risk genes, we applied the evolutionary action (EA) equation to compute the likely functional impact of coding variants in whole exome sequencing (WES) of CAD patients in the UK Biobank (UKB). We used four complementary EA-based algorithms which incorporate evolutionary information and either statistical principles or machine learning to identify genes associated with CAD risk. We selected 15,657 CAD cases and 15,657 sex-matched, healthy controls in the UKB and applied the four algorithms to identify CAD-driver genes. Using the Cauchy-combination test to integrate our results, we prioritized 126 genes. These genes overlap with known CAD genes in the literature (n=11, p=2.3e-12) including PCSK9, APOE, LDLR, LPL, NOS3, MIA3, ADAMTS7, IL6R, WDR12, SH2B3, and SCARB1, and they diffuse to known CAD

genes in the STRING interaction network (AUROC=0.74, z=4.83). Among the prioritized genes, CACNA1A and FDPS have known interactions with anti-epileptic medications and bisphosphonates, respectively. In our cohort, CAD cases taking Gabapentin or Pregabalin (n = 581) have a 1.5-year reduced average age of onset of CAD (log-rank p=5.0e-5) compared to CAD cases not taking these medications (n=29,598). Conversely, CAD cases taking Risedronate sodium or Alendronate sodium (n = 528) have a 2.7-year delayed age of onset of CAD (log-rank p=5.6e-11) compared to those not taking these medications. To identify sex-specific CAD risk variants, we applied our algorithms to early onset males (n=2,294 cases) and early onset females (n=2,382 cases) in the UKB. We identified 24 variants in women and 37 variants in men with significant odds ratios (AF > 0.05). These variants were validated in early-onset CAD cohorts from All of Us where 5/24 variants in women and 7/37 variants in men maintained significance. Our findings highlight the power of these EA-based methods to recover known CAD biology and identify novel genes and variants associated with CAD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5022W: Application of polygenic risk scores to discriminate different types of diabetes in early onset diabetics and to predict the prediabetic stage in the general population

Authors:

J. Klovins, I. Atava, R. Rescenko, M. Briviba; Latvian BioMed. Res. and Study Ctr., Riga, Latvia

Abstract:

Background: Genetic factors play a crucial role in the stratification and etiology of multifactorial diseases, including Diabetes Mellitus. This stratification is vital for the diagnosis and prevention of these diseases. Our study aimed to evaluate the performance of polygenic risk scores (PRS) for type one (T1D) and type two diabetes (T2D) in differentiating diabetes types in early-onset diabetics and predicting the prediabetic stage in the general population.**Methods:** We combined whole genome sequencing (WGS) and array genotyping on 5,267 participants from the Genome Database of the Latvian Population (LGDB) to calculate PRS for T1D and T2D and identify rare inherited variants of early-onset diabetes. Our sample set included both general population participants and representative subsets of diabetic patients with diverse etiologies. **Results:** We first assessed previously published PRSs for metabolic diseases using disease-specific cohorts from the Latvian population and compared their performance with population-specific PRSs. We present two practical applications of PRS. Firstly, we found a strong correlation between the PRS for T2D and the age of onset for metabolic syndrome, suggesting

its utility as an independent metric for determining metabolic impairment. Secondly, we identified and confirmed MODY causative mutations (including novel variants) in 25 early-onset diabetes patients. Additionally, patients without genetically confirmed MODY had a significantly higher T1D risk compared to controls. The 75th percentile of T1D-PRS included only 8.7% of genetically confirmed MODY patients but encompassed 34% of patients without mutations, demonstrating good specificity for identifying indicative T1D within this PRS range. **Conclusion:** Our study showcases the advantages of WGS in identifying both monogenic variants and estimating polygenic risk, enhancing diagnostic precision for early-onset diabetes.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5023W: Assessment of a novel syndromic T2D subset based on IGF1 Resistance with high unmet medical need

Authors:

G. Andreoletti[1], W. Huynh[1], R. Gupta[2], A. Tebbe[1], K. Allison[1], E. Weiss[1], B. CUMMINGS[1]; [1]Marea Therapeutics, San Francisco, CA, [2]Harvard Med. Sch. / Broad Inst., Cambridge, MA

Abstract:

Insulin-like growth factor-1 (IGF-1) and its binding proteins are implicated in type 2 diabetes mellitus (T2D) and coronary heart disease (CHD). Gardner et al linked damaging missense variants in *IGF1R* with increased IGF-1 circulation to T2D risk, suggesting a role for IGF-1 resistance in its pathogenesis. This is the first rare variant genetic evidence highlighting the IGF-1 system's critical role in T2D and CHD, presenting it as a viable therapeutic target. This work suggests that populations with syndromic T2D with IGF-1 resistance may provide a proof of concept for drug response, underscoring the therapeutic potential of targeting this pathway. To evaluate if IGF1 resistance could constitute a new syndromic T2D population with unmet medical needs, we leveraged large-scale biobank datasets to replicate and further characterize the population carrying damaging *IGF1R* variants. To improve association analyses, we evaluated variants within biologically informed domains rather than relying solely on computational predictions for pathogenicity. We present independent replication of the association between T2D and *IGF1R* pathogenic missense variants across biobanks (UK Biobank $p = 3.0 \times 10^{-7}$ OR 1.9; Flannick *et al* (20,791 T2D) OR=1.5, p = 0.046). We additionally present results from a pheWAS of *IGF1R* variants, clinically informed to include features of *IGF1R* resistance-related phenotypes, such as acromegaly. We also assessed the unmet medical need in this IGF1-resistant diabetic population using primary care records. Our analysis reveals that 12.5% of *IGF1R* carriers with T2D have unmet medical needs. We show that

syndromic patients exhibit a higher severity of illness compared to non-syndromic patients and poor responsiveness to treatments, as evidenced by consistently high HbA1c levels over 6.5% across multiple visits, despite treatment with combinations of metformin and insulin. Finally, we functionally characterize *IGF1R* variants in the most severe patients to establish a lack of functional IGF1R signaling in these patients. Our findings demonstrate how to leverage rare genetic variants in biobank datasets to stratify common complex cardiometabolic disorders and define novel patient populations for therapeutic development. Incorporating multiple biobank datasets and functional variant characterization enhances our understanding of syndromic subtypes within complex disorders. Identifying patients with syndromic features highlights the potential to develop and repurpose drugs targeting the underlying genetic cause of disease, addressing unmet medical needs through personalized medicine based on genetic profiles.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5024W: Association of Single Nucleotide Polymorphism in OCT1 Gene in type II diabetes mellitus patients from North Indian population.

Authors:

H. Parveen; King George Med. Univ., Lucknow, Lucknow, India

Abstract:

Background: Diabetes mellitus has become a most important health problem worldwide in recent times. **Aims and objective:** To investigate the possible association of organic cation transporter- 1(OCT1) gene polymorphism (A-G) in type II Diabetes mellitus. **Methodology**: 100 diagnosed T2DM patients were recruited for the study and the genotypes for OCT-1 gene polymorphism using polymerase chain reaction (PCR) followed by restriction fragment length polymorphism (RFLP) were done. **Result:** We did not find any significant association between GG homozygous alleles ($p<0.079$) and AG heterozygous alleles ($p=0.209$) in type 2 diabetes mellitus patients compared to the control. It this case-control study, the frequency of the G allele of OCT-1 was found significant in alcoholic type 2 diabetes mellitus patients ($p=0.026$*). **In conclusion,** OCT-1 gene G allelic polymorphism is associated with type 2 diabetes mellitus patients in the north Indian population.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5025W: Associations between a polygenic predisposition to later pubertal timing and diagnostic odysseys for patients with abnormal growth

Authors:

J. Shelley[1], M. Shi[2], S. Van Driest[3], J. Simmons[2], J. Peterson[2], J. Mosley[2]; [1]Vanderbilt Univ. Sch. of Med., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]NIH, Bethesda, MD

Abstract:

Background: Delayed puberty can be a sign of underlying chronic illness or endocrine dysfunction, but in most patients, no underlying cause is found. Delayed puberty unrelated to disease, termed constitutional delay, is a diagnosis of exclusion that clusters in families and is thought to be polygenic. We hypothesized that a polygenic score for age at menarche (PGSAAM) could be used to identify patients with a benign, genetic predisposition to delayed puberty.
Methods: Using LDpred2, PGS_{AAM} weights were estimated for 979,457 SNPs using a GWAS for age at menarche in 329,345 subjects. We curated a retrospective cohort of 13,038 pediatric subjects of European ancestry with clinical data and genotyping in BioVU. We tested whether the PGSAAM associated with clinical diagnoses (PheWAS) and specialty referrals, orders for imaging or labs, and medication prescriptions related to the evaluation of delayed puberty (multivariable logistic regression). We then curated a cohort of 193 children and adolescents aged 8 and older with an evaluation by endocrinology for abnormal growth (defined as an ICD code for short stature or delayed puberty) and a delayed bone age (defined as a bone age more than 1 s.d. below the mean). Diagnoses were retrieved through chart review and classified as pubertal delay secondary to constitutional delay or pathologic causes. Logistic regression was used to test the association between the PGS_{AAM} and benign (vs. pathologic) delayed puberty. All analyses were adjusted for age, sex, and the top 5 principal components.
Results: Among pediatric participants in BioVU, a polygenic predisposition to later puberty (PGS_{AAM}) was associated with diagnoses of short stature (OR = 1.27 [95% CI: 1.2 - 1.35], p<0.001) and delayed puberty (1.50 [1.31 - 1.72], p<0.001) but not precocious (early) puberty (0.89 [0.78 - 1.02], p=0.09). Additionally, a predisposition to later puberty was associated with an endocrine evaluation for short stature or delayed puberty (OR = 1.28 [1.20 - 1.38], p<0.001), testing for testosterone among boys (1.25 [1.10-1.41], p<0.001), and prescriptions for testosterone among boys (1.32 [1.13 - 1.53], p<0.001). Among patients evaluated by endocrinology for delayed puberty, the PGS_{AAM} was not associated with a diagnosis of constitutional delay (1.28 [0.92-1.80], p=0.15).
Conclusion: We demonstrated that children with a polygenic predisposition to later pubertal timing are more likely to undergo evaluation for and diagnosis of delayed puberty. However, the

PGS_{AAM} did not associate with benign, constitutional delay among a subset of patients with both benign and pathologic delayed puberty.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5026W: Autism diagnosis at birth via functional analysis of IP_3receptor Ca^{+2} signaling in fibroblasts

Authors:

J. Gargus[1], D. Justus[1], E. Chao[2], I. Smith[2]; [1]NeuroQure, Irvine, CA, [2]UC Irvine, Irvine, CA

Abstract:

ASD has no defined biomarkers for diagnostics or novel drug discovery. A major cluster of ASD-associated loci encode calcium signaling proteins. Functionalizing these signals in fibroblasts showed only intracellular Ca^{2+} release altered in rare forms of monogenic ASD syndromes (fmr1, tsc1,tsc2), and a molecular defect in IP_3 channel gating was resolved with single-channel optical patch clamp showing all forms have a reduced mean open-time (an unstable flicker opening) and therefore low flux, while channel abundance, selectivity, conductance, and latency are unchanged, similar to kinetic changes seen in genetic channelopathies. A high-throughput FLIPR screen was used to image global Ca^{2+} signals from the channel, revealed as depressed signals from P2Y purinergic stimuli, first in a broader set of monogenic ASD fibroblasts, but also in fibroblasts from typical, sporadic ASD subjects, validated with research-grade ADOS testing. Several other monogenic brain disorders signaled like neurotypicals. iPSC- derived gaba neuronal precursors from ASD patient fibroblasts share this signaling defect. ROC curves can distinguish at 85% AUC high sensitivity and specificity, between syndromic and sporadic ASD samples, which signal with a similar small, slow wave in this assay, and the fast, large neurotypical signal. Therefore, IP_3R signaling appears to be at a node in a signaling pathway at which many forms of ASD are unified into a shared defective output that limits intracellular Ca^{2+} release. It is well established that IP_3R Ca^{2+} release flux is required to maintain proper mitochondrial functioning. Defects in mitochondrial bioenergetics are a phenotype long associated with ASD and abnormalities in mitochondria have recently been discovered via cortical brain imaging in areas responsible for social/facial stimuli. IP3R signaling impacts the full ASD phenotype including GI and immune dysfunctions and the excitation/inhibition ratio as the final step in axonal synaptic neurotransmitter release. The ER is also a key stress sensor, perhaps contributing to environmental inputs to ASD.
This IP3R functional defect is suggestive that ASD is yet another organellar disease, along with those of mitochondria, lysosomes and peroxisomes, and similarly diagnosable in fibroblasts. It

also establishes this IP_3R signal as a biomarker useful as a clinical trial surrogate outcome measure or entry criterion, and a screen for novel therapeutics. Most importantly it is now available as an adjunct early diagnostic, greatly shortening diagnostic odyssey and allowing early intervention with therapy shortly after birth , long before many major developmental milestones are missed.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5027W: Brain proteome of well-being: A proteome-association study unveils novel proteins related to increased well-being

Authors:

N. Enduru, O. Donmez, B. Fernandes, Z. Zhao; The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract:

Background: Well-being is a complex and multifaceted construct comprised of the presence of positive emotions and moods, the absence of negative emotions, satisfaction with life, fulfillment, and positive functioning, being a dimension of mental life beyond the absence of disease. Well-being is a moderately heritable and stable trait, and recent GWAS on different aspects of well-being showed an association with several genes, suggesting that well-being has a biological component. Proteins are the final products of gene expression and are the main functional components of cells and biological processes, and protein levels of proteins expressed in the brain can potentially influence well-being. We aimed to investigate if genetic variants associated with well-being exert their effect through altering brain protein expression levels. With that in mind, we performed a protein-wide association study (PWAS) to identify protein brain expression potentially mediating the relationship between genetic variants and well-being phenotypes. After, we performed a transcriptome-wide association study (TWAS) to provide a second layer of confidence that the discovered brain proteins contribute to well-being. **Methods**: We obtained GWAS summary statistics on subjective well-being, well-being spectrum, positive affect, and life satisfaction). For the PWAS analyses, we used cis-regulated protein weights from the postmortem human dorso pre-frontal cortex (dPFC) brain proteomes from ROSMAP (n = 376 individuals, n = 12,691 proteins quantified) and Banner (n = 152 individuals, n = 11,518 proteins quantified) as the discovery and validation datasets, respectively. For the TWAS, we used brain mRNA weights from AMP-AD (783 participants) and CMC (452 participants). All analyses were corrected for multiple comparisons, and p-values < 0.05 were considered statistically significant. **Results**: We identified 121 potentially causal genes

that might affect well-being by regulating their brain protein abundances in the discovery PWAS. Of those, 17 genes whose cis-regulated brain protein abundances were associated with well-being in the discovery PWAS were validated in the replication PWAS, in addition to having their cis-regulated mRNA levels also associated with well-being in the TWAS analysis: *ALDH16A1, B3GALTL, CDH13, CSDC2, CSE1L, GMPPB, GPX1, LRP4, MGST1, MMAB, PACSIN3, PLEKHB1, SHMT1, TREX1, and VKORC1.* **Discussion**: We identified several proteins expressed in the brain that contribute to well-being, providing valuable insight into its underlying biological mechanisms. If validated, these proteins could be promising drug targets for increasing well-being.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5028W: Characterization of the obese proteome and downstream sequalae in Hispanic Latino study participants

Authors:

M. Yaser[1], M. Graff[1], E. Frankel[2], R. Roshani[2], W. Zhu[2], H-H. Chen[2], P. Gordon-Larsen[1], M. Krishnan[1], L. Petty[2], V. Buchanan[3], M. Lee[4], A. D. Guiterrez[4], H. Highland[1], K. Young[1], J. McCormick[4], S. Fisher-Hoch[4], J. Below[5], K. North[1]; [1]Univ North Carolina, Chapel Hill, NC, [2]Vanderbilt Univ., Nashville, TN, [3]Univ of North Carolina, Chapel Hill, NC, [4]UTH SPH Brownsville, Brownsville, TX, [5]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

Background: Obesity is a major driver of metabolic diseases (e.g., type 2 diabetes [T2D], cardiovascular diseases [CVD] etc.) that disproportionately impact historically marginalized populations, such as self-identified Hispanic/Latino (H/L) populations. H/L are generally underrepresented in multiomic studies, designed to enable mechanistic insights into inherent and mutable factors that impact obesity-related metabolic disease. Accordingly, we assessed the association and the broader clinical significance of circulating proteins with obesity traits in 259 self-identified Mexican Americans from the Cameron County Hispanic Cohort (CCHC). **Methods**: We performed mixed model differential abundance analyses (DAA) with 2,921 plasma proteins quantified using Olink panel and 13 obesity-related phenotypes including anthropometrics and dual-energy x-ray absorptiometry body composition. For proteins significantly associated with obesity, we performed colocalization, and 2 sample Mendelian Randomization (2SMR) using independent samples. We additionally performed 2SMR with ten obesity-related clinical traits in independent populations including coronary heart disease, type 2 diabetes, chronic kidney disease, low-density lipoprotein, hypertension, colorectal cancer,

Alzheimer's disease, heart failure, atrial fibrillation, and breast cancer. Lastly, we functionally annotated top proteins using tissue/cell gene expression and overrepresentation analyses. **Results**: One hundred twenty-two plasma proteins were associated with at least one obesity related trait (at false discovery rate <0.05). Twenty-two proteins colocalized with either obesity related traits or clinical disorders associated genes at posterior probability >70%. In 2SMR analyses, 39, 22, and 23 proteins displayed nominal evidence of causal, consequential, and bidirectional associations with obesity related traits, respectively. Sixty-six proteins displayed significant associations with downstream disease. At least 30 proteins were associated with the incidence of T2D and dyslipidemia prospectively. Genes encoding proteins potentially causally associated with overall obesity displayed high levels of RNA expression in brain tissues, whereas those associated with central/visceral adiposity traits displayed high RNA expression levels in adipose tissues and arteries. **Conclusion**: Key proteins with potentially causal roles in obesity-related disease sequelae were identified which can be potential targets for biomarkers and clinical intervention. Colocalization and 2SMR results suggest findings are generalizable across populations.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5029W: Characterizing the genetic architecture of Autism from a multi-ancestry perspective ⭐

Authors:

M. Mendes de Aquino[1,2], C. Xu[1,2], W. Engchuan[1], B. Trost[1,2], B. Thiruvahindrapuram[1,2], T. de Paiva[3], X. Zhou[1,2], E. Tarazona-Santos[4], V. Borda Pua[5,6], S. W. Scherer[1,2,7,8]; [1]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [2]Genetics and Genome Biology Program, The Hosp. for Sick Children, Toronto, ON, Canada, [3]Núcleo de Pesquisas em Ciências Biológicas (NUPEB) Univ.e Federal de Ouro Preto (UFOP), Ouro Preto, Brazil, [4]Departamento de Genética, Ecologia e Evolução, Inst. de Ciências Biológicas, Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil, [5]Inst. for Genome Sci. - Univ. of Maryland, Baltimore, MD, [6]Inst. for Hlth.Computing - Univ. of Maryland, Baltimore, MD, [7]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [8]McLaughlin Ctr., Toronto, ON, Canada

Abstract:

Utilizing genetic variant data from one continental population restricts the discovery of novel Autism Spectrum Disorder (ASD) variants crucial for both clinical diagnostic testing, as well as in discovery research, mainly due to differences in allele frequency and linkage disequilibrium patterns.To extend genomic advances in ASD research to diverse populations, we applied a

genome-wide Admixture Mapping (AM) approach in an integrative way with rare CNVs and SNVs. We used whole-genome sequencing data from ASD probands in MSSNG, SSC and SPARK cohorts, to conduct AM using two approaches to capture genetic diversity: (i) admixed North American samples (1,747 cases, 1,606 controls) from 5 ancestral origins, and (ii) Latin American samples (1,505 cases, 3,214 controls) from 3 ancestral origins, including controls from a Brazilian cohort. We utilized reference datasets from the 1KGP and the Peruvian Genome Project. Acknowledging the growing recognition of rare (<1%) variants in ASD, we assessed the frequency of rare CNV deletions and predicted damaging SNVs impacting ASD-relevant genes identified by AM in ASD-probands compared to their parents. Utilizing the five-way admixture approach, we identified loci significantly associated with ASD from the African ancestry (chr1p21.2, p-value=9.65×10^{-5}), and another to Native American ancestry (chr11q13.4, p-value=5.54×10^{-5}), encompassing *DHCR7* (SFARI 1S) and *SHANK2* (SFARI 1S, EAGLE 18.55), respectively. Fine-mapping of chr11 implicated rs77695321, exhibiting genome-wide significance (p-value=1.10×10^{-8}, OR=0.832474), and similar allele frequencies in Americans (0.14) and Europeans (0.10). In the Latin data, two Amerindigenous ancestry loci were identified: chr1p33-p32.3 (p-value=5.94×10^{-5}), encompassing *AGBL4* (SFARI 2) and chr11p15.2 (p-value=7.78×10^{-5}), covering *ARNTL*, *FAR1* and *PLEKHA7*. Rare variants significantly differed in frequency across ancestries with *AGBL4* CNV frequency being highest in Americans (1.76) and Europeans (1.15). Predicted damaging SNVs frequency was also higher in Americans for *DHCR7* (0.55) and *OSBPL9* (0.21), corroborating the AM findings. Current ASD clinical diagnostic panels primarily rely on genes identified in European populations. Among the 8 AM-associated genes linked to American ancestry and featuring rare exon deletions and damaging SNVs in ASD probands, only 2 (*SHANK2* and *DHCR7*) are included in Latin population diagnostic panels. This underscores the necessity of considering genetic ancestry in ASD research, which can facilitate the development of polygenic scores applicable across populations and tailored diagnostic panels.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5030W: Circulating Mfge8 levels increase the risk of human T2DM and Cardiovascular disease

Authors:

D. Sanghera, M. Rout, M. W. Malone-Perez, G. Park, M. Lerner, J. Frazer, B. Apple, A. Vaughan, M. Payton, S. Stavrakis, K. A. Fung; Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Abstract:

MFGE8 is a major exosome (EV) protein known to mediate inflammation and atherosclerosis in type 2 diabetes mellitus (T2DM) in animal studies. The pathophysiological role of this protein in obesity, T2DM, and cardiovascular disease is controversial and less investigated in humans. Earlier, we reported a rare Asian Indian population-specific missense variant (rs371227978; Arg148His) in the MFGE8 gene associated with increased circulating Mfge8 and T2DM. Here, we have further investigated the role of Mfge8 with T2DM risk in additional Asian Indians (n=4897) and Europeans and other multiethnic cohorts from UK Biobank (UKBB) (n= 455,808) and the US (n=1150). We also evaluated the exposure of Mfge8-enriched human EVs in zebrafish (ZF) for their impact on the cardiometabolic organ system. Most individual carriers of Arg148His variant not only had high circulating Mfge8 but also revealed a positive significant correlation with glucose ($r= 0.42$; $p = 4.9x10^{-04}$), while the non-carriers showed a negative correlation of Mfge8 with glucose ($r = -0.38$; $p = 0.001$) in Asian Indians. However, the same variant was monomorphic in Europeans and other ethnic groups of UKBB and US cohorts. Even without the Arg148His variant, serum Mfge8 correlated significantly with blood glucose in other non-South Asian ethnicities ($r = 0.47$; $p = 2.2x10^{-13}$). Since Mfge8 is an EV marker, we tested the exposure of Mfge8-enriched human EVs to ZF larvae as an exploratory study. The ZF larvae showed rapid effects on insulin-sensitive organs by developing fatty liver disease and heart hypertrophy and exhibiting redundant growth with poor muscular architecture with and without the high-fat diet (HFD). In contrast, the control group fishes developed fatty liver disease and heart hypertrophy only after the HFD feeding. Backed with strong support from animal studies on the role of Mfge8 in obesity, insulin resistance, and atherosclerosis, the current research suggests that circulating Mfge8 may become a potential marker for predicting the risk of T2DM and cardiovascular disease in humans. **Funding** The Asian Indian Diabetic Heart Study/Sikh Diabetes Study was supported by National Institute of Health grants-R01DK082766 and R01DK118427 (National Institute of Diabetes and Digestive and Kidney Diseases, NIDDK) and College of Medicine Alumni Association (COMAA), the Presbyterian Health Foundation, and the Leinbach Foundation grants, and from Dr. Geoffrey Altshuler Endowment funds from the Children's Health Foundation of the University of Oklahoma Health Sciences Center.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5031W: Clinical Variability Among Three Siblings with Tubulin-Folding Cofactor D-Related Encephalopathy: Further Delineation of Clinical Presentation

Authors:

A. Alghamdi[1], A. Shawli[2], R. Abualnaja[3], A. Alsudais[1], S. Almuntashiri[1], M. Bhader[1]; [1]King Saud bin Abdulaziz Univ. for Hlth.Sci., Jeddah, Saudi Arabia, [2]King Khalid Natl. Guard Hosp., Jeddah, Western, Saudi Arabia, [3]King Saud bin Abdulaziz Univ. for Hlth.Sci., Makkah, Saudi Arabia

Abstract:

Background: Tubulin-folding cofactors D-related encephalopathy (PEBAT) is a rare neurogenetic disorder associated with mutations in the TBCD gene. This disorder leads to progressive encephalopathy, brain atrophy, and thin corpus callosum, resulting in a spectrum of neurological manifestations. However, limited information is available regarding the clinical features, severity of he symptoms, diagnostic challenges, and management strategies of PEBAT, particularly in clinical practice. **Objectives:** This paper aims to provide a comprehensive analysis of the clinical characteristics, diagnostic investigations, genetic findings, and outcomes of patients with PEBAT. Presenting these cases contribute to the growing body of knowledge on this rare neurogenetic disorder and enhance understanding of its clinical presentation and management. **Study Design:** retrospective review involving three Saudi siblings diagnosed with PEBAT. Clinical data, including patient demographics, medical history, and physical examination findings, were collected. Diagnostic investigations, including EEG, CT scan, and MRI, were reviewed, and genetic analysis was performed to identify variants in the TBCD gene. Detailed laboratory findings were also included. **Methods:** Clinical presentations, diagnostic investigations, genetic analyses, and outcomes were analyzed for each patient. Descriptive statistics were used to summarize the data. Genetic variants were classified according to established guidelines. **Results:** The three patients presented with various neurological manifestations, including seizures, cognitive impairment, motor and social development regression, delayed speech development, and microcephaly. Diagnostic investigations revealed abnormalities consistent with PEBAT. Genetic analysis identified homozygous variants in the TBCD gene, classified as likely pathogenic. The patients exhibited varying outcomes, with controlled seizures, neurocognitive delays, and refractory epilepsy. **Conclusion:** The outcomes varied among patients, highlighting the need for individualized diagnostic tools and management approaches. Further research is necessary to elucidate the clinical significance and implications of these genetic mutations in relation to the patients' neurodevelopmental regression and associated features.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5032W: Comparisons of subcutaneous adipose and plasma metabolite associations with BMI and with genome-wide genetic variation

Authors:

E. Hill-Burns[1], X. Yin[2], H. Stringham[3], A. Locke[4], C. Fuchsberger[5], M. Erdos[6], L. Bonnycastle[6], a. etheridge[7], S. Vadlamudi[8], C. King[8], F. Collins[9], C. Burant[1], E. Fauman[10], M. Boehnke[3], M. Laakso[11], K. Mohlke[8], L. Scott[1]; [1]Univ Michigan, Ann Arbor, MI, [2]Nanjing Med. Univ., Nanjing, China, [3]Univ. of Michigan, Ann Arbor, MI, [4]Regeneron Genetics Ctr., Tarrytown, NY, [5]Eurac Res., Bolzano, Italy, [6]NHGRI/NIH, Bethesda, MD, [7]Univ. of north carolina at chapel hill, chapel hill, NC, [8]Univ North Carolina, Chapel Hill, NC, [9]NIH, Bethesda, MD, [10]Pfizer, Cambridge, MA, [11]Univ. of Eastern Finland, Kuopio, Finland

Abstract:

Subcutaneous adipose tissue (SAT) maintains lipid energy balance, and SAT composition and function modulate an individual's risk of developing cardio-metabolic diseases (CMD). Connections between SAT metabolic state and cardiometabolic health are less well understood. Our goal was to understand the relationships of SAT metabolite levels with BMI and genetic variation, and to compare these findings to plasma-based analyses.

In 805 Finnish men (METSIM study, SAT biopsy age: 65±5 yrs; plasma collection age: 55±5 yrs), we quantified levels of 722 SAT and 1,540 plasma (previous work) metabolites using the Metabolon platform. For 568 SAT and 1,145 plasma metabolites with ≥50% non-missing values, we tested for BMI-metabolite associations, adjusting for age, batch, and median sample metabolite level (FDR<5%). We found significant associations with BMI for 249 SAT and 432 plasma metabolites. The SAT lipid metabolites were enriched for BMI associations (59% of 278 tested lipids), and plasma amino acids were enriched for BMI associations (47% of 196 tested amino acids). Of 445 metabolites tested in both tissues, 24% were significantly associated with BMI only in adipose, 17% only in plasma, and 22% in both tissues (78% of which had effects in the same direction).

We tested for association between genome-wide genetic variants and metabolite levels in SAT and plasma in 704 whole genome sequenced samples. We included variants with minor allele count ≥15, and adjusted for age and batch in both, metabolite median in SAT, and lipid medications in plasma (Bonferroni corrected in each tissue for independent tests, $5x10^{-8}/228$ for SAT and $5x10^{-8}/327$ for plasma). Levels of 11 SAT and 118 plasma metabolites were significantly associated with one or more genetic variants.

We detected shared SAT and plasma variant-associations for two amino acids and one lipid. In SAT, four sphingomyelins and one ceramide were associated with variants

in/near *CERS4* (Ceramide Synthase 4, min $p=4x10^{-29}$), likely reflecting synthase activity in SAT; in plasma, these lipids were significantly associated only in a larger METSIM sample (N=9,413). Two of the significant SAT metabolites were not detected in plasma; one, nicotinamide ribonucleotide (NMN), was associated with variants in the *BST1*/*CD38* gene region (min $p=2x10^{-13}$). *BST1* and *CD38* are homologous genes (37% amino acid sequence identity) which both encode enzymes capable of converting NAD+ to nicotinamide, which can then be converted to NMN.
We found both unique and shared associations of SAT and plasma metabolites with BMI and genetic variants, highlighting the value of performing metabolomics in multiple tissues.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5033W: Comprehensive analysis of the genetic variation in the *LPA* gene from short-read sequencing

Authors:

R. Betschart[1], G. Koliopanos[1], P. Garg[2], L. Guo[3], M. Rossi[4], S. Schönherr[5], S. Blankenberg[3], R. Twerenbold[3], T. Zeller[3], **A. Ziegler**[1]; [1]Cardio-CARE, Davos Wolfgang, Switzerland, [2]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, Hess Ctr. for Sci. and Med., New York, NY, [3]Dept. of Cardiology, Univ. Heart and Vascular Ctr. Hamburg, Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, [4]Illumina Inc, San Diego, CA, [5]Inst. of Genetic Epidemiology, Med. Univ. of Innsbruck, Innsbruck, Austria

Abstract:

Lipoprotein (a) [LP(a)] is a risk factor for cardiovascular diseases and mainly regulated by the complex *LPA* gene. We investigated the types of variation in the *LPA* gene and their predictive performance on LP(a) concentration. We determined the Kringle IV-type 2 (KIV-2) copy number (CN) using the DRAGEN LPA Caller (DLC) and a read-depth based CN estimator in 8351 whole genome sequencing samples from the GENESIS-HD study. The pentanucleotide repeat in the promoter region was genotyped with GangSTR and ExpansionHunter. LP(a) concentration was available in 4861 population-based subjects. Predictive performance on LP(a) concentration was investigated using random forests. The agreement of the KIV-2 CN between the two specialized callers was high (r=0.9966; 95% confidence interval [CI] 0.9965-0.9968). Allele-specific KIV-2 CN could be determined in 47.0% of the subjects using the DLC. Lp(a) concentration can be better predicted from allele-specific KIV-2 CN than total KIV-2 CN. Two single nucleotide variants 4925G>A and rs41272114 further improved prediction. The genetically complex LPA gene can be analyzed with excellent agreement between different callers. The allele-specific KIV-2 CN is

more important for predicting LP(a) concentration than the total KIV-2 CN. It would be important that the allele-specific KIV-2 CN is determinable in all subjects.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5034W: Comprehensive Genomic Analysis of Healthy Aging in Super Seniors

Authors:

R. Hoque[1], S. Leach[1], O. Vishnyakova[1], A. Brooks-Wilson[2]; [1]Canada's Michael Smith Genome Sci. Ctr., BC Cancer, Vancouver, BC, Canada, [2]Dept. of BioMed. Physiology and Kinesiology, Simon Fraser Univ., Burnaby, BC, Canada

Abstract:

The Super Seniors Study aims to identify genetic factors that influence healthy aging. 'Super Seniors' are individuals 85 years of age or older who have never been diagnosed with cancer, cardiovascular or major pulmonary disease, diabetes or dementia. We conducted genome wide association and pathway analysis of the Super Seniors' phenotype to identify variants and biological pathways for which genetic variation may contribute to healthy aging. The analysis was performed using 6.6 M imputed variants in individuals of European ancestry. Cases were 541 Super Seniors; controls were 373 population-based mid-life individuals ascertained without regard to health status. Genome-wide association study (GWAS) was conducted using an additive model and logistic regression, adjusting for sex and genetic principal components. Thirty-one variants were associated with healthy aging under the additive model, with p-values ≤ 1e-5. These include variants from the genes IFIH1, LPCAT3, and OR10P1, which have been previously associated with type 1 diabetes mellitus, plasma phospholipid fatty acid proportions, and systolic blood pressure, respectively. A two-step process was employed to conduct genome-wide pathway analysis (GWPA). In the first step, gene-based associations were analyzed. In the second step, scores from the first step were utilized to test for pathway-based associations, comparing genes within a specific pathway to the rest of the genome. Gene sets for pathway analyses were defined using the KEGG database. The pathways were ranked based on their P values. Top ranked pathways included several relevant to healthy aging or longevity, including Alzheimer disease (P = 0.003), P53 signaling (P = 0.004), insulin signaling (P = 0.008), Type 1 diabetes mellitus (P = 0.01), and Huntington disease (p = 0.02). While the results from both the GWAS and GWPA did not achieve genome-wide significance, the identification of variants and pathways with known relevance among the lowest p-values suggests that several genetic

variants could play a role in promoting healthy aging within this distinctive cohort of Super Seniors.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5035W: Contribution of copy number variants to schizophrenia in East Asian populations

Authors:

Y. Chen[1], Q. Feng[2], M. Lam[3], M. Yu[4], Y. Sun[5], O. Shanta[6], B. Thiruvahindrapuram[7], Y-W. Jen[8], J. Wang[1], J. Xu[9], Stanley Global Asia Initiatives[13], W. Shi[9], S. Scherer[10], F. Zhu[11], C-M. Liu[12], Z. Guo[1], D. Howrigan[13], D. Howrigan[14], M. Daly[14], B. Neale[14], A. Sawa[15], J. Sebat[6], J. Tang[16], X. Ma[17], W-j. Chen[12], S. Qin[18], W. Yue[5], T. Ge[14], H. Huang[14]; [1]Broad Inst., East Cambridge, MA, [2]MGH, Worcester, MA, [3]Inst. of Mental Hlth., Singapore, Singapore, [4]Broad Inst., Cambridge, MA, [5]Peking Univ. Sixth Hosp., Peking Univ. Inst. of Mental Hlth., NHC Key Lab. of Mental Hlth.(Peking Univ.), Natl. Clinical Res. Ctr. for Mental Disorders (Peking Univ. Sixth Hosp.), Beijing, China, [6]UC San Diego, San Diego, CA, [7]The Hosp. for Sick Children, Peter Gilgan Ctr. for Res. and Learning, Toronto, ON, Canada, [8]Inst. of Epidemiology and Preventive Med., Coll. of Publ. Hlth., Taibei, Taiwan, [9]Digital Hlth.China Technologies Co, China, China, [10]The Hosp. for Sick Children, Toronto, ON, Canada, [11]Xi'an Jiaotong Univ., Belmont, MA, [12]Natl. Taiwan Univ. Hosp., Taibei, Taiwan, [1314]Massachusetts Gen. Hosp., Boston, MA, [15]Johns Hopkins Univ., Baltimore, MD, [16]The 2nd Xiangya Hosp. of Central South Univ, Changsha, China, [17]The First Affiliated Hosp. of Xi'an Jiaotong Univ., Xi'an, China, [18]Shanghai Jiao Tong Univ., Shanghai, China

Abstract:

Studies have established an important role for rare copy number variants (rCNVs) in the etiology of schizophrenia, showing that rCNVs have larger effect sizes (OR of 4-100) compared to common variants (OR<1.3). However, most studies have been conducted on individuals of European descent (EUR). The extent to which rCNVs contribute to schizophrenia in non-EUR populations remains unclear. Here, we report findings from the largest rCNV study in individuals of East Asian ancestry (EAS), including 20,903 schizophrenia cases and 23,258 controls.
We found that genome-wide rCNV burden is significantly elevated in EAS individuals with schizophrenia compared to controls (OR = 1.021 [1.015 - 1.027], P=2.74E-12). A comparison between EAS and EUR populations found that the genome-wide rCNV effect on schizophrenia is generally consistent across populations. Notably, the rCNV burden remains significant after excluding previously reported rCNVs in the EUR population (OR= 1.012 [1.006 - 1.018], P=1.60E-05), suggesting the presence of schizophrenia-associated rCNVs in EAS that were not discovered

in previous EUR studies.
Subsequent association analysis identified nine rCNV loci significantly associated with schizophrenia in the EAS population, six of which reached genome-wide significance for the first time, including deletions at 8p21.3, 11q13.1, 19p13.3, and 19q13.42 and duplication at 15q11.2 and 15q12. These novel rCNVs have a much higher carrier frequency in the EAS population compared to the EUR population, underscoring the pressing need to expand the population diversity in genetic studies to fully capture the genetic risk factors underpinning schizophrenia.
A meta-analysis across EAS and EUR cohorts, with a total sample of 38,409 cases and 40,009 controls, increased the number of rCNV loci associated with schizophrenia at genome-wide significance from 8 (from the largest published EUR study) to 15. We found that genes within these 15 rCNV loci had significantly elevated intolerance to protein-truncating variants across EAS and EUR populations.
In summary, for the first time, we found that the genome-wide burden of rCNVs is comparable between EAS and EUR populations. At the locus level, rCNVs have a consistent effect on schizophrenia risk across the two populations. These observations support a broadly similar genetic impact of rCNVs on schizophrenia between EAS and EUR ancestries. This large-scale schizophrenia rCNV study represents a vital advancement in schizophrenia genetics by uncovering novel rCNV loci and providing insights into the comparative genetic architecture of rCNVs in schizophrenia across populations.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5036W: Cross-population statistical fine-mapping and exome-wide association study identify causal condensate variants for type 2 diabetes

Authors:

R. Tian[1], A. Domingues[2], F. Carpenter[1], I. Klein[1]; [1]Dewpoint Therapeutics, Boston, MA, [2]Dewpoint Therapeutics, Dresden, Germany

Abstract:

Disease-causing mutations in protein intrinsically disordered regions (IDRs) perturb normal phase separation, leading to aberrant condensate formation and subsequent dysregulation of biological processes and pathogenesis. Type 2 diabetes (T2D) is a common and polygenic disease, with over 600 risk loci identified by GWAS. However, the role of IDR mutations in T2D remains largely unknown.
To identify both common and rare T2D causal variants in IDRs, we used ‘metapredict’ to predict

18,985 protein IDRs based on their amino acid sequence on the canonical transcript, followed by cross-population fine-mapping of 611 T2D risk loci, along with gene-level and single-variant exome-wide association studies of T2D in the UK Biobank European population.
Common-variant statistical fine-mapping was performed using SuSiEx, integrating ancestry-specific GWAS from European (N_{cases}=242,283, $N_{controls}$=1,569,734), East Asian (N_{cases}=88,109, $N_{controls}$=339,395) and South Asian (N_{cases}=16,832, $N_{controls}$=33,767) populations. SuSiEx identified a total of 23,681 putatively causal variants across 574 loci, 408 of which have posterior inclusion probability (PIP)>0.5. Among these, 34 are missense variants, with 7 located in IDRs. Apart from rs2229840 in *NCOR2* and rs1042522 in *TP53*, the remaining 5 variants (rs1800961 in *HNF4A*, rs41277236 in *NEUROG3*, rs2066827 in *CDKN1B*, rs12437434 in *NYNRIN*, rs9379084 in *RREB1*) have all been previously reported as T2D causal variants and validated in functional studies, suggesting the involvement of condensate mechanisms in T2D.
We further performed single-variant and gene-based association tests using REGENIE to examine the association between T2D risk and single variants and aggregated variant burden within IDRs (N_{cases}=37,484, $N_{controls}$=395,843). We primarily focused on protein-truncating and missense variants. T2D cases were defined by combining information on biomarkers (hbA1c >= 48), self-reported and ICD9/10 coded diagnosis, doctor diagnosis and diabetes medication usage. Samples with type 1 diabetes were excluded. While we did not identify any IDR-enriched risk genes reaching exome-wide significance (P<2.5e-6), we identified 150 exome-wide significant rare variants (minor allele frequency <0.001; P<8e-9), 23 of which across 23 genes are located within IDRs.
In conclusion, both common and rare T2D-associated variants were found within IDRs, and in some instances, they converged on the same gene (e.g. *HNF4A*), suggesting a potential role of condensate in T2D pathogenesis. We plan to functionally validate candidate condensate genes and causal IDR mutations in disease-relevant cell models.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5038W: Decoding the Genetic Basis of Autoimmune Gastritis and Pernicious Anemia.

Authors:

B. Morris, A. Reynolds; Baylor Univ., Waco, TX

Abstract:

Autoimmune gastritis (AIG) is a condition characterized by inflammation of the stomach lining due to autoimmune destruction of parietal cells that affects between 0.5-2.5% of the US

population. Parietal cells are specialized epithelial cells, found in the fundus of the stomach, aid in digestion and nutrient absorption by secreting hydrochloric acid and intrinsic factor. AIG leads to multiple disease end-stage phenotypes, including iron deficiency anemia, pernicious anemia, and increases the risk of developing gastric cancers; but it is unclear why some patients develop different end-stages of the disease and what risk factors influence this. Pernicious anemia (PA), one end-stage of AIG, causes malabsorption of vitamin B12 (cobalamin), an essential micronutrient that helps maintain healthy blood cells, nerves, DNA synthesis and structural stability, methyltransferase activity, and many other important metabolic processes. Quick diagnosis of PA is vital, as left untreated, the condition can cause permanent neurological damage and even death. However, a positive diagnosis of PA can be difficult because of slow progression and nondescript symptoms that can mask the underlying disease, leaving many patients undiagnosed or misdiagnosed for 5 - 10 years. To better understand the genetic risk factors of AIG and PA we used whole-genome sequence data from the All of Us research program to conduct a genome-wide association study (GWAS) on 3761 AIG cases, 979 PA WGS cases, and 245,388 healthy controls. Using a statistical fine-mapping approach, we identified credible sets of putatively causal variants associated with AIG and PA. We compared our findings with previously published GWAS of other autoimmune diseases to shed light on common autoimmune disease risk factors. These findings contribute to ongoing efforts to characterize polygenic and pathway risk scores for AIG, PA, and other autoimmune diseases, laying the foundation for future improvements in clinical guidelines and diagnostic and therapeutic strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5039W: Decreasing polygenic prediction accuracy with age is driven by age-dependent genetic and non-genetic disease liability

Authors:

A. Durvasula[1], X. Jiang[2]; [1]Univ. of Southern California, Los Angeles, CA, [2]Cambridge Univ., Cambridge, United Kingdom

Abstract:

Polygenic risk score (PRS) accuracy decreases with age for many complex diseases (Jiang et al 2021 *PLOS Genet*), but the cause is unclear. Here, we propose three mechanisms to explain the decreasing accuracy under a unified age-dependent liability-threshold model with three parameters: age-specific heritability, genetic correlation between young and old groups, and non-genetic liability correlation between young and old groups. Using this model, we attribute

the decreasing PRS prediction accuracy with age (ΔR^2)to three sources: 1) decreases in SNP-heritability (i.e. GxAge), 2) changes in the genetic liability and non-genetic liability correlation across age, and 3) changes in the calibration of the PRS with age. Extensive simulations indicate that our approach accurately estimates the contributions of these distinct mechanisms.
We applied our framework to 7 disease-quantitative trait pairs (average N=180K) in the UK Biobank; examples of pairs include type 2 diabetes and HbA1c as well as hypertension and systolic blood pressure (SBP). We analyzed electronic health record (combined primary care and hospital) data to identify the age of diagnosis. We confirmed that PRS accuracy decreased significantly with age for all diseases (average ΔR^2across traits= -24.5% ±1.5% per 10 years). Applying our model, we first attributed 16% of ΔR^2 to decreases in SNP-heritability of the disease liability; for example, we found SNP-heritability for LDL decreased 36.9%±1.8% in the oldest age bin compared to the youngest age bin, implicating changes in genetic liability with age as a driver of ΔR^2. Second, we attributed 59.8% of ΔR^2 to the difference of genetic liability and non-genetic liability correlation between young and old groups. On average, the genetic correlation between young and old groups is 1.36x±0.03 higher than the phenotypic correlation between young and old group. We estimate this mechanism will cause a decrease in PRS prediction R^2 of 14.7%±1.9% on average under a liability threshold model. Third, we found differences in the calibration of the PRS between young and old groups for LDL and SBP where the youngest bin has more extreme phenotype values in the highest PRS bin compared with the oldest group. For example, observed minus expected SBP in the top PRS decile is 1.91x higher in the youngest bin compared to the oldest bin, implicating changes in PRS calibration. Taken together, our results suggest decreases in PRS accuracy with age are mainly driven by differences between the genetic liability and non-genetic liability correlation between young and older groups, and partially driven by the age-dependent heritability (GxAge).

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5040W: Delineating ancestry-specific genetic patterns in type 2 diabetes: Implications for treatment efficacy from real-world and trial data

Authors:

A. Dawed[1], C. Maroteau[1], K. E. Hanghøj[2], M. S. Rasmussen[3], J. M. Justesen[3], J. M. Howson[1], Y. Jamshidi[1]; [1]Dept. of Genetics, Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom, [2]Integrated Omics, Novo Nordisk, Søborg, Denmark, [3]Biostatistics Omics, Novo Nordisk, Søborg, Denmark

Abstract:

Significant advancements in our understanding of type 2 diabetes (T2D) have revealed a diverse range of clinical profiles, disease progressions, and therapeutic outcomes. Although distinct genetic subgroups within T2D have been identified, their validation using randomized datasets are notably absent. Moreover, the relationship between these subgroups and the glycaemic response to glucose lowering drugs has yet to be fully explored. Our study addressed this gap by elucidating the genetic heterogeneity within T2D using comprehensive real-world data from the UK Biobank along with findings from two pivotal semaglutide trials (STEP1 and STEP2) which include individuals with overweight and obesity with/without T2D. Additionally, we aimed to evaluate the potential links between the genetic subgroups and glycaemic response to glucose lowering drugs. We constructed T2D partitioned polygenic scores (pPS) based on variants reported by Smith et al (PMID: 37886436) for each participant in the UK Biobank and STEP trials. Subsequently, we examined the respective distributions of pPS amongst varied ancestries - European (454,285), East Asian (461), South Asian (8,600), and Africans (5,308). Finally, we assessed their association with glycaemic response to metformin and sulphonylureas in individuals diagnosed with T2D, using published summary level data from the MetGen consortium. Compared to Europeans, Asians and Africans have higher genetic burden of both beta cell dysfunction and insulin resistance (lipodystrophy, hyper-insulin secretion) pPS, $p<2.2\times10^{-16}$. Africans have the highest genetic burden of the beta-cell dysfunction and hyper-insulin secretion pPS, while East Asians have the highest genetic burden of lipodystrophy, $p<2.2\times10^{-16}$. These findings were consistent in the STEP 1 and STEP 2 trials, where East Asians have greater genetic burden of beta cell dysfunction and insulin resistance pPS compared to Europeans, $p<2.2\times10^{-16}$. Notably, different T2D risk-increasing alleles in these pPS were associated with varied glycaemic response. Notably, higher beta cell pPS was significantly associated with reduced glycaemic response to metformin (p=0.0001), whereas increased lipodystrophy was associated with greater response to metformin (p=0.04).

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5041W: Detecting gene and isoform expression changes in the IBDs using long-read sequencing

Authors:

M. Kennedy Ng, M. Bauer, G. Lau, S. Silverstein, G. Lian, C. Beasley, M. Schaner, J. Wang, S. Z. Sheikh, T. Furey; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Crohn's disease (CD) and ulcerative colitis (UC), two types of inflammatory bowel disease (IBD), cause chronic intestinal inflammation. While the diseases differ in their clinical presentation, side effects, and complications, short RNA sequencing (RNA-seq) studies have discovered significantly altered expression patterns in both CD and UC. This includes genes encoding antimicrobial peptides (*DEFA5*, *DEFA6*), mucins (*MUC12*), tight junction molecules (*CLDN8*), the REG protein family (*REG1A*, *REG3A*), and glycoproteins (*OLFM4*). These studies generally combine all isoforms together in a single quantification. The emergence of long-read RNA-seq allows us to better characterize unique transcript isoforms and accurately quantify them across samples, and even discover transcripts that are not part of a standard genome reference. We hypothesize that performing long-read sequencing on patients with and without IBD will lead to a better understanding of the expression changes that occur at both the gene and isoform level. We performed Oxford Nanopore sequencing on six colonic samples: two from uninflamed regions from CD patients, two from inflamed regions of UC patients, and two non-IBD controls. We aligned samples using Minmap2, then used bambu for multi-sample transcript discovery and quantification. Finally, we performed differential expression (DE) testing with DESeq2 and differential exon usage (DEU) testing with DEXSeq. We found 33 DE genes between IBD and non-IBD, including previously characterized genes as well as newly discovered genes. Of 1688 genes newly assembled by bambu, one was DE. We validated its assembly in an independent genome-guided de-novo assembly pipeline using bulk RNA-seq from 140 CD and non-IBD patients. Furthermore, we found 239 DEU events from 174 genes in IBD tissues compared to non-IBD controls, including the increased usage of transcripts with exon skipping (*USP12, NUP153*), alternative 3' acceptor sites (*PPP3CB-AS1*), alternative transcript start sites (*SMIM27*, *OSGIN2*), and alternative transcript end sites (*GNPTAB*, *ANKLE2*, *TRPM6*). We even discovered a newly assembled isoform of *PCCA*, a gene that encodes a subunit of propionyl-CoA carboxylase (PCC); PCC plays a role in the breakdown of some amino acids, lipids, and cholesterol. This isoform was highly truncated in comparison to its other isoforms and was only expressed in the IBD tissues.This is the first study to analyze long-read RNA-seq data in intestinal mucosal tissues. Through this work, we both confirmed findings in the previous literature, as well as found both novel genes and transcript isoforms that are differentially expressed in the IBDs.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5042W: Detecting genetic effects on variability in body fat distribution in UK Biobank to uncover gene-environment interactions

Authors:

X. Zhang[1], J. Min[2], O. Pain[3], J. Bell[1]; [1]King's Coll. London, London, United Kingdom, [2]Univ Bristol, Bristol, United Kingdom, [3]King s Coll. London, London, United Kingdom

Abstract:

The increased rate of obesity worldwide has led to a greater prevalence of insulin resistance, metabolic disease, and cardiovascular disease. Specific cardio-metabolic phenotypes, such as visceral fat accumulation, are stronger risk factors for disease than measures of overall obesity. Here, we explored the genetic architecture of magnetic resonance imaging (MRI) based measures of body fat distribution, including visceral fat.

Adiposity phenotypes are influenced by genetic and environmental factors, and gene-environment interactions (GxE). However, little is known about GxE interaction effects on body fat distribution. Variance quantitative trait loci (vQTLs) are genetic variants associated with phenotypic variance. Analysis of vQTLs is an efficient approach for the detection of GxE interactions, because interactions increase the phenotypic variance of particular genotype groups. The current study aims to identify vQTLs for MRI-based estimates of body fat distribution in the UK Biobank to uncover GxE interaction effects.

We analysed genotype data from up to 49,463 European individuals from UK Biobank with MRI-based fat distribution measures. We considered eight body fat phenotypes including four measures of abdominal adiposity, as well as visceral adipose tissue (VAT), abdominal subcutaneous adipose tissue (ASAT), liver proton density fat fraction (PDFF), and pancreas PDFF. We employed the Brown-Forsythe test (BF), the deviation regression model (DRM), and the squared residual value linear model (SVLM) to identify vQTLs of these eight MRI-based adiposity phenotypes.

Our analysis identified 23 unique vQTLs influencing at least one of the eight adiposity phenotypes, both with and without correction for BMI. Among the top-ranked vQTLs was rs429358 in *APOE*, which affects liver PDFF. This signal was consistently detected by all three vQTL detection methods, both with and without correction for BMI, suggesting its effect on liver PDFF variability is not affected by overall obesity.

Ongoing work is focused on GxE interaction analyses with the identified vQTLs, incorporating a comprehensive list of potential environmental factors and exposures collected by the UK Biobank. Future work will consider the predictive utility of vQTL-based polygenic risk scores for metabolic disease risk, as well as the longitudinal stability of the observed vQTLs and GxE interaction effects in a subset of samples.

In conclusion, this study identified vQTLs to discover GxE interaction effects for MRI-based regional fat measures. The results provide novel insights into the genetic and environmental basis of cardio-metabolic health and disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5043W: Development of a multi-ancestry type 1 diabetes genetic risk score and application to genomic changes in the pancreas

Authors:

R. Melton[1], C. McGrail[1], M. Miller[1], Q. Yang[1], H. Mummey[1], L. Tucciarone[1], Y. Lee[1], M. Shapiro[2], T. Brusko[2], I. Kusmartseva[2], M. Atkinson[2], A. Wang[1], K. Gaulton[1]; [1]Univ. of California San Diego, La Jolla, CA, [2]Univ. of Florida, Gainesville, FL

Abstract:

Genetic risk scores (GRS) for type 1 diabetes (T1D) have primarily been developed in individuals of European ancestry, making it difficult to integrate multi-ancestry data to determine changes associated with T1D genetic risk. To address this challenge, we created a multi-ancestry GRS using a composite set of 131 cross-ancestry T1D risk variants genome-wide and at the MHC locus. This 131-variant T1D GRS (named multi-T1GRS) was highly predictive of T1D in both Europeans (AUC= 0.894) and African Americans (AUC=0.870), and we generated a percentile scale for each ancestry to enable comparisons across multi-ancestry data. We used this T1D GRS to examine the effect of genetic risk on genomic changes in the pancreas using single-cell multi-omic (RNA-seq, ATAC-seq) profiles generated from 129 mixed ancestry individuals from the Network for Pancreatic Organ donors with Diabetes (nPOD) and Human Pancreas Analysis Program (HPAP) biorepositories. In total, we integrated 724,134 gene expression and chromatin profiles from all 129 donors, which were mapped to 19 pancreatic cell types and subtypes. From these, we derived cell type-specific gene regulatory profiles for each donor. We then calculated two multi-T1GRS percentiles for each donor, one using all T1D risk variants (total GRS) and another excluding variants at the MHC locus (polygenic GRS). We then determined whether genomic profiles of each pancreatic cell type were associated with either total or polygenic GRS. We identified changes in gene expression in endocrine alpha, beta, and delta cells, as well as immune populations like T cells associated with multi-T1GRS polygenic risk in non-diabetic individuals. In beta cells, higher polygenic GRS was associated with protein degradation and biogenesis of lysosome-related organelles, which is linked to antigen processing and presentation and increased stress. In other endocrine cell types, including alpha and delta cells, GRS was associated with genes related to TGF-beta signaling and NF-kB activity. In addition, in T cells, there was altered activity of anti-inflammatory cytokine production pathways with a higher risk score. Finally, we identified gene expression changes in multiple subtypes of exocrine acinar cells, associated with GRS, including iron uptake and transportation pathways and genes such as *FTH1* and *FTL*. Together, these results provide a multi-ancestry genetic risk score for

T1D, facilitating analysis of mixed ancestry cohorts and enabling insight into disease mechanisms in at-risk individuals.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5044W: Diagnostic potential of miR-548c-3p in CKD pathogenesis by integrated computational and experimental analyses

Authors:

M. Khan[1], M. Ejaz[1], A. Liaquat[2], H. Salim[1]; [1]COMSATS Univ. Islamabad, Islamabad, Pakistan, [2]Shifa Tameer-e-Millat Univ. Islamabad, Pakistan, Islamabad, Pakistan

Abstract:

Micro RNAs (miRNAs) are small non-coding sequences that play potent role in regulating gene expression. The dysregulation of gene expression can disrupt many important biological processes and pathways fostering onset of several diseases including kidney diseases e.g., chronic kidney disease (CKD). The main aim of this research was to identify and explore the potential role of miR-548c-3p as a biomarker in diagnosis of CKD during both early and later stages. The potential role of miR-548c-3p in CKD pathophysiology was explored through *in silico* analysis and expression profiling. Datasets of gene expression profiling by array were obtained from Gene Expression Omnibus (GEO), subjected for statistical analysis in GEO2R to identify differentially expressed genes (DEGs). miR-548c-3p indicated strong association with CKD regulating expression of *SIRT1* and *STAT3* in several biological processes, molecular functions, and cellular components. Furthermore, we found FOXO signaling pathway as the most enriched pathway through KEGG pathway analysis. However, the stage wise relative expression of miR-548c-3p and its target genes revealed aberrant expression in CKD pathogenesis by regulating expression of *SIRT1* and *STAT3*. The *in silico* and wet laboratory approaches employed in this research embodies inclusive role of ascertained miR-548c-3p as potential biomarkers for CKD diagnosis. Further research can be carried out to develop miR-548c-3p as biomarker to modulate the dysregulated pathways for diagnosis of CKD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5045W: Differential gene expression of opioid overdose death across human brain regions

Authors:

C. Willis[1], B. Quach[1], B. Maher[2], N. Gaddis[1], A. Palmer[3], E. Chesler[4], J. A. Bubier[4], S. Sanchez-Roige[5,3], L. Davis[5], V. Troiani[6], D. Jacobson[7], D. Hancock[1], S. Akbarian[8], O. Corradin[9], E. Johnson[1]; [1]RTI Intl., Durham, NC, [2]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [3]Univ. of California San Diego, La Jolla, CA, [4]The Jackson Lab., Bar Harbor, ME, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Geisinger Clinic, Danville, PA, [7]Oak Ridge Natl. Lab., Oak Ridge, TN, [8]Icahn Sch. of Med., New York, NY, [9]MIT, Cambridge, MA

Abstract:

Opioid addiction and overdose deaths (OOD) remain public health crises. However, our understanding of the genetics and neurobiological mechanisms associated with these conditions remain limited and constrain development of new treatments. The nucleus accumbens (NAc) and prefrontal cortex (PFC) are two key brain regions in the reward pathway that are associated with addiction. Here we meta-analyzed a new cohort of NAc RNAseq samples from post-mortem human brains (N=92) with an existing cohort (Seney et al., N=40) using a random effects model to account for inter-dataset variability. Cases were defined by OOD among opioid misusers and controls had negative toxicology for opioids and no history opioid misuse at time of death. Of the 16,183 genes analyzed in both cohorts, 192 differentially expressed genes (DEGs) were identified in our meta-analysis (FDR < 0.10). We then compared the NAc meta-analysis DEGs to a meta-analysis of DEGs associated with OOD in the PFC brain region (Carter et al.) across 4 datasets (Corradin et al, Seney et al., Sosnowski et al., Mendez et al., N=285, 335 DEGs). When testing whether the 191 available NAc DEGs extended in the PFC meta-analysis, 16 genes showed statistically significant replication ($p < 0.05/191$), indicating predominantly region-specific but some cross-region differential gene expression by OOD status. The 16 cross-brain region DEGs were *ARL4D, CITED2, DUSP2, DUSP4, DUSP6, DYRK2, EGR4, FSTL1, GPR19, KAT6A, LBH, NPAS4, NR4A2, RGS2, SRSF6, TMEM198.* Overrepresentation analyses of cross-region OOD DEGs with Reactome database gene set terms identified several signal transduction pathways (e.g., RAF-independent MAPK1/3 activation and Signaling by NTRK1 (TRKA)) and NPAS4 regulation pathways (e.g., Regulation of NPAS4 gene expression) enriched for DEGs at FDR < 0.05. Although we didn't identify any drug compounds that target the 16 cross-region genes, we identified 288 FDA approved drug compounds that target the 445 significant DEGs (72 DEGs are targeted by approved drug compounds) from the NAc and PFC meta-analyses and assessed their drug repurposing potential. This list of drugs contains drugs like methadone, a drug already used to treat opioid use disorder, as well as many drugs that treat other diseases and disorders. By leveraging a new NAc dataset and making use of existing data, we identified DEGs robustly associated OOD, examined their potential role in the neurobiology of OOD and identify drugs that target DEGs for potential drug repurposing opportunities for opioid use disorder treatment.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5046W: Disentangling allele-specific effects of the HLA class II (HLAII) haplotype DR15/DQ6 in the development of inhibitory anti-drug-antibodies to therapeutic factor VIII proteins (tFVIIIs) using a peptidomics approach

Authors:

T. Howard[1,2,3,4], B. W. Luu[3], M. A. Almeida[1,2], M. Hofmann[5], E. G. Manusov[1,2], J. A. Galan[1,2], J. S. Powell[3,6], L. V. Dinh[3], R. Rajalingam[7], H. Mead[8], J. M. Peralta[1,2], P. V. Lehmann[9,10], S. Kumar[1,2], E. J. Fine[11], J. E. Curran[1], M. A. Escobar[12,13], S. Williams-Blangero[1,2], E. Maraskovsky[14], J. Blangero[1,2], V. P. Diego[1,2]; [1]Dept. of Human Genetics, Sch. of Med., Univ. of Texas Rio Grande Valley, Brownsville, TX, [2]South Texas Diabetes and Obesity Inst., Sch. of Med., Univ. of Texas Rio Grande Valley, Brownsville, TX, [3]Haplogenics Corp., Brownsville, TX, [4]Dept. of Pathology and Lab Med., VA Valley Coastal Bend Hlth.care Ctr., Harlingen, TX, [5]CSL Innovation GmbH, Marburg, Germany, [6]Div. of Hematology and Oncology, Dept. of Internal Med., Sch. of Med., Univ. of California at Davis, Davis, CA, [7]Immunogenetics and Transplantation Lab., Dept. of Surgery, Sch. of Med., Univ. of California at San Francisco, San Francisco, CA, [8]Global Med. Affairs, BioMarin, Novato, CA, [9]Dept.s of Pathology and Neurology, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [10]Cellular Technology Ltd, Shaker Heights, OH, [11]Consultant, Atlanta, GA, [12]Div. of Hematology and Oncology, Dept. of Med., Sch. of Med., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [13]Gulf States Hemophilia and Thrombophilia Ctr., Houston, TX, [14]Bio21 Inst., CSL Limited Res., Melbourne, Australia

Abstract:

HLA-DR15/DQ6 is a haplotype associated with immune mediated diseases and adverse drug events, most notably including multiple sclerosis and the anti-drug antibodies that develop against the therapeutic FVIII proteins (tFVIIIs) infused to prevent bleeding in hemophilia patients. This HLA class II (HLAII) haplotype expresses three polymorphic β-chain genes, *DRB5*, *DRB1*, and *DQB1*, among which DRB5*01:01, DRB1*15:01, and DQB1*06:02 are the most prevalent alleles. Though results from recent studies show that the constituent DR15/DQ6 alleles may differentially affect the immunogenicity of tFVIIIs, the question as to which is the true risk allele is difficult to answer by genetic association analysis due to the strong linkage disequilibrium (LD) across the MHC. We address this by analyzing DR- and DQ-bound peptides derived from tFVIIIs in two independently performed sets of dendritic cell (DC)-protein processing/presentation assays. We performed log linear mixed model analyses under 2

models, where the dependent variable is the logarithm of the expected peptide count. Under Model 1, we analyzed a single HLAII allele predictor variable consisting of 10 levels (4 distinct DRB1 alleles and 6 DQB1 alleles) in the fixed effect component (FEC), and variables for individuals (8 levels from one set and 9 from the other), experiments (2 levels from both sets), and HLAII isomers (2 levels for DR and DQ) in the random effects (REs) component, which maximizes the signal to noise ratio by reducing the error (noise) in our estimation of the predictive effect of HLA alleles for peptide counts by accounting for the clustering due to individuals, experiments, and isomers. Model 2—which was more focused, as the single HLA allele predictor variable in the FEC contained only DRB1*15:01 and DQB1*06:02, with the same REs as in model 1. Relative to the baseline (DQB1*02:01) allele of Model 1, both DRB1*15:01 and DQB1*06:02 significantly increased risk of contributing to the overall peptide count, and DRB1*15:01 contributed significantly more than DQB1*06:02. Specifically—reported as risk ratios (RRs) and 95% confidence interval (CI) lower bound (LB) and upper bound (UB)—we found a RR (95% CI LB, 95% CI UB) of 13.9 (10.2, 18.9) and 2.2 (1.6, 3.1) for DRB1*15:01 and DQB1*06:02. Under Model 2, the head-to-head comparison, the RR for DRB1*15:01 vs DQB1*06:02 (baseline) is 7.1 (5.8, 8.5). Our results show that: 1) a log linear mixed model analysis of MHC-associated peptidomics data can help answer questions concerning HLA alleles that are intractable to association analysis due to the strong and complex LD of the extended MHC; and 2) DRB1*15:01 and/or DRB5*01:01 are likely the true HLAII risk allele(s).

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5047W: Dissecting type 1 diabetes etiology with human genetics and population-scale multiome profiling of human islets under basal and pro-inflammatory conditions

Authors:

C. Robertson[1], X. Wang[1], P. Orchard[1], A. Chi Nok Chang[2], Z. Meng[2], V. Dolgachev[1], N. Narisu[3], M. Erdos[3], F. Collins[3], S. Chen[2], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Weill Cornell, New York, NY, [3]NIH, Bethesda, MD

Abstract:

Type 1 diabetes (T1D) is defined by loss of insulin production due to autoimmune attack on beta cells in pancreatic islets. Inflammation in the islet may be an early trigger of beta cell stress leading to islet autoimmunity. However, why proinflammatory exposures lead to autoimmune attack on beta cells, but not other islet cell types, remains an open question. Genetic susceptibility to T1D likely shapes multiple components of T1D pathophysiology, including beta

cell response to inflammation and immune cell autoreactivity. To investigate genetic regulation of beta cell response to inflammation, we performed multi-omic profiling on 16 human islet samples, including control and cytokine-treated (IL-1beta and TNF-alpha) for each of 8 donors, using 10X Genomics single nucleus multi-omic sequencing, which jointly profiles gene expression and chromatin accessibility on the same nuclei. We are currently generating multi-omic profiling on an additional 194 samples (control and treated samples for each of 97 donors). We are pooling samples for single nucleus experiments using a randomized block design, which reduces sample-to-sample variability and eliminates batch confounding in downstream analyses. Specifically, we control for cytokine treatment, islet distribution center, and donor age, sex, and BMI in our randomization scheme. Sequencing analysis of 37,170 high quality nuclei from 16 samples profiled to date identified transcriptional response to cytokines in major islet cell types (547, 887, and 2,998 significantly differentially expression genes in ductal, alpha, and beta cells, respectively, FDR < 0.05). Marker gene expression was significantly downregulated in beta cells (*INS* fold change (FC) = 0.55, $p = 3 \times 10^{-5}$) but unchanged in alpha (*GCG* FC = 0.68, p = 0.07) and ductal cells (*KRT19* FC = 0.82, p = 0.60). Similarly, pathway analyses using the human MSigDb hallmark gene sets indicated upregulation of inflammatory response pathways in all three cell types and downregulation of the “hallmark beta cell” pathway in beta cells only (p = 0.005). In conclusion, we found beta cell-specific responses to proinflammatory cytokines, which may contribute to beta cell demise in T1D. Future analyses with larger sample sizes will explore the genetic basis of beta cell response to inflammatory triggers through population-scale analysis of genetic effects on molecular response to cytokine treatment (response-eQTL and -caQTL analyses).

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5048W: DNA damage-associated protein co-expression modules provide insight into genetic variation tolerance and cardiovascular disease risk

Authors:

O. Johnson[1], S. Paul[2], J. A. Gutierrez[3], W. K. Russell[1], M. C. Ward[4]; [1]the Univ. of Texas Med. Branch, Galveston, TX, [2]Univ. of Texas Med. Branch Galveston, Galveston, TX, [3]the Univ. of Texas Med. Branch at Galveston, Galveston, TX, [4]Univ. of Texas Med. Branch, Galveston, TX

Abstract:

Cardiovascular disease (CVD) is the leading cause of death worldwide, and is influenced by complex interactions between an individual’s genetics and environmental risk factors that

modify the molecular environment of CVD cell types. Although Genome-Wide Association Studies (GWAS) identify risk loci associated with complex CVD traits that can be mapped to genes, the molecular mechanisms by which those risk genes are linked to disease processes is unclear. This is in part due to the absence of the appropriate cellular context. One unifying consequence of the molecular risk factors in CVD is DNA damage in the heart, which must be repaired by DNA damage response proteins. However, the impact of DNA damage on the heart proteome, and its relationship to CVD risk remain unknown. We therefore treated induced pluripotent stem cell-derived cardiomyocytes (iPSC-CM) from three healthy individuals with a sub-lethal dose of the DNA damaging agent Doxorubicin (DOX) and a vehicle control for 24 hours. Induction of DNA damage was confirmed by the presence of γ-H2AX in DOX-treated cells. We identified 4,178 proteins by tandem liquid-chromatography mass spectrometry and found the iPSC-CM proteome to be highly correlated with the proteome from human left ventricle heart tissue. Weighted Gene Co-expression Network Analysis resulted in 12 co-expressed modules, with five significantly correlated with DOX treatment. These DNA damage-associated modules represent distinct biological processes including RNA processing, chromatin regulation, translational regulation and metabolism. To connect network features with measurements of genetic variation, we identified 403 hub proteins (202 DOX-correlated) with high intramodular connectivity. DOX-correlated hub proteins are depleted for cis and trans pQTLs but are enriched for genes that are intolerant to mutations (pLI > 0.9), highlighting the importance of these proteins. While CVD traits such as arrhythmia and dilated cardiomyopathy are enriched in specific DOX-correlated modules, DOX-correlated hub proteins are not enriched for CVD GWAS risk proteins. Instead, DOX-correlated hub proteins are enriched for proteins that physically interact with CVD GWAS risk proteins. Further, CVD GWAS risk proteins are tolerant to mutation, while their physical interactors are mutation intolerant. Our data demonstrate that DNA damage in cardiomyocytes induces diverse effects on biological processes through protein co-expression modules that are relevant for CVD, and that the level of protein connectivity in DNA damage-associated modules influences the tolerance to genetic variation.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5049W: Doubling the number of genetic risk loci for mosaic loss of chromosome Y in 544,112 diverse ancestry participants from the Million Veteran Program ⭐

Authors:

M. Francis[1], B. R. Gorman[1], T. B. Bigdeli[2,3], G. Genovese[4], G. Voloudakis[5,6], J. Bendl[5], B. Zeng[5], S. Venkatesh[5,6], C. Chatzinakos[2], E. McAuley[1], S-G. Ji[7], K. Markianos[1], P. Schreiner[1], E. Partan[1], Y.

Shi[1], P. Devineni[1], VA Million Veteran Program, A. Bick[8], P. Natarajan[4], T. Assimes[9], P. Tsao[10], D. Klarin[11], C. Tcheandjieu[12], N. Peachey[13,14], S. Iyengar[13,15], P. Roussos[5,6], S. Pyarajan[1,16]; [1]VA Boston Hlth.care System, Boston, MA, [2]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [3]VA New York Harbor Hlth.care System, Brooklyn, NY, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]New York/New Jersey VA Hlth.Care Network, Bronx, NY, [7]BridgeBio Pharma Inc., Arlington, MA, [8]Vanderbilt Univ. Med. Ctr., Nashville, TN, [9]Stanford Univ Sch Med., Palo Alto, CA, [10]VAPAHCS/Stanford Univ., Palo Alto, CA, [11]Massachusetts Gen. Hosp., Boston, MA, [12]Gladstone Inst., San Francisco, CA, [13]VA Northeast Ohio Hlth.care System, Cleveland, OH, [14]Cleveland Clinic, Cleveland, OH, [15]Case Western Reserve Univ., Cleveland, OH, [16]Harvard Med. Sch., Boston, MA

Abstract:

Background. Mosaic loss of the Y chromosome (mLOY) is the most common type of mosaic chromosomal alteration (mCA), observable in upwards of 40% of males above age 70, and 70% of males over 85. Inherited factors govern many processes which can lead to somatic mCA acquisition and subsequent clonal expansion in hematopoietic cell lineages, including cell-cycle regulation, DNA damage response, and susceptibility to cancer. Estimates of single nucleotide polymorphism-based heritability (SNP-h^2) for mLOY are as high as 31.7%. **Methods.** We studied 544,112 male participants in the Million Veteran Program (MVP), a biobank in the Department of Veterans Affairs (VA) healthcare system, which combines ancestrally diverse genetic data with extensive electronic health records. Using the MoChA software tool, we identified 126,108 MVP participants with detectable mLOY (cases: 106,054 European; 13,927 admixed African; 6,127 Hispanic) and 418,004 without (controls: 294,916 European; 85,176 admixed African; 37,912 Hispanic). Genome wide association studies (GWAS) and multi-ancestry meta-analysis were performed. **Results.** We identified 323 genome-wide significant mLOY risk loci, 167 of which were novel, including five AFR-specific loci. Tract-based ancestry deconvolution resolved local inflation at AFR lead SNPs. Transcriptome-wide association studies (TWAS) of mLOY summary statistics highlighted 2,297 significant genes; these genes were significantly enriched for sister chromatid cohesion, cell cycle, and DNA replication pathways. Mendelian randomization (MR) implicated six traits as causal influences on mLOY: triglycerides, high-density lipoprotein, smoking, body mass index, testosterone, and sex hormone-binding globulin; and indicated a causal influence of mLOY on plateletcrit, prostate cancer, lymphocyte percentage, and neutrophil percentage. Multivariable MR confirmed that the inferred causality identified in univariable MR was independent of predisposition to cigarette smoking. **Conclusion.** These findings significantly advance our understanding of the inherited risk and broader health implications of mLOY. Additionally, our results highlight the benefits of including more diverse populations in large-scale GWAS of complex polygenic traits.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5050W: Educational Attainment as a Cognitive Reserve Factor in Alzheimer Disease: Implications for *APOE e4* Carriers

Authors:

F. Rajabli[1], A. Emekci[2], A. Sequencing Project[1]; [1]Univ. of Miami, Miami, FL, [2]Pioneer High Sch., San Jose, CA

Abstract:

Background: Cognitive reserve posits that certain life experiences, such as education, build resilience in the brain, enabling individuals to better cope with brain changes related to neurodegenerative diseases. Education as a proxy for cognitive reserve in Alzheimer disease (AD) has been a key area of research, aiming to understand how educational attainment (EA) may serve as a buffer. However, the relationship between EA, the presence of an AD genetic risk factor, and cognitive manifestation remains understudied. This study assessed the effect of EA on cognitive outcomes among individuals with amyloid plaques post-autopsy and investigated the protective effect of EA in the presence of the APOE ε4 allele, the strongest genetic risk factor for late-onset AD. **Methods:** The study uses publicly available data from the National Alzheimer's Coordinating Center, with a sample size of 1,513 individuals. We assessed whether EA provides protection despite the presence of neuritic plaques and if this effect varies between ε4 carriers and non-carriers. Utilizing EA as a cognitive reserve proxy, CDRSUM as a measure of cognitive impairment, and CERAD score (C) as indicators of AD-related brain changes. We performed the non-parametric Mann-Whitney U test to evaluate the relationship between EA (less than or equal to 8 years (EA1), 9 to 12 years (EA2), and over 12 years (EA3)) and CDRSUM in the presence of ε4 among individuals with neuritic plaques. **Results:** We observed a significant difference in CDRSUM scores among individuals with frequent neuritic plaques (C3) between the EA1 and EA2 groups ($p = 0.023$) and between the EA1 and EA3 groups ($p = 0.030$). Additionally, there was a significant difference in CDRSUM scores between the EA2 and EA3 groups in individuals with sparse neuritic plaques (C=1) ($p = 0.006$). When stratifying EA groups by *ε4* allele carriers and non-carriers, we noted significant differences in CDRSUM scores within the EA3 group across the C1, C2, and C3 subgroups. Individuals in the EA3 group with C1, C2, and C3 plaques had significantly higher CDRSUM scores if they were carriers of the ε4 allele ($p = 0.008, 0.04, 0.002$). This indicates that higher EA serves as a limited cognitive reserve factor in ε4 carriers. **Conclusion:** This study highlights the significance of EA as a modifiable risk factor, with higher education playing a protective role in the symptoms of AD. However, EA does not offer as much reserve in individuals with the genetic risk factor *APOE* ε4 allele. The findings

advocate for the inclusion of educational background in clinical assessments as a critical variable, emphasizing its role in cognitive reserve in moderating the impact of dementia-related pathologies.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5051W: Elucidating Sex-Differential Genetic Associations on Plasma Metabolites and Coronary Artery Disease Risk: Insights from Sex-stratified GWAS Analyses in the UK Biobank. ⭐

Authors:

X. Wang[1,2], J. Hu[1,3,4], Y. Li[1], J. Li[2,5], L. Liang[1,6]; [1]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Div. of Preventive Med., Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [3]Ctr. for Genomic Med. and Dept. of Anesthesia, Critical Care and Pain Med., Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [4]Med. and Population Genetics, The Broad Inst. of MIT and Harvard, Boston, MA, [5]Dept. of Nutrition, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [6]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Sex differences in coronary artery diseases (CAD) are widely recognized and are attributable to genetic, environmental, and societal factors. Identifying sex-differential metabolites associated with CAD, could provide key evidence for discovering novel disease mechanisms and facilitating precision risk prediction and intervention. Here, we employed REGENIE to identify sex-stratified metabolite quantitative trait loci (mQTLs), for 251 nuclear magnetic resonance (NMR)-measured metabolomic markers (145 lipids, 11 amino acids, and 95 markers of other categories), in 106,275 men and 124,272 women with European ancestry from the UK Biobank (UKB). We identified 12,236 mQTLs in women and 18,199 mQTLs in men. Heritability estimates for individual biomarkers ranged from 0.03 to 0.20 among men, and from 0.02 to 0.23 among women. The between-sex genetic correlations for metabolomic biomarkers ranged from 0.60 to 1.00, indicating varying degrees of shared genetic architecture between sexes. We then conducted a between-sex genome-wide heterogeneity test for each biomarker. The analysis revealed 1,095 genetic loci exhibiting statistically significant sex differences in genetic effects ($p < 5e-08$) across 231 metabolites (numbers of sex-differential loci ranged from 1 to 11 for each metabolomic maker), suggesting broad sex differences in the genetic regulations of the metabolism. We focused on 3 lipid biomarkers with the least shared genetics between sexes (i.e., between-sex genetic correlations < 0.7), namely, triglycerides-to-total lipids ratio in

chylomicrons and extremely large VLDL, phospholipids/total lipids ratio in very large VLDL, and cholesterol/total lipids ratio in small LDL. The 3 biomarkers had 10 sex-differential loci across the genome. To examine the clinical relevance of these sex differential effects, we conducted sex-stratified GWAS for CAD in 187,879 women (9,877 cases) and 220,298 men (21,491 cases) in the UKB. We performed a sex-specific colocalization analysis for these 3 biomarkers and CAD and identified an mQTL (rs7412; mapped to *APOE*) for 2 VLDL markers colocalized with CAD in both sexes (PPH4 > 0.90). Specifically, rs7412 showed greater effects on levels of the 2 VLDL markers (beta = -0.05, p = 2.33e-11 for male, and beta = 0.01, p = 0.07 for female) and CAD risk in males (beta = 0.15, p = 1.8e-16) compared to females (beta = 0.10, p = 1.2e-04). In summary, our study identified broad sex differences in genetic regulations on the metabolism. Additionally, sex-differential regulations on extreme and very large VLDL, led by rs7412 on *APOE*, may contribute to sex differences in CAD risk.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5052W: *EMILIN2* dissociates obesity from abdominal fat distribution by enhancing adipocyte differentiation and function

Authors:

J. Belanich[1,2], M. Urena[1], R. S. Martinez[1], A. Hecq[3], A. Mosnier[4], B. Emanuelli[1], R. Gibson[2], T. Batista[2], R. Cho[2], F. dos Santos[2], H. Dashti[2], M. Claussnitzer[5,2], T. Kilpeläinen[6,2]; [1]Univ. of Copenhagen, København, Denmark, [2]Broad Inst., Cambridge, MA, [3]Univ. of Mons, Mons, Belgium, [4]Oniris École Natl.e vétérinaire, Nantes, France, [5]Massachusetts Inst Technology, Cambridge, MA, [6]Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

Obesity is linked to compromised cardiometabolic health, but the strength of this relationship varies greatly among individuals. The distribution of adipose tissue, particularly the ratio of abdominal to gluteofemoral fat, plays a key element. At population level, a higher body mass index (BMI) is correlated with a higher waist-hip ratio (WHR), a marker of abdominal fat distribution. Intriguingly, some genetic loci inverse the relationship between BMI and WHR, improving cardiometabolic health. The biological mechanisms underlying this genetic dissociation between BMI and WHR are not well understood. We have focused on *EMILIN2*, located at a locus inverting BMI and WHR. The lead-variant of this locus is also associated with increased gluteofemoral adipose tissue volume, improved insulin sensitivity and liver markers, and decreased triglyceride levels, blood pressure, and risk of type 2 diabetes. This variant, located in the promoter of *EMILIN2*, colocalizes with a subcutaneous adipose tissue eQTL for

the gene. Gene knockdown with siRNA in human A41 subcutaneous adipocyte cultures decreased differentiation and function of the cells, affecting cellular metabolism and adipogenic markers. Further investigations using CRISPRi perturbations and the LipocyteProfiler for hundreds of image-based cellular features revealed a reduction in lipid storage capacity, with no effect on cell proliferation, size, or apoptosis. Currently, we are performing RNA-Seq and CRISPRa to draw further connections between the modified function of *EMILIN2* and cellular phenotypes. Our study sheds light on how *EMILIN2* may uncouple obesity from abdominal fat distribution and adverse cardiometabolic outcomes by influencing adipocyte function, improving our understanding of the complex interplay between adipose tissue and cardiometabolic health.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5053W: Engineering 3D corneal organoids from patient-derived iPSCs to model and investigate underlying genetic influences on keratoconus pathophysiology

Authors:

A. Chen[1], L. Szczotka-Flynn[1], K. Brustoski[1], S. Iyengar[1], P. Hysi[2], A. Schaffer[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]KCL St. Thomas Hosp., London, United Kingdom

Abstract:

Keratoconus is a rare ocular disease characterized by progressive changes in corneal structure and morphology resulting in distortions of visual function and increased sensitivity to light and glare. As it stands currently, keratoconus remains a clinical diagnosis with much that remains nebulous regarding its pathophysiology. Previous studies have identified genetic risk alleles and risk factors for keratoconus, amongst which family history is highest followed by eye rubbing, eczema, asthma, and allergy, respectively. In this study, we hypothesize that genetics influence keratoconus disease severity ranging from keratoconus refractory to corneal transplantation to mild and moderate keratoconus treated via corneal cross-linking. To investigate this hypothesis, patients with progressive recurrent keratoconus (n=10), mild to moderate keratoconus (n=10), and age-matched controls (n=10) were recruited and consented for blood and saliva collection to be used for genetic analysis and reprogramming into iPSCs. Three-dimensional corneal organoids will be developed using unique patient-derived iPSCs *in vitro* to model keratoconus pathogenesis and pathophysiology and validated using known cell markers of corneal maturity. Single-cell RNA-sequencing of keratoconus organoids and discarded patient cornea are expected to reveal genes with differential expression and splicing profiles that correlate with keratoconus

severity and identify shared molecular pathways associated with disease. Collectively, this work represents the novel application of engineered three-dimensional corneal organoids to model ocular diseases and untangle the genetic influences on keratoconus pathophysiology and phenotypic severity.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5054W: Enrichment of rare copy number variations in metabotropic glutamate receptor gene networks in ADHD patients with comorbid neurodevelopmental phenotypes

Authors:

J. Glessner[1], M. Khan[2], X. Chang[3], Y. Liu[4], S. Torkamandi[5], A. Deborah[6], F. G. Otieno[6], M. Lemma[6], F. Mentch[6], J. Li[7], C. Kao[5], M. March[8], H. Qu[6], J. Connolly[5], H. Hakonarson[9]; [1]Childrens Hosp. of Philadelphia/PENN, Philadelphia, PA, [2]Children s Hosp. of Philadelphia, Wallingford, PA, [3]CHOP, Bryn Mawr, PA, [4]The children s Hosp. of philadelphia CHOP, Philadelphia, PA, [5]Children s Hosp. of Philadelphia, Philadelphia, PA, [6]Childrens Hosp. of Philadelphia, Philadelphia, PA, [7]the Children s Hosp. of Philadelphia, Philadelphia, PA, [8]Childrens Hosp. of Philadelphia, Lansdowne, PA, [9]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Copy number variations (CNVs) in metabotropic glutamate receptor genes (mGluRs) have been implicated in various neurodevelopmental disorders (NDD) and psychiatric disorders. We conducted an analysis integrating multiple psychiatric query approaches and utilizing multiple sources of phenotype classification, including eMERGE algorithms using International Classification of Diseases (ICD) codes, PheCodes, and updated Python/SQL queries, in children and adolescents aged 6 to 18 years. A total of 72,626 samples biobanked and genotyped by the Center for Applied Genomics at CHOP on arrays with genome-wide coverage and meeting CNV quality control standards were analyzed, ensuring no duplicates were included. We leveraged our high density Omni2.5 array data with CNVs of high curation quality as a reference for calling other array data versions for CNVs by utilizing PennCNV script detect_cnv options validate and candlist, we identified CNVs and filtered for CNV confidence score exceeding 10. Previously we reported enrichment in four out of eight mGluR genes (mGluR1, mGluR5, mGluR7, mGluR8) exhibiting significant CNV associations individually. Moreover, while CNVs in individual mGluR interacting genes showed marginal significance (such as CNTN4, MC4R, CTNNA2, SNCA, MAPT, ADRA2A, and CA8), they collectively revealed robust significance (p=1.03E-20). Our findings underscore the key role of mGluRs in attention-deficit/hyperactivity disorder (ADHD) and

comorbid NDDs. We updated the mGluR interacting network gene list using the latest string-db protein-protein interaction (PPI), resulting in retained significant enrichment of CNVs in NDDs, especially samples with multiple comorbid NDDs, which made up 2/3 of cases. In addition to ADHD and autism, multiple other comorbid conditions were observed including (sorted by frequency) developmental delay, anxiety, speech and language disorder, conduct disorder, learning disorder, epilepsy, tics and stuttering, depression, intellectual disability, bipolar, and psychosis. This study elucidates the genomic complexity underlying psychiatric and NDDs, emphasizing the important role of mGluRs and their interacting genes as therapeutic targets. Furthermore, we utilize advancements in variant detection, quality scoring, and the relevance of comprehensive comorbidity assessments, particularly between autism spectrum disorder and other mental health conditions. We conclude that genomic microarray and sequencing serve as a guiding tool toward personalized medicine, highlighting the intricate interplay between genetic variations such as mGluRs and clinical phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5055W: Establishment and Characterization of Mouse Models for the Study of Human Major Histocompatibility Complex Class I Gene Effects

Authors:

D. Shi[1], P-L. Chen[1,2,3], C-M. Hsieh[4], Y-Y. Pan[1], K-M. Kam[1], W-T. Lu[1]; [1]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ.,, Taipei, Taiwan, [2]Genome and Systems Biology Degree Program, Academia Sinica and Natl. Taiwan Taiwan, Taipei, Taiwan, [3]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, Taipei, Taiwan, [4]Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, Taipei, Taiwan

Abstract:

Human leukocyte antigen (HLA) genes are the most polymorphic genes in the human genome, and they are associated with numerous immune-related diseases or traits. Graves' disease (GD) is an autoimmune disease that is the leading cause of hyperthyroidism. Previous studies have shown that Graves' disease patients with antithyroid drug-induced agranulocytosis (TiA) are highly associated with two alleles: *HLA-B*38:02* (odds ratio: 21.48) and *HLA-DRB1*08:03* (odds ratio: 6.13). In addition, carrying both *HLA-B*38:02* and *HLA-DRB1*08:03* increases odds ratio to 48.41. We aim to develop multi-gene humanized composite mice to study GD, TiA, and other autoimmune diseases. We have successfully developed human *CD8*, human *B2M* double knock-in (KI) mice. We are currently incorporating humanized HLA-B alleles into the double-KI mice to

make a triple-KI mouse model. All of these three genes are critical for major histocompatibility complex (MHC) class I functionality. Using CRISPR/Cas9 technology, we constructed humanized mice and confirmed the successful knock-in via Southern blotting. Physiological assessments included regular body weight measurements and organ histology examinations. Immunophenotyping via flow cytometry demonstrated that double KI mice (n=8) had normal myeloid and lymphoid immune cell populations, similar to wild-type (WT) mice (n=8). Specifically, the double KI mice showed comparable levels of splenic and peripheral CD4+, CD8+ T cells, and B cells, with unchanged numbers of naive and memory T cell subsets. T cell development in double KI mice revealed similar numbers and distributions of double-positive (DP) and single-positive CD4+ and CD8+ thymocyte subsets to WT mice, with a slight reduction in the double-negative (DN) population. Innate immune cells such as natural killer cells, NKT cells, monocytes, neutrophils, macrophages, and dendritic cells were found in normal proportions in the spleen and peripheral blood of double KI mice. To analyze TCR gene usage, we constructed cDNA libraries from splenic T cells of double KI mice and performed TCR repertoire sequencing using next-generation sequencing technology. The double KI mice utilized diverse *TRAV*, *TRAJ*, *TRBV*, and *TRBJ* genes. In conclusion, this humanized *CD8* and *B2M* double KI mouse model exhibits normal immune and physiological functions, making it a powerful preclinical platform for investigating autoimmune disorders in the future if disease-susceptibility HLA alleles can be further integrated.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5056W: Estimated selection coefficients of protein coding variants support the large fitness impact of pathogenic variants in neurodevelopmental disorders

Authors:

X. Lin[1], J. Wright[2,3], P. Feliciano[2,3], Y. Shen[1,4]; [1]Dept. of Systems Biology, Columbia Univ. Irving Med. Ctr., New York, NY, [2]Div. of Genetics and Genomics, Boston Children's Hosp., Boston, MA, [3]Dept. of Pediatrics, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [4]Dept. of BioMed. Informatics, Columbia Univ. Irving Med. Ctr., New York, NY

Abstract:

Early onset severe conditions, such as neurodevelopmental disorders and congenital anomalies, lead to substantial reduction on reproductive fitness. As a result, pathogenic variants that contribute to these conditions with large effect are under strong negative selection in human populations. Recently a few methods have been developed for estimating selection coefficients

of protein truncating variants (PTV) and missense variants using biobank-scale human population genome data, including MisFit (for missense) and GeneBayes (for PTVs). In this study, we investigated how informative these estimated selection coefficients (S) are in the genetic diagnosis of neurodevelopmental disorders, using a cohort of 654 autistic individuals from SPARK with a clinically-confirmed likely pathogenic or pathogenic (P/LP) genetic variant. Nearly all the P/LP PTVs and more than half of P/LP missense variants have an estimated S greater than 0.05, which is larger than the average selection coefficient of a variant under strongest selection in individuals in the general population. The overall distribution of the estimated S for these P/LP variants are significantly ($p<1e-20$) larger than random de novo PTV and missense variants in a published autism cohort. We then ranked S in all coding variants in each individual with a genetic diagnosis. We found that 88% of P/LP PTVs are ranked at the top in corresponding individuals and 99% are within the top three. 41% of P/LP missense variants are ranked at the top and 63% are within the top three. The P/LP missense variants whose S is not within the top ranked have other evidence (predicted pathogenicity and gene-condition association) that establishes their pathogenicity. Our analysis suggests that estimated selection coefficients of rare coding variants may be informative for identifying and prioritizing likely pathogenic/pathogenic genetic diagnoses in neurodevelopmental disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5057W: Evaluation of sex-stratified polygenic risk scores for type 2 diabetes and glycemic traits in the Framingham Heart Study

Authors:

N. Wang[1], Y. Zhang[1], P. Schroeder[2,3], C-T. Liu[1], J. Dupuis[4,1], J. M. Mercader[2,3,5]; [1]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]McGill Univ., Montreal, QC, Canada, [5]Harvard Med. Sch., Boston, MA

Abstract:

Diabetes is a multifactorial disease with a significant genetic predisposition. Polygenic risk scores (PRS) have been developed to estimate an individual's genetic risk of a disease. Most PRS are based on sex-combined genome-wide association studies (GWAS) results, assuming same effect sizes in males and females. This study investigates the impact of incorporating sex-specific effects in PRS development for type 2 diabetes (T2D) and glycemic traits by comparing PRS performance derived from sex-combined versus sex-stratified GWAS. We conducted sex heterogeneity tests on sex-specific GWAS and identified nine T2D loci with sex-specific effects. We developed PRS[sex-combined] and PRS[sex-stratified] using either sex-combined or sex-

specific summary statistics from GWAS meta-analyses for T2D (41,444 T2D cases and 354,539 controls), fasting glucose (FG, N=120,595) and fasting insulin (FI, N= 98,210). We assessed these PRS in the Framingham Heart Study (1,272 T2D cases and 6,928 controls). PRS for prevalent and incident T2D were evaluated in all participants, and PRS for FG and FI were evaluated in non-diabetic individuals. PRS[sex-combined] shows better prediction than PRS[sex-stratified] across all traits examined. Our study suggests that PRS[sex-stratified] did not improve the prediction of T2D and glycemic traits compared to PRS[sex-combined]. This could largely be due to the reduced sample sizes involving calculating PRS[sex-stratified], highlighting the importance of larger sample sizes for the development of more powerful sex-stratified PRS models.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5058W: Evidence for widespread gene-environment interactions in genome-wide data

Authors:

B. Verhulst; Texas A&M Univ., Bryan, TX

Abstract:

Gene-environment interaction (GxE) plays a central role in the genetic etiology of many phenotypes. Twin and model organism studies have found pervasive GxE for a wide variety of traits. These findings suggest that GxE depends on the specific combination of the phenotype, the moderator, and the locus, meaning that associations may only be identifiable in some environments. Furthermore, phenotypes are more heritable (i.e., genetic factors play a larger role) in adverse environments. These findings imply that while current genome-wide association studies (GWAS) may capture the average influence of an allele on a phenotype across all levels of a potential moderating variable, at some levels of the moderator the genetic association may be substantially larger (or smaller) than average. Accordingly, while there may not be a main effect of the allele on a phenotype, it is very possible that, at some levels of an environment, there is a robust genetic association. To explore the possibility of wide-spread gene-environment interactions, we conducted a series of genome-wide gene-environment interaction studies (GxE GWAS) for a set of phenotypes and moderators from the UK Biobank. Specifically, we examined whether body mass index was moderated by lifestyle factors, whether lipids (LDL, HDL and triglycerides) were moderated by statins, and whether neuroticism, depression and anxiety were moderated by stressful life events. Our results suggest that the genetic architecture of anthropometric, cardiometabolic, and psychological variables are all sensitive to relevant moderators. Specifically, on a genome-wide level, we demonstrate that phenotypes are

more heritable in “risky” environments. On a locus level, we show interactions are detectable based on a combination of the phenotype, the moderator, and the genomic locus. Accordingly, ignoring GxE in GWAS has the potential to overlook critical aspects of the genetic architecture of many phenotypes. From a treatment perspective, GxE analyses can provide insights into potential situations where environmental factors amplify or mute the genetic associations, highlighting variants that may be sensitive to intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5059W: Examining the shared genetic architecture of keloid scars and uterine fibroids

Authors:

C. Greene[1], T. J. Lewis[2], G. Hampton[1], J. Hellwege[2,3], T. Edwards[2], D. R. Velez Edwards[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]VA Tennessee Valley Hlth.care System (626), Nashville, TN

Abstract:

Keloid scars and uterine leiomyomata, or uterine fibroids, are benign fibroproliferative disorders hypothesized to share genetic etiology. Despite their disparate anatomy and pathology, these conditions are frequently comorbid and are both characterized by uncontrolled cell proliferation and overgrowth of connective tissue. They are also both highly heritable and have similar disparities in prevalence. Given these parallels, we aimed to investigate the shared genetic architecture of keloids and uterine fibroids, leveraging summary statistics from recent genome-wide association study (GWAS) meta-analyses. We performed cross-ancestry and ancestry-stratified correlation analyses to determine potential differential effects. We estimated global genetic correlation using Linkage Disequilibrium Score Regression (LDSC) and utilized stratified LDSC to examine commonalities in partitioned heritability. We also conducted local genetic correlation analyses using Local Analysis of [Co]Variant Association (LAVA), filtering 2,495 pre-defined LD blocks to 124 containing significant variants from either the keloids or uterine fibroids GWAS. Finally, we performed gene set enrichment analyses of implicated regions using the Functional Mapping and Annotation of GWAS (FUMA) GENE2FUNC module. We observed strong global correlations between keloids and uterine fibroids. The cross-ancestry result was 0.63 (0.05, p=3.98×10-39), and ancestry-stratified analyses yielded similar values (mean=0.67 [0.11]). Both keloids and fibroids exhibited significant (p<9.4×10-4) partitioned heritability enrichment in functional annotations associated with increased enhancer activity (histone marks H3K4me1 and H3K27ac, SuperEnhancers), suggesting gene regulation via enhancers may

be important in fibroproliferative disease. Univariate analyses with LAVA resulted in significant ($p<4.03\times10-4$) local heritability for both traits in 81 loci, which were further considered in comparative bivariate analyses. We identified two significant ($p<6.2\times10-4$) and 15 nominally significant ($p<0.05$) locally correlated loci. GWAS variants meeting the suggestive threshold ($p<1\times10-5$) within nominally correlated regions were mapped to genes using the Ensembl Variant Effect Predictor. Enrichment analysis identified overlaps with GWAS Catalog gene sets related to obesity and immune measures, in addition to enrichment of gene sets involved in oncogenic pathways. This preliminary characterization provides functional insight into the shared etiology of keloids and uterine fibroids and creates opportunities for further investigation.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5060W: Exome sequencing of >1 million individuals identifies 232 genes associated with adult height,bridging the gap between common traits and rare diseases

Authors:

J. Kosmicki[1], L. Ganel[1], T. Joseph[1], J. Mbatchou[1], A. Ziyatdinov[1], D. Sharma[1], R. Collins[2], J. Torres[2], J. Emberson[2], J. Berumen[3], J. Alegre-Díaz[3], R. Tapia-Conyer[3], P. Kuri-Morales[4], O. Melander[5], J. Rotter[6], México City Prospective Study, GHS-RGC DiscovEHR Collaboration, MAYO-RGC Project Generation, UCLA-RGC ATLAS Collaboration, Colorado Center for Personalized Medicine - RGC Collaboration, Penn Medicine BioBank, S. Balasubramanian[1], J. Reid[1], A. Baras[1], G. Abecasis[1], J. Marchini[1], M. Ferreira[1], A. Locke[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Univ. Natl. Autónoma de México, México City, Mexico, [4]Tecnológico de Monterrey, Monterrey, Mexico, [5]Malmo Univ Hosp, Malmo, Sweden, [6]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA

Abstract:

Adult height is the model trait of human genetics and the first for which genome-wide association studies (GWAS) saturated the heritability attributable common variants. Yet, rare (minor allele frequency [MAF]<1%) variants remain relatively unexplored. To address this, we performed the largest exome sequencing study of height in 1,068,853 individuals (discovery: N=826,066 [23% non-European]; replication: N=242,787 [30% non-European]).
Overall, gene-based tests identified 232 genes (Bonferroni $P<1.75x10^{-9}$) independent of GWAS loci. The rarest and most deleterious variant class, singleton putative loss-of-function variants (pLoFs) in 18 genes were associated with the largest effects on height ranging from -20.9cm to

+12.9cm. While 57% of genes are the nearest gene to the GWAS locus, singleton pLoF effect sizes were 64x larger than the GWAS variant. Combining missense and pLoFs yielded the most genes, albeit with 34-65% lower effect sizes. Only six genes were associated solely through missense variants, half of which harbor known gain-of-function mechanisms.
Almost half of our associated genes also cause Mendelian disorders. Yet, the Mendelian disorders for 61 genes lack a stature component. For example, the autism gene, *CHD8*, increases height by 12.8cm ($P=3x10^{-10}$), but it is currently unknown whether the same variants contribute to both autism and height. Remarkably, we also identified 8 genes with effects on height larger than known Mendelian genes yet lack a Mendelian disorder.
While most gene-based associations encompass variants sufficiently rare to be independent of common variants, we show that conditioning on GWAS loci eliminated 11% of associations, largely from SKAT-O and ACAT-V. Despite this, SKAT-O and ACAT-V yielded 26 genes (11%) missed by burden tests, largely explained by variants with directionally opposing effects. We find a positive correlation between stronger gene constraint and larger pLoF effects in both directions suggesting height is under balancing selection. Recessive tests found only 1 significant variant independent of common variants: *CFTR* **Δ**;;F508 that causes cystic fibrosis (-7.7cm; $P=2x10^{-9}$). Lastly beyond solely identifying genes, rare pLoFs and deleterious missense variants explain 1.9% and 1.1% height's heritability, respectively.
Overall, our analysis of height in >1 million exomes identified 232 genes, constitutes the most comprehensive description of the rare variant architecture of height to date and bridges the gap between "common" traits, such as height, and rare diseases by both discovering genes with Mendelian-like effect sizes and expanding the characterization of known Mendelian disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5061W: Expanding genetic insight into steatotic liver disease contributed by Finnish founder effect

Authors:

O. Dwivedi[1,2], l. kullamaa[3], S. Kurki[1,3], E. Lahtela[1], E. Abner[4], J. Mehtonen[1], I. Karhiaho[3], S. Koskelainen[5], A. Bhowmik[6], R. Mettala[7], V. Mannisto[8], J. Koffert[6], J. Arola[9], M. Farkkila[3], EstBB, P. Palta[4], A. Palotie[1,10], N. Matikainen[3], M. Daly[1,10], T. Tuomi[1,2,3]; [1]Inst. for Molecular Med. Finland (FIMM), Helsinki Inst. of Life Sci. (HiLIFE), Univ. of Helsinki, Helsinki, Finland, [2]Folkhälsan Res. Ctr., Helsinki, Finland, [3]Abdominal Ctr., Helsinki Univ. Hosp., Helsinki, Finland, [4]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [5]Finnish Inst. for Hlth.and Welfare (THL), Helsinki, Finland, [6]Turku Univ. Hosp., Turku, Finland, [7]Tampere Univ. Hosp., Tampere, Finland, [8]Kuopio Univ. Hosp., Univ. of Eastern Finland, Kuopio, Finland, [9]Helsinki Univ., Helsinki,

Finland, [10]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA

Abstract:

Steatotic liver disease (SLD) affects 25-30% of the population but is underdiagnosed in hospital-based setting, with <3% *international classification of disease* (ICD) based prevalence, hindering genetic exploration (previous largest study: 9,491 non-alochalic fatty liver-NAFL cases). In the FinnGen SLD Study, we implemented phrase mining algorithms to abdominal imaging reports (>30% with multiple imaging readouts) of 55,387 individuals and identified radiological 'incidental' cases with liver steatosis or cirrhosis (8,475 cases). Additionally, we identified ICD-10 code based SLD cases (N=6,458). We performed a FinnGen SLD-GWAS with 14,933 cases vs. 485,415 controls, and a comprehensive meta-analysis (>10M variants), of 25,654 vs. 1,098,974 subjects by aggregating UK and Estonia biobank SLD case-control summary data.
The 'incidental' SLD cases were associated (OR=1.17; 1.14-1.19) with a genetic score (GS) of previously known NAFL-GWAS loci (n=7), and the GS was further enriched (top vs. bottom quintile, OR=1.64; 1.42-1.91) in those with multiple imaging diagnoses. The FinnGen GWAS identified 16 GWAS loci including a) six new Finnish enriched (>9 fold) variants (missense mutation in *CIDEB* and *GPT*, splice acceptor in *TRIM37* , non-coding in/near *CDS1*, *SFRP2* and intergenic-locus *MYVOE-FGF19*), and b) ten known GWAS loci previously associated with clinically-diagnosed NAFL or liver fat measures
(in/near *PNPLA3*, *TM6SF2*, *APOE*, *TRIB1*, *GCKR*, *MTARC1*, *AKNA*, *FTO* and *GPAM*). Conditional analysis of FinnGen SLD-GWAS further revealed 4 independent (region-wise significant) signals in known GWAS regions (*GPAM*, *AKNA*, *TM6SF2* and *PNPLA3*). The meta-analysis identified 11 additional loci: 5 new (in/near *ARHGEF19*, *PREB*, *MYOZ2*, *PLXNA4* and *KPNA2*) and 6 known loci (*COBLL1*, *TMC4*/*MBOAT7*, *MTTP*, *ERLIN1*, *PPARG* and *SERPINA1*).
The current study expands the genetic architecture of clinically diagnosed SLD (>2-fold increase in current NAFL-GWAS loci, total 27 with 11 new loci). Finnish-enriched variants implicate genomic regions with putative candidate genes (*CIDEB*, *CDS1*, and *FGF19*) in pathways (lipid droplet size regulation and fibroblast growth factor) targeted for new SLD drug development, offering opportunity to gain genetic-based insight into new potential SLD drug targets.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5062W: Exploring the genetic complexity of Alzheimer's disease in African ancestry using single-cell multi-omics integrative analysis ⭐

Authors:

D. Gingerich, D. Hodgson, E. K. Shwab, Z. Man, M. Garrett, G. E. Crawford, A. E. Ashley-Koch, O. Chiba-Falek; Duke Univ., Durham, NC

Abstract:

Recent single-cell multi-omics analyses have produced informative data advancing our mechanistic understanding of the genetic factors involved in the risk to develop Late Onset Alzheimer's Disease (LOAD). However, most studies were conducted in European subjects, while other populations are largely understudied. It's imperative to explore the underpinning genetics of LOAD in diverse populations and to gain insights into the shared (pan-ethnic) and distinct (ancestry-specific) genetic drivers of LOAD across multiple populations. Our overarching goal is to untangle the genetic complexity and genomic architecture of LOAD in diverse populations and to translate LOAD genetic association discoveries from different racial groups to causal mechanisms of disease. In this study, we analyzed 28 rapidly autopsied cortical tissues from African ancestry LOAD cases and matched controls and profiled simultaneously the transcriptome (snRNA-seq) and chromatin accessibility (snATAC-seq) using the 10x Genomics Multiome technology. 175,962 nuclei were clustered into 32 cell subtypes representing 8 major cell types. Differential expression analysis identified 1047 differentially expressed genes (DEGs) across all cell-subtypes, with the largest number of DEGs in the GABAergic neuron4 and Oligo2 (435, 307). Comparison analysis with European ancestry, revealed shared (pan-ethnic) and distinct (ancestry-specific) cell-subtype specific DEGs. Notable, GABAergic neuron4 showed the highest overlapped transcriptomic changes (86%), while OPC6 and Oligo1 demonstrated minimal overlaps. Next, using the snATAC-seq data we identified cell subtype specific candidate cis regulatory elements (cCRE) and integration with the corresponding cell types DEGs discovered cCRE-DEG pairs, providing in silico evidence for regulatory interactions and information regarding the functionality and directionality of gene dysregulation associated with LOAD. Furthermore, we pinpointed common and ancestry-specific impaired biological pathways in specific cell-types including disruption of synapse organization and postsynaptic signaling. The results enhance our understanding of the shared and distinct cell-subtype gene dysregulation networks and biological processes underlying LOAD in African vs European ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5063W: Exploring the impact of Life's Essential 8 and APOE4 allele on brain aging in the UK Biobank

Authors:

L. Feng[1], Z. Ye[2], S. Chen[2], T. Ma[1]; [1]Univ. of Maryland Coll. Park, College Park, MD, [2]Univ. of Maryland Baltimore, Baltimore, MD

Abstract:

Age-associated brain disorders and dementia are increasingly burdening society, yet brain aging does not progress uniformly across populations. The study examines the main effect of the American Heart Association's Life's Essential 8 (LE8) on the white matter brain age gap (WM BAG) and its potential interaction with the genetic risk factor APOE4, aiming to understand how these factors contribute to brain health and dementia prevention. The study analyzed the main effect of LE8 on WM BAG in 18,817 participants and the LE8-APOE4 interaction in 18,259 participants, all non-pregnant individuals of European ancestry (mean age 55.45) aged 40-69 without extreme white matter hyperintensities, from the UK Biobank. Multiple linear regression models adjusted for demographic, socio-economic, environmental, and medical conditions were used. WM BAG was computed using a machine learning model from regional fractional anisotropy measures derived from diffusion tensor imaging data (2014+). LE8 scores were based on lifestyle behaviors (diet, physical activity, smoking, sleep) and health factors (Body Mass Index, blood lipid, blood glucose, and blood pressure) at baseline (2006-2010). APOE genotype was categorized into APOE4 carriers (ε3/ε4, ε4/ε4) and non-carriers (ε2/ε2, ε2/ε3, ε3/ε3). Higher LE8 scores and being non-APOE4 carriers each correlated with a reduced WM BAG. Participants with ideal CVH status showed 490 days younger WM BAG than those with poor status. We also found that the joint health factors had a 31% greater effect on reducing WM BAG than the joint lifestyle factors. The APOE4 genotype showed complex modification effects; non-APOE4 carriers consistently showed reduced WM BAG with higher LE8 scores, whereas effects in APOE4 carriers were more variable and exhibited reduced efficacy. Furthermore, sex differences and specific interactions between LE8 and APOE4 status were observed, with APOE4 attenuating the protective effect of LE8 on WM BAG in early middle-aged women. The study underscores the importance of CVH in mitigating brain aging, as reflected by WM BAG. It emphasizes a holistic approach in brain health strategies, considering both modifiable factors and genetic predispositions, particularly in different genders.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5065W: FGFR2-related syndromes and de novo variants in a sub-Saharan African cohort with craniosynostosis

Authors:

L. Gowans[1], S. Obiri-Yeboah[1], G. O. Mensah[1], M. D. Asante[1], D. K. Sabbah[1], R. N. Larmie[1], A. A. Oti[1], G. Plange-Rhule[1], A. Butali[2], A. A. Adeyemo[3], M. L. Cunningham[4], P. Donkor[1]; [1]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [2]Univ. of Iowa, Iowa City, IA, [3]NIH, Bethesda, MD, [4]Univ. of Washington and Seattle Children's Res. Inst., Seattle, WA

Abstract:

Background: Metopic, coronal, sagittal and lambdoid sutures of human cranium remain patent in early postnatal life to allow cranium growth commensurate with brain development. Craniosynostosis (CS), or premature closure of sutures, can negatively impact brain development. CS has multifactorial etiology, and ~15% of cases are syndromic. Common genetic factors underlying CS etiology include variants in FGFR1, FGFR3, RUNX2, TCF12, TWIST1 and FGFR2. The current study ascertained the contribution of FGFR2 and other genetic variants to CS in a Ghanaian cohort presenting with phenotypes characteristic of FGFR2-related syndromes.
Methods: Whole exome sequencing (WES) was carried out on DNA from 8 Ghanaian families employing Illumina HiSeq at a read depth of 100X based on GRCh38. After quality control, alignment of reads and variant calling was carried out using Sentieon 202112.01 workflow and VCF files were normalized using bcftools 1.13. Variants with MAF ≤ 0.01 were filtered and pathogenicity of variants was ascertained using tools that include Ensembl VEP, CADD, SIFT, Polyphen-2 and ClinVar annotation.
Results and Discussions: Genomic analysis revealed that de novo genetic variants in FGFR2 may underly CS in all the 8 probands. A novel, de novo FGFR2 variant c.815_817del (p.Gly272del) was observed in an individual with Crouzon Syndrome. Three other FGFR2 variants were observed in three different probands with Crouzon Syndrome: c.1007A>G (p.Asp336Gly), c.1025G>C (p.Cys342Ser) and c.1024T>G (p.Cys342Gly). In the four probands with Apert Syndrome, three had c.755C>G (p.Ser252Trp) whereas one had c.758C>G (p.Pro253Arg) in FGFR2. Notably, all these variants occur very close to or within the extracellular Ig-like C2-type 3 domain of FGFR2, a domain that directly interacts with fibroblast growth factors and whose alternative splicing is crucial for ligand selectivity. We also observed interesting variants in other craniofacial genes, such as a de novo c.2269A>C (p.Asn757His) variant in ARVCF, a gene associated with Velocardiofacial Syndrome, in a proband with Crouzon Syndrome who had the p.Asp336Gly variant in FGFR2. However, the exact contribution of these other variants to the craniofacial phenotypes is uncertain.
Conclusions: Our observations suggest that de novo FGFR2 variants and related genetic syndromes contribute to the genetic etiology of CS in Africans. These observations are crucial for molecular diagnostics, genetic counseling, and clinical management of probands.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5066W: Functional dissection of genetic interactions contributing to phenotype variability of complex disorders

Authors:

J. Sun[1,2,3], S. Noss[1], M. Das[1], D. Banerjee[1], C. Smolen[1], F. Canzar[1], H. Bhavana[1], S. Girirajan[1,2,4]; [1]The Pennsylvania State Univ., University Park, PA, [2]Dept. of Biochemistry and Molecular Biology, University Park, PA, [3]Molecular, Cellular and Integrative BioSci.s Graduate Program, Univeristy Park, PA, [4]Dept. of Anthropology, Univeristy Park, PA

Abstract:

Phenotypic variability among individuals carrying the same recurrent copy-number variant (CNV) has posed significant challenges for genetic counseling, clinical management, and understanding the functional relevance of the CNV. Using the 16p12.1 deletion associated with a heterogeneous group of neurodevelopmental and psychiatric features as a paradigm, we proposed a “two-hit” model of CNV pathogenicity which suggests that secondary variants in the genetic background interact with the primary CNV to determine the phenotypic trajectory. To test this hypothesis, we used human induced pluripotent stem cell (iPSC) models, CRISPR-based methods, multi-omics strategy to understand the mechanisms of 16p12.1 deletion and to dissect the combined effects of the deletion in the context of rare genetic background. Cellular assays using neural progenitor cells (NPCs), immature and mature neurons differentiated from iPSC derived from 12 individuals from three large families and two healthy donors as well as isogenic CRISPR/Cas9-mediated deletion lines showed distinct dysregulations of neural function and neurogenesis that were specific to family and individual phenotypes. For example, we identified excitatory/inhibitory imbalance, early maturation of neurons and aberrant proliferation of NPCs that correlated with autism and head-size phenotypes observed in these families. Whole genome sequencing (WGS) data revealed that these changes also correlated with the presence of secondary variants affecting known neurodevelopmental genes such as *MECP2*, *SHANK3, EGF,* and *CTNND1*. Furthermore, RNA sequencing across all four cell types and ATAC-seq on iPSCs and NPCs showed shared, as well as cell-type- and secondary-variant-specific differential gene expression, including alterations in known risk genes (e.g., *FOXG1*, *CNTN6*, *KCNQ5*). Weighted gene co-expression network (WGCNA) analysis showed family-specific alterations in gene expression trajectory across differentiation stages and functional enrichments of co-expression genes were unique to families such as Wnt signaling pathway, Notch signaling pathway and translation-related pathways. Overall, by integrating WGS, RNA-seq, ATAC-seq, and cellular assays across families with variably expressive phenotypes, our study provides a framework to dissect the complex genetic basis and mechanistic consequences of complex disorders. These strategies are key to deriving genotype-

phenotype correlations and providing insights into developing potential personalized therapies in the future.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5067W: Functionally-informed fine-mapping and gene prioritization in height-associated loci

Authors:

A. Eliasen[1,2,3], L. Yengo[4], S. Vedantam[5], E. Marouli[6], A. Seth[7], Y. Okada[8], A. Wood[9], Finngen, GIANT Consortium height working group, J. Hirschhorn[7,10]; [1]Boston Children's Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Boston, MA, [3]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Copenhagen, Denmark, [4]The Univ. of Queensland, Brisbane, Australia, [5]Regeneron Pharmaceuticals, Tarrytown, NY, [6]Barts and The London Sch. of Med. and Dentistry Queen Mary Univ. of London, London, United Kingdom, [7]Boston Children s Hosp., Boston, MA, [8]Osaka Univ. / The Univ. Tokyo / RIKEN, Osaka, Japan, [9]Coll. of Med. and Hlth., Exeter Univ., Exeter, United Kingdom, [10]Dept. of Pediatrics and Genetics, Harvard Med. Sch., Boston, MA

Abstract:

Adult height is the classic model polygenic trait. It is highly heritable and easily quantifiable. In 2022, the GIANT consortium conducted a genome-wide association study (GWAS) of height in 5.4 million individuals of diverse ancestries and identified 12,111 independent genetic variants, which capture most of the heritability explained by common genetic variants. The associated genetic loci encompass nearly 1/4 of the human genome, implicating many potential biological mechanisms. Using a combination of computational approaches and additional genetic data sets, we aimed to more definitively prioritize likely causal genes and mechanisms. Specifically, we conducted a genome-wide fine-mapping analysis using SuSIE and polyfun in the loci harboring the 12,111 height-associated genetic variants using individuals of European ancestry from the UK Biobank (UKB). We integrated the fine-mapping results with genes prioritized by all 3 methods we used (PoPS, MAGMA, and DEPICT), a curated list of "OMIM genes" known to be involved in monogenic disorders of skeletal growth, and fine-mapping results from Finngen. The fine-mapping analysis in UKB identified 2,105 95% credible sets (CSs). Of these, 330 contained a single SNP, of which 15% were missense variants, compared to all lead variants where 4.65% were missense. We observed single SNP CSs to be more likely missense variants when an OMIM or prioritized gene was the nearest gene - 30% when OMIM and 26% when a prioritized gene. 10% of noncoding single SNP CSs had an OMIM or prioritized gene as the nearest gene,

generating specific hypotheses about causal noncoding elements and their target genes. By focusing on fine-mapped variants enriched in the Finnish population, we found examples of larger CSs in UKB that overlap with a single SNP missense variant CS in Finngen. For example, a CS in UKB near the *PAPPA2* gene overlaps with a single SNP CS in Finngen containing a missense variant in *PAPPA2* (which underlies short stature, Dauber-Argente type). In summary, we combined several computational and genetic approaches to nominate putative height-associated genetic variants and the genes they implicate. We find that single SNP CSs with missense variants are more likely to identify an OMIM or computationally prioritized gene, suggesting that single SNP CSs are useful signposts for relevant genes and variants. Lastly, we show that fine-mapping can further implicate specific genes by integrating the results with rare missense height-associated variants that are more prevalent in Finland.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5068W: G6PD deficiency masks the diagnosis of abnormal glucose tolerance: insight from the Africans in America Study

Authors:

A. Bentley[1], C. Kabeza[2], K. Ntabadde[2], C. Dubose[2], D. B. Sacks[3], C. Rotimi[4], A. Adeyemo[5], A. Sumner[2]; [1]Natl. Human Genome Res. Inst./NIH, Bethesda, MD, [2]Natl. Inst. of Diabetes and Digestive and Kidney Disorders, Bethesda, MD, [3]Clinical Ctr., NIH, Bethesda, MD, [4]NIH, Woodstock, MD, [5]NIH, Bethesda, MD

Abstract:

Introduction: Genetic association studies report lower hemoglobin A1C levels with G6PD deficiency variants. As A1C is sometimes used as a diagnostic test for abnormal glucose tolerance (AbGT; i.e. diabetes or prediabetes), it is critical to determine the clinical impact of G6PD deficiency on the diagnostic efficacy of A1C, especially in Africans, in whom deficiency variants are common and risk of AbGT is high. In this study, we aimed to: 1. compare the ability of A1C to diagnose AbGT in the presence and absence of G6PD deficiency; 2. test the concordance between the clinical assay for G6PD deficiency and the often unavailable, more expensive genotyping for *G6PD* risk variants. **Methods:** AbGT was determined by oral glucose tolerance test in 493 African-born individuals currently living in the Washington, DC area enrolled in the Africans in America cohort. Deficiency was predicted based on *G6PD* A-haplotype (n=451; rs1050828, rs1050829, rs76723693, and rs137852328) and/or clinical assay (n=243; fluorescent spot test). Diagnostic performance of A1C for AbGT (A1C ≥ 5.7%) was evaluated in those with and without G6PD deficiency. Concordance of haplotype-predicted and

clinically measured deficiency was evaluated in those with both measures (n=188). **Results:** AbGT occurred in 41% of participants (201/493). G6PD deficiency was present in 10% of participants (51/493), with the expected skew by sex of an X-linked trait (men: 14% [43/308]; women: 4% [8/185]). A1C was lower with G6PD deficiency (4.6 ± 0.5 vs. 5.5 ± 0.6, $P<0.001$). Sensitivity of A1C for diagnosis of AbGT in those without deficiency was 50%. In contrast, sensitivity in those with deficiency was 0% (0/16), indicating that no individuals with AbGT were detected based on A1C. Among those with AbGT and G6PD deficiency, A1C ranged from 4.2 to 5.5 (mean 4.9 ± 0.4). In concordance analysis, the clinical assay outperformed genotyping in 13 A- heterozygous women, as X-inactivation made deficiency status based on haplotype ambiguous. Otherwise, there was perfect concordance between haplotype-predicted and clinically measured deficiency. **Conclusions:** A1C levels were 0.9% lower in those with G6PD deficiency vs. those without. These falsely low A1C levels could lead to late diagnosis, undertreatment, and increased complications, contributing to existing disparities. Although deficiency predicted by genotype and by clinical assay was perfectly concordant in most individuals, genotype data is insufficient to predict deficiency in women heterozygous for the risk haplotype. Therefore, in screening this African population for AbGT, the clinical assay for G6PD was more cost-effective and informative than genotyping.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5069W: Gene clusters and patterns of neurodevelopmental phenotypes in cerebral palsy

Authors:

S. Myers[1], H. Shimelis[1], L. K. Walsh[1], M. T. Oetjens[1], C. L. Martin[1], D. H. Ledbetter[2]; [1]Geisinger, Danville, PA, [2]U. Florida Coll. of Med.-Jacksonville, Fernandina Beach, FL

Abstract:

The diagnostic yield of exome sequencing (ES) in cerebral palsy (CP) is ~30%. Genes implicated in cognitive and neurobehavioral developmental disabilities (intellectual disability (ID) and autism spectrum disorder (ASD), respectively) and epilepsy (EP) are shared across diagnoses. However, it is unclear whether genes implicated in CP are similarly non-specific in their phenotypic expression. This study examines the frequency and distribution of neurodevelopmental disability (NDD) phenotypes among patients with pathogenic variants to determine whether certain genes are more likely to cause CP or specific profiles of NDDs. Cases included in this study were identified from published reports with curated genomic and phenotype data available in the Geisinger Developmental Brain Disorder (DBD) Gene Database

(https://dbd.geisingeradmi.org/). For each gene, we calculated the percentage of cases with each of 4 NDDs: CP, EP, ID, and ASD. K-means clustering was used for hypothesis-free identification of groups of genes with similar phenotype profiles, first for 190 high-confidence (Tier 1) NDD genes and then for the 22 most common CP genes based on a published meta-analysis of ES studies. Clusters were named based on the NDDs observed in >30% of individuals with disease-causing variants in the included genes. Phenotype-based clustering of 190 Tier 1 NDD genes revealed 4 gene clusters: CP/ID (47%/45%; N=8 genes), EP/ID (75%/65%; N=41), ID only (88%; N=93), and ASD/ID (70%/48%; N=48). Analysis of the 22 CP genes revealed 5 clusters: CP only (90%; N=4), CP/ID (58%/53%; N=5), EP/ID (73%/ 48%; N=8), ASD/ID (55%/48%; N=5), and ID only (80%; N=3). The 2 clusters enriched for CP included 9 genes, 2 of which overlapped the CP/ID cluster in the previous analysis. Together, phenotype-based clustering of the 190 NDD and 22 CP genes showed that disease-causing variants in 15 unique genes (*ARID2*, *ARX*, *ATL1*, *ATP7A*, *CAMTA1*, *COL4A1*, *CTNNB1*, *GNAO1*, *HPRT1*, *KIF1A*, *KMT2B*, *SPAST, TH, TUBA1A, TUBB4A)* are associated with high prevalence of CP and ID but low prevalence of ASD. Like genes implicated in cognitive and neurobehavioral disabilities, those underlying CP are also associated with a spectrum of NDD phenotypes. Our results suggest that certain genes may be associated with a high rate of CP in addition to ID, whereas the prevalence of ASD appears to be low. However, determining whether certain genes are relatively specific for the combination of motor and cognitive disability phenotypes (as opposed to neurobehavioral and cognitive phenotypes) will require large-scale studies that include adequate phenotyping of all patients, allowing equal opportunity for diagnosis of each condition.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5070W: Gene-diet interaction analysis in UK Biobank identified genetic loci that modify the association between fish oil supplementation and the incidence of dementia

Authors:

Y. Lu, H. Xu, Y. Sun, S. Ihejirika, S. Song, Y. Shen, K. Ye; Univ. of Georgia, Athens, GA

Abstract:

Background: Dementia is a common disease in the elderly, caused by both genetic and environmental factors. *APOE* ε4 is well-known to increase the risk of dementia, and it has been shown to attenuate the protective association of fish oil supplementation and the incidence of dementia. However, the interaction between fish oil supplementation and other genetic factors is less well studied.

Methods: We first performed time-to-event genome-wide association study (GWAS) of all-cause dementia and two of its subtypes, Alzheimer's disease (AD) and vascular dementia, in the UK Biobank. GWAS were performed in all participants (N = 357,631) and in two subgroups with or without fish oil supplementation (N = 113,267; 244,364). The single nucleotide polymorphisms (SNPs) suggestively associated with dementia were then used to evaluate their interactions with fish oil status in Cox-regression models. Furthermore, we conducted gene set enrichment analysis to identify the tissue or cell-type specific patterns for these interaction signals.
Results: We identified 6, 5, and 2 genome-wide significant loci ($p < 5e-8$) for the incidence of all-cause dementia, AD, and vascular dementia, respectively. Most of them overlapped with previously known GWAS loci for AD and related dementia. A total of 1,819 SNPs with suggestive GWAS signals ($p < 1e-5$) were passed onto interaction analysis. 81 SNPs in 43 loci significantly modify the association between fish oil supplementation and dementia incidence ($p < 2.8e-4$ with Bonferroni correction). One locus overlapped with a known AD GWAS locus (*EED*/*PICALM*) and two overlapped with GWAS loci for circulating omega-3 fatty acids. Gene set enrichment analysis found that candidate genes of interaction signals demonstrated tissue or cell-type specificity in brain.
Conclusion: We identified 43 genetic loci which modify the association between fish oil supplementation and dementia.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5071W: Genetic analysis of chromosome X nominates a new risk locus in females with Lewy body dementia

Authors:

R. Chia; NIH, Bethesda, MD

Abstract:

Lewy body dementia (LBD) is the second most common neurodegenerative dementia in the western hemisphere. This includes both Parkinson's disease dementia and dementia with Lewy bodies. Epidemiological and neuropathological observations show that there is a sex-driven difference in the prevalence, symptom manifestation, and pathology in the brain of female and male patients with LBD. This suggests that beyond the autosomal genetic risks identified to date, there may be risk factors on the X chromosome that may contribute to these sex-related differences. To address this, we performed an X chromosome-wide association study (XWAS) and gene burden testing using whole-genome sequence data from 2591 LBD cases and 4391 controls. Given the difference in the recombination rate on the X chromosome compared to

autosomes, we calculated the number of haploblocks to set the appropriate threshold of significance for the analysis (0.05/6296 haploblocks = 7.94×10^{-6}). XWAS was performed on males and females independently, followed by a joint analysis of all samples. Given the known role of the APOE ε4 allele in dementia, an APOE ε4-conditional analysis was performed to investigate if there were independent risk variants with differential effects on males and females. We found a significant signal in intron 1 of the *MAP3K15* gene when conditioned on APOE ε4 dosage, which increases LBD risk in females by 2.4x (rs141773145, OR = 2.42, 95% CI = 1.65-3.56, P = 7.0×10^{-6}). We mapped the signal to a nearby enhancer that could regulate the expression of MAP3K15 in ganglionic eminence-derived primary cultured neurospheres. Ganglionic eminence cells give rise to the basal ganglia, thalamic, olfactory, and majority of cortical interneurons. Of relevance, learning and memory impairment was previously shown to be affected by dysfunction of the basal forebrain cholinergic neurons and γ-aminobutyric acid (GABA) interneurons, which are the very cells that stem from the medial ganglionic eminence structures in the brain during development. These findings implicate the MAPK pathways involved in neuroinflammatory responses, neuronal death triggered by alpha-synuclein aggregates, and tau and amyloid precursor protein phosphorylation. Rare variant burden testing showed nominal significance in differential enrichment of missense mutations in *TEX13A* in female LBD cases (P = 1.34×10^{-4}). In aggregate, our findings complement previously reported autosomal genetic risk factors for LBD and support the role of genetic factors as important drivers for sex differences in LBD. Sex-specific effects of genetic changes may start very early in life and eventually lead to Lewy body dementia.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5072W: Genetic factors associated with depression and cognitive decline

Authors:

J. Dennis, K. Singh, S. Bonnor, G. Boucher; Univ. of British Columbia, Vancouver, BC, Canada

Abstract:

Background: Depression in dementia is common, yet poorly treated by available therapies. We hypothesize that depression in dementia is etiologically distinct from primary depression, and that genetic analyses will elucidate these differences. Since the neurodegenerative processes that lead to dementia begin decades before diagnosis, we focus on depression and cognitive decline, before dementia onset. **Methods**: We used data on 11,958 participants in the Canadian Longitudinal Study on Aging who were aged 65-85 at baseline in 2010-2015, and who have been

followed prospectively thereafter. Cognitive function was assessed at up to three different time points. Depressive symptoms were assessed at up to five different time points. We used growth curve modeling to quantify cognitive function and depressive symptom trajectories spanning up to nine years. We created polygenic scores for Alzheimer's disease (AD) and major depressive disorder (MDD) using the most recent GWAS summary statistics. **Results**: Participants who were older at baseline experienced the fastest rate of cognitive decline, with no differences between males and females. Increasing genetic risk for AD (i.e., PGS for AD and/or *APOE* e4 genotype) associated with cognitive function trajectories, validating these trajectories as AD endophenotypes. In preliminary analyses, people who experienced new onset depression were more likely to experience cognitive decline than people with persistent or no depression. Next, we will test genetic risk for AD and PGS for MDD for association with new onset depression, while accounting for cognitive function trajectories. Our findings will help guide new treatments for depression in dementia.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5073W: Genetic Landscape of Postpartum Psychosis from Common to Rare Variants

Authors:

S. Jung, M. Caballero, A. Kępińska, T. Robakis, V. Bergink, B. Mahjani; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Postpartum psychosis (PP) is a rare but severe psychiatric condition, affecting about one to three per 1,000 mothers after childbirth. This condition poses significant risks, including increased chances of suicide and infanticide. Symptoms typically appear within one to two weeks post-delivery and include mood disturbances, psychotic episodes, cognitive disorganization, severe anxiety, and sleep disturbances. Using epidemiological family data, we have found a high relative recurrence risk among female full siblings of individuals with PP, indicating a substantial genetic or shared environmental effect on the risk of PP. However, the full impact of common and rare genetic variants on PP risk is not yet fully understood. Our recent study leveraged single-nucleotide polymorphism (SNP) array-based and whole-genome sequencing (WGS) data to assess the impact of command and rare protein-coding variants on the risk of PP.For the SNP array-based data, the dataset consisted of 368 cases and 1,840 matched controls after QC. To estimate heritability, the restricted maximum likelihood analysis compared genetic similarity variance to total phenotypic variance, with prevalence set at 0.003

and principal components as covariates.For the WGS data, the dataset included 301 cases and 3,010 matched controls after QC. To identify risk genes for PP, we performed the transmission and de novo association test (TADA), a Bayesian model to generate gene-level indicators of association evidence using rare variant counts, converted into the false discovery rate (FDR). Heritability estimates were 46% for autosomes and 11% for the X chromosome. Rare protein-truncating variants (PTVs), missense, and synonymous variants were compared between cases and controls. Average total counts of all variant types were comparable, but PTV signals were concentrated in the most conserved genes among cases (P=0.02, binomial test), suggesting rare PTVs may contribute to PP risk. This was confirmed by a SKAT-O burden test (P=0.038). The TADA result showed that HMGCR (3-Hydroxy-3-Methylglutaryl-CoA Reductase) was the PP risk gene with FDR<0.1.Our study provides compelling evidence for the significant genetic contributions to PP, highlighting the roles of both common and rare genetic variants. We identified HMGCR as a potential PP risk gene. HMGCR is vital for cholesterol and isoprenoid biosynthesis, crucial for neurotransmitter and hormone production, and is linked to autoimmune responses. Considering the autoimmune aspect of PP and hormonal changes postpartum, disruptions in cholesterol metabolism and immune function may significantly contribute to PP pathophysiology.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5074W: Genetically Predicted Clonal Hematopoiesis of Indeterminate Potential and Telomere Length and Risk of Multiple Primary Cancers

Authors:

S. He, A. Hubbard, W-Y. Huang, M. J. Machiela, S. Berndt; Natl. Cancer Inst., Rockville, MD

Abstract:

With the improvements in early detection and treatment, multiple primary cancers (**MPC**) incidence has risen in many countries. Clonal hematopoiesis of indeterminate potential (**CHIP**) is the abnormal expansion of clonally derived hematopoietic stem cells without evidence of hematological malignancy. Recent studies have identified common genetic variants associated with susceptibility to CHIP, including some related to telomere length. We previously discovered that longer genetically predicted leukocyte telomere length (**TL**) was associated with a strong increased risk of MPC. We sought to evaluate if genetically predicted CHIP was associated with MPC risk and if polygenic scores (PGS) for CHIP and TL could be used to predict MPC risk in two large cohort studies, the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial

(N=79650, including 4215 MPC cases) and the U.K. Biobank (N=488000, including 3246 cases). Multiple primary cases were individuals with two or more confirmed primary cancers at different sites. We estimated PGS for telomere length using up to 160 previously identified variants and the PGS for CHIP using 58 genetic variants. We used Cox proportional hazard models to estimate the association between the PGS and MPC risk. To assess the value of including these PGS in the risk assessment for MPC, we estimated the area under the curve (AUC) and evaluated three different models by ancestry: model 1 (M1): baseline risk factors, including age, sex, smoking, body mass index, and family history; model 2 (M2): baseline risk factors + TL PGS; model 3 (M3): baseline risk factors + CHIP PGS + TL PGS. Genetically predicted TL and CHIP were associated with an increased risk of MPC in a meta-analysis across ancestries (HR(95% CI):1.91 (1.63-2.24), $P=1.72x10^{-15}$ for TL PGS and 1.13 (1.07-1.19), $P=3.01x10^{-6}$ for CHIP PGS). The AUCs of the three models were 0.65, 0.71, and 0.71 for Europeans; 0.71, 0.73, 0.73 for Africans; 0.65, 0.71, and 0.73 for East Asians; 0.67, 0.68, and 0.71 for Hispanics for M1, M2, and M3, respectively. The addition of the TL PGS to baseline risk factors improved the AUC for all populations. The added contribution of the CHIP PGS to the AUC was small for the European and African populations, but it improved risk prediction for the East Asian and Hispanic populations. Overall, the best predictive model was the model including CHIP PGS and TL PGS (with AUC ~0.71-0.73 across ancestries). Our study establishes that both genetically predicted CHIP and TL are associated with an increased risk of developing MPC. This study demonstrates the potential clinical utility of including PGS in assessing the future risk of multiple cancer primaries.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5075W: Genome wide association study investigating increased risk of postpartum depression among women affected by immune-mediated inflammatory conditions

Authors:

A. Linerud[1], M. Davis[2], J. Denny[3]; [1]Brigham Young Univ., Provo, UT, [2]Brigham Young Univ, Provo, UT, [3]NIH, Bethesda, MD

Abstract:

There is a gross lack of research investigating minority health, not excluding women's health. Among some of the greatest risk factors for women includes postpartum depression (PPD). However, despite lasting consequences of PPD for both women and families, limited research has been allocated towards alleviating this condition. Furthermore, previous studies indicate a

need to investigate the significantly higher risk of PPD among vulnerable populations affected by immune-mediated inflammatory diseases (IMIDs), including MS, CD, RA, axSpA, and PsA. While 11-16% of new mothers develop a diagnosis of PPD, studies show individuals with IMID conditions have significantly higher rates of PDD (MS: aOR 2.0; RD (axSpA/PsA/RA): aHR 1.22; CD: aHR 1.12). Using EHR medical histories, health history surveys, and whole genome sequencing from the *All of Us* dataset we have identified a cohort of 741 IMID-affected women with a history of deliveries. The presence of at least two ICD-10, ICD-9, and SNOMED codes were used to identify IMID-conditions. The postpartum period was defined as 12 months following a first recorded instance of labor/delivery (CPT, ICD, and SNOMED codes specific to a labor and delivery event); postpartum depression was defined as an ICD or SNOMED code for PPD or major depressive disorder during the postpartum period. 186 individuals were affected by PPD. Our results are consistent with these previous studies; 25% of the women in our study have been affected by PPD, with disease specific rates ranging from 24-28%, compared to a general rate of 18% for the 25105 women with a history of deliveries across the *All of Us* database. Our genome-wide associate study will act as a novel inquiry, as previous research has studied IMID conditions only separately and clinically, without any genetic investigation involved. Our analyses include covariates of a history of obesity, history of alcohol abuse, access to health resources, living alone, number of pregnancies, duration between first pregnancy and occurrence of PPD. GWAS will identify variants specific to postpartum depression in individuals with IMID. Ultimately, this project takes a progressive stand for diversifying scientific research, especially the field of genetic research, to serve those of all societal classes, especially women and minorities, and provide the clinical understanding in hopes of improving the well-being and happiness of all.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5076W: Genome-wide association study of dermatochalasis reveals novel susceptibility loci

Authors:

M. Karjalainen[1,2], K. Rajueni[1], F. Koskimäki[3,4], FinnGen, Estonian Biobank Research Team, A. Elnahas[5], A. Reigo[5], K. Reis[5], T. Esko[5], P. Palta[5], J. Liinamaa[3,4], V. Saarela[3,4], J. Kettunen[1]; [1]Res. Unit of Population Hlth., Univ. of Oulu, Oulu, Finland, [2]Inst. for Molecular Med. Finland, Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, [3]Dept. of Ophthalmology and Med. Res. Ctr., Oulu Univ. Hosp., Oulu, Finland, [4]Res. Unit of Clinical Med., Univ. of Oulu, Oulu, Finland, [5]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia

Abstract:

Introduction: Dermatochalasis, an excess of skin on the eyelids manifesting as sagging eyelids, is a phenomenon commonly associated with aging. The degree of eyelid sagging is affected by genetic factors. The aim of this study was to detect genetic factors associated with dermatochalasis using a genome-wide association study (GWAS). **Methods:** We performed a GWAS meta-analysis of dermatochalasis using participants from FinnGen, Estonian Biobank and UK Biobank (total *n*=975,713; case *n*=13,200). GWAS was followed by identification of potential candidate genes at the associated loci. In addition, a phenome-wide association study (PheWAS) was performed and genetic correlations of dermatochalasis with different diseases and traits were assessed. **Results:** We identified eighteen dermatochalasis-associated loci at genome-wide significance. The associated loci comprised multiple genes having direct roles in skin biology. As an example, one of the genes, *ELN*, encodes elastin, which is a component of elastic fibers associated with elasticity of different tissues including the skin. Many of the associated loci also had genes implicated in different rare skin diseases, such as *ELN* and *LTBP1*, encoding latent-transforming growth factor beta-binding protein 1, associated with cutis laxa, a connective tissue disorder. Some of the associated loci had genes with roles in lipid-related biology, including *SORT1* and *MTX2* encoding sortilin 1 and metaxin 2, respectively. In PheWAS, we detected that the dermatochalasis-associated loci had previously reported associations mainly with eye and face morphology-related phenotypes. Finally, we detected that dermatochalasis had low-to-moderate genetic correlations with multiple traits, many of which were linked to smoking and pain. **Conclusions:** Our GWAS meta-analysis of dermatochalasis revealed genetic associations in eighteen loci. Many of these loci had biologically plausible candidate genes involved in skin biology. Our findings clarify the biological pathways associated with dermatochalasis and provide a solid foundation for further dermatochalasis-related studies.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5077W: Genome-wide association study of multiple neuropathology endophenotypes identifies novel risk loci and provides insights into genetic risk of dementia

Authors:

D. Fardo; Univ. of Kentucky, Lexington, KY

Abstract:

Background: Genome-wide association studies (GWAS) have identified >80 Alzheimer's disease and related dementias (ADRD)-associated loci. However, the clinical outcomes used in most prior studies belie the complex nature of underlying neuropathologies. As a complementary

approach, we performed GWAS on eleven ADRD-related neuropathology endophenotypes from three sources: National Alzheimer's Coordinating Center, Religious Orders Study and Rush Memory and Aging Project (ROSMAP), and Adult Changes in Thought (n=7,804 total participants with autopsy confirmation).
Method: Endophenotypes were harmonized across data sources and included Alzheimer's-related pathologies (tau neurofibrillary tangles, neuritic plaques, amyloid plaques, cerebral amyloid angiopathy [CAA]), non-Alzheimer's proteinopathies (Lewy body and TDP-43), vascular pathologies (microinfarcts, gross infarcts, cerebral atherosclerosis, and brain arteriolosclerosis), and hippocampal sclerosis. GWAS were performed separately on each data source and then underwent fixed-effect meta-analyses. We then used lead variants from a recent GWAS of ADRD to test association between ADRD-associated variants and neuropathology endophenotypes. Finally, we performed in silico functional genomic and epigenomic studies to prioritize genes and identify potential functional pathways by which genetic variants affect risk of neuropathology endophenotypes.
Result: Eight independent loci were identified with significant associations. Four loci were previously associated with ADRD (APOE, TMEM106B, GRN, BIN1) and four were novel (COL4A1, PIK3R5, LZTS1), including a new independent locus near APOE (APOC2) associated with CAA independently of APOE e diplotype. Thirty-nine additional loci showed suggestive associations. Directly testing the remaining known ADRD loci, 19 loci were significant for at least one neuropathology after FDR adjustment, 18 of which were concordant in effect direction with ADRD. Genetic colocalization analyses identified pleiotropic effects and various quantitative trait loci (QTL). For example, multiple methylation QTL and expression QTL also colocalized to either TMEM106B or APOC2. Methylation at two CpG sites colocalizing with CAA were in turn significantly associated with CAA in ROSMAP.
Conclusion: Our findings highlight genetic studies of neuropathology endophenotypes as a necessary next step to understand the mechanisms underlying genetic risk of ADRD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5078W: Genome-wide association study of non-alcoholic fatty liver disease based on ultrasound diagnosis in Korean population

Authors:

J. Park[1], S. Kwak[2], J. Choi[1], J. Lee[2], H. Lee[1], S. Lee[2], Y. Jo[2]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract:

Non-alcoholic fatty liver disease (NAFLD) is one of the most common liver diseases worldwide, caused by complex genetic and environmental factors related to fat accumulation in the liver. This study aims to uncover new genetic associations by meta-analyzing genome-wide association study (GWAS) data to identify significant single nucleotide polymorphisms (SNPs) associated with NAFLD using ultrasound images in multiple Korean cohorts. We performed a meta-analysis of GWAS data from four cohorts including a total of 11,730 participants. People with high alcohol consumption and patients with hepatitis B or C were excluded from the analysis. Covariates included age, age^2, sex, bmi and first 10 genetic principal components. NAFLD was assessed using abdominal ultrasound grading (0-3). The meta-analysis identified several SNPs that are genome-wide significance for NAFLD, resulting in four prominent peaks on the Manhattan plot, including rs738408 in the *PNPLA3* gene (p=4.29E-34) and rs2965185 in the *TM6SF2* gene (p=2.05E-12). Additionally, while not reaching the stringent genome-wide significance threshold (5E-8), the following SNPs showed notable associations: rs1260326 in the *GCKR* gene (p=2.88E-7) and rs2980888, a TRIB1-associated lncRNA near the *TRIB1* gene (p=4.01E-7). In this study, the SNP rs116979849, which has not been well-reported in previous NAFLD research, showed a p-value of 5.73E-08. This variant is an intronic variant of *SLC39A11*, notable for its involvement in the activation of zinc ion transmembrane transporter activity, which plays a crucial role in maintaining proper zinc homeostasis within cells. The participation of *SLC39A11* in zinc transport suggests that variants in this gene could potentially disrupt metabolic pathways in the liver, thereby cautiously indicating a risk for metabolic issues such as NAFLD. Further research involving larger cohorts is needed to validate these findings and understand the genetic underpinnings of NAFLD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5079W: Genome-wide association study of proteomic aging reveals a shared genetic architecture with longevity, reproductive timing, and age-related diseases

Authors:

A. Argentieri[1,2], C. Liao[1,2], F. Ivankovic[1,2], R. Ye[2], D. Bennett[3], Z. Chen[3], C. van Duijn[3], B. Neale[1,2], M. Daly[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Univ. of Oxford, Oxford, United Kingdom

Abstract:

We previously developed a proteomic age clock using 2,897 plasma proteins (Olink Explore) in the UK Biobank (n=45,441) that accurately predicts chronological age (Pearson r: 0.93) and is

associated with incidence of 18 major chronic diseases, multimorbidity, and all-cause mortality. Here, we describe the first genome-wide association study of proteomic aging, using genomic and proteomic data from European ancestry participants in the UK Biobank (n=38,865). Using imputed genotype dosages, we assessed common genetic variant (MAF > 0.01) associations with participants' proteomic age gap (defined as the difference between plasma protein predicted age and calendar age), adjusting for age, sex, and 10 genetic principal components. We identified twelve genome-wide (p<5e-8) significant loci associated with proteomic aging, including variants in CELSR2, CTRB2, SCARF2, and the HLA region that have been previously associated with parental longevity, fat mass, lipids, diabetes, heart disease, inflammation, liver/kidney function, and IGF-1.

We estimated the SNP heritability (h2) of proteomic aging to be 0.13 (SE: 0.02), indicating a large role of the environment in determining variation in proteomic aging. Genetic correlation analyses (using LD score regression) identified a strongly shared genetic architecture of proteomic aging with longevity (rg: -0.82; p: 0.01), defined as those living to an age above the 90th percentile based on census life tables (age range 87-116 years). We further used LD Hub to scan for genetic correlation between proteomic aging and 832 diverse traits, identifying 96 correlated traits at FDR p-value < 0.05. We found evidence for a shared genetic architecture between proteomic aging and earlier reproductive timing, including age when voice broke in males (rg: -0.22; p: 1.0e-4) and age at menarche in females (rg: -0.19; p: 4.3e-5). We also found a shared genetic architecture between proteomic aging and numerous age-related traits, including adult body fat mass (rg: 0.18; p: 6.8e-6), statin use (rg: 0.25; p: 4.6e-5, hospital diagnosis of acute myocardial infarction (rg: 0.27; p: 2.2e-3), and decreased cognitive function (rg: 0.21; p: 1.8e-3).

Overall, we demonstrate that the genetic architecture of proteomic aging is greatly shared with that of lifespan and longevity, and that this genetic architecture is further shared across reproductive and age-related traits. We plan to further: (1) replicate our GWAS and meta-analyze results across FinnGen (n=1,990), China Kadoorie Biobank (n=3,977), and MESA (n=6,431); and (2) create a proteomic aging polygenic risk score and explore its use for predicting future disease morbidity and mortality.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5080W: Genome-wide association study of responsiveness by SARS-CoV-2 mRNA vaccination in Japanese population

Authors:

K. Yamazaki[1], T. Naito[2], Y. Mashimo[1], T. Kageyama[1], S. Tanaka[1], T. Taniguchi[3], K. Matsushita[3], H. Igari[1,3], H. Hanaoka[3], K. Yokote[1], H. Nakajima[1], Y. Okada[2,4], Y. Onouchi[1]; [1]Chiba Univ. Graduate Sch. of Med., Chiba, Japan, [2]Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [3]Chiba Univ. Hosp., Chiba, Japan, [4]Graduate Sch. of Med., the Univ. of Tokyo, Tokyo, Japan

Abstract:

Recently, large-scale genome-wide association studies (GWAS) for antibody response against COVID-19 vaccines have reported significant associations within HLA region, but most studies were done in Western populations. To identify genetic determinants for antibody response against COVID-19 vaccine in the Japanese population, we conducted a GWAS. We genotyped 1,512 Japanese healthcare workers who received two doses of BNT162b2 mRNA COVID-19 vaccine in Chiba University Hospital from March to April in 2021 using Illumina Asian Screening Array, performed quality control for both participants and SNPs, and then conducted a whole genome imputation. Finally, association between 6 million SNPs with accuracy(R2) > 0.6 and MAF > 0.05 and post-vaccination antibody titers against SARS-CoV-2 S protein normalized by rank-based inverse normal transformation in 1,456 participants were assessed by a generalized linear regression analysis. As a result, we observed significant association signals in HLA class I (SNP X, $\beta = -0.279$, $P = 3.54\times10^{-14}$) and HLA class II (SNP Y, $\beta = -0.382$, $P = 1.37\times10^{-17}$) regions. In both loci, effects of the associated SNPs tended to be larger in females, and especially, the trend was prominent in the HLA class I region (females: n = 950, $\beta = -0.366$, $P = 9.95\times10^{-16}$; males: n = 506, $\beta = -0.114$, $P = 0.0665$). We are conducting HLA imputation and going to determine the leading HLA alleles in both loci. Our findings might lead to a better understanding of the mechanisms of sex differences in the antibody response and potentially of that in adverse effects after vaccination.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5081W: Genome-wide linkage analysis reveals a novel locus for increased plasma p-tau 181 levels on Chromosome 1 in the Amish

Authors:

P. Wang[1], Y. Song[1], A. Lynn[1,2], K. Miskimen[1], P. Whitehead[3], M. Prough[3], D. Dorfsman[3,4], A. Gulyayev[3], R. Laux[1], S. Fuzzell[1], S. Hochstetler[1], A. Zaman[3], L. Adams[3], L. Caywood[3], J. Clouse[3], S. Herington[3], Y. Liu[5], N. Moore[5], P. Ogrocki[6,7], A. Lerner[6,7], A. Griswold[3,4], J. Vance[3,4], M. Cuccaro[3,4], W. Scott[3,4], M. Pericak-Vance[3,4], J. Haines[1,2]; [1]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [3]John P. Hussman Inst. for Human Genomics, Univ.

of Miami Miller Sch. of Med., Miami, FL, [4]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [5]Dept. of Nutrition, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [6]Dept. of Neurology, Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [7]Dept. of Neurology, Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Abstract:

Plasma phosphorylated tau at threonine 181 (p-tau181) is a promising blood-based biomarker for Alzheimer Disease (AD). Emerging studies have shown that elevated p-tau181 is highly predictive for tau pathology at the early stages of AD. However, the genetics underpinning p-tau181 plasma levels remain elusive. Here, we investigated the linkage of genetic loci to plasma p-tau181 in the Midwestern Amish population with multigenerational pedigrees.
Genotyping was performed using the Illumina Expanded Multi-Ethnic Genotyping Array or Global Screening Array. Plasma p-tau181 levels were measured using the Simoa™ Advantage V2 assay. We applied extensive quality control to both genotypic and biomarker data, then standardized p-tau181 data using a rank-based inverse normal transformation after adjusting for age, sex, and study centers to remove potential confounding effects. Linkage analysis was conducted using the variance component model implemented in MERLIN, and loci with a logarithm of the odds (LOD) ≥ 1.86 (suggestive) and ≥ 3.30 (significant) were further examined.
Plasma p-tau181 was available for 977 Amish individuals aged 60 - 99 (mean age = 81.0, 59.1% females). Among them, 25.6% were classified as cognitively impaired by a clinical adjudication board, including 103 individuals diagnosed with AD. The AD group had significantly higher plasma p-tau181 compared to the cognitively unimpaired group (2.1 ± 0.9 pg/mL vs. 1.9 ± 0.8 pg/mL, p = 0.04). The linkage analysis conducted on 129 Amish sub-pedigrees identified 38 suggestive SNPs on Chromosome 9, and 104 suggestive and 16 significant SNPs on Chromosome 1, with a maximal LOD score of 3.53 at 139 cM (rs7365459). Although the significant loci were not found within protein-coding regions or near previously known AD loci, they were located close to the *AHCYL1* and *CSF1* genes, which are involved in regulating microtubule-associated protein tau and microglia function.
Our results revealed significant novel loci on chromosome 1 linked to plasma p-tau181 levels in the Amish, suggesting plasma p-tau181 is a valuable surrogate trait to complement the discovery of genetic components of AD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5082W: Genomic Signatures of Aging for Frailty in the NHLBI TOPMed Program ⭐

Authors:

L. Yanek[1], M. T. Lynch[2], A. K. Aragaki[3], T. Bartz[4], K. Lunetta[5], B. G. Windham[6], R. Casanova[7], D. Kiel[8], F. Liu[9], S. Liu[10], A. Newman[11], M. Odden[12], A. P. Reiner[13], S. Sahni[14], J. W. Wende[15], D. Arking[1], K. Bandeen-Roche[16], J. Walston[1], C. Sarnowski[17], A. Bick[18], J. Weinstock[19], J. Coresh[9], C. Kooperberg[3], J. Murabito[20], B. M. Psaty[4], R. A. Mathias[21], M. A. Taub[16]; [1]Johns Hopkins Univ Sch. of Med., Baltimore, MD, [2]NIH, Bethesda, MD, [3]Fred Hutchinson Cancer Ctr., Seattle, WA, [4]Univ. of Washington, Seattle, WA, [5]Boston Univ SPH Crosstown Ctr, Rm 313, Boston, MA, [6]Univ. of Mississippi Med. Ctr., Jackson, MS, [7]Wake Forest Univ. Sch. of Med., Winston-Salem, NC, [8]Marcus Inst. for Aging Res., Roslindale, MA, [9]Johns Hopkins Univ, Baltimore, MD, [10]Brown Univ., Providence, RI, [11]Univ. of Pittsburgh, Pittsburgh, PA, [12]Stanford Univ., Palo Alto, CA, [13]Univ of Washington, Seattle, WA, [14]Harvard Med. Sch., Roslindale, MA, [15]SUNY Buffalo, Buffalo, NY, [16]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [17]Univ. of Texas Hlth.Sci. Ctr., Houston, TX, [18]Vanderbilt Univ. Med. Ctr., Nashville, TN, [19]Univ. of Michigan, Ann Arbor, MI, [20]Boston Univ. Sch. of Med., Framingham, MA, [21]NIAID, Bethesda, MD

Abstract:

Background: Markers of somatic variation including shorter telomere length (TL), presence of clonal hematopoiesis of indeterminate potential (CHIP), and lower mitochondrial DNA copy number (mtDNAcn) can be assessed from whole genome sequencing (WGS) and reflect biological aging. To assess association of these dynamic genome metrics with frailty, we used WGS generated through the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. Methods: Four TOPMed cohorts (ARIC, CHS, FHS, WHI) were included in this study. Frailty was assessed with the Fried criteria (scored 0-5: 0 classified as robust, 1-2 as pre-frail, and 3-5 as frail). Association analyses used a three-level ordinal model. Measures of the dynamic genome were centrally called at the TOPMed Informatics Research Center. Cross-sectional multivariable ordinal regression models were run in each cohort with frailty as the dependent variable and TL, mtDNAcn, or CHIP as the independent variable of interest, adjusting for the difference between age at blood draw and age at frailty assessment, age at frailty assessment, sex, self-reported race, education, BMI, smoking, and cohort-specific covariates (clinic, pedigree). Results were meta-analyzed using regression coefficients and standard errors. Results: Participants (n=8788: n=727 frail, n=4350 prefrail and n=3711 robust) were 65.2 ± 10 yrs at time of blood draw, 73.6 ± 7.6 yrs at time of frailty assessment, with mean age difference of 8.4 ± 9.4 yrs between the two timepoints; 61.7% female; 85% European ancestry, 14.3% African ancestry and 0.7% Hispanic. Adjusting for all covariates, lower mtDNAcn was significantly associated with risk of worse frailty

status in FHS (β; 95% CI = -0.466; -0.7927, -0.1393) and the meta-analysis across the four cohorts (Z = -2.5, p = 0.01, three of four studies with consistent negative effect). Shorter TL showed significant association with risk of worse frailty in unadjusted models for CHS (-0.15; -0.25, -0.01) and FHS (-0.22; -0.38, -0.07); with the adjustment for all covariates effects were no longer significant. Neither TL nor CHIP showed significant association with frailty in meta-analysis. Conclusions: Only mtDNAcn revealed association with frailty after covariate adjustment, perhaps due to survival bias or interrelatedness of metrics. Future work will extend investigation to mosaic chromosomal alterations, epigenetic aging, and collapsed accelerated aging across multiple measures. Identification of specific aging related biological changes that may influence frailty may help to identify novel pathways for prevention and intervention development.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5083W: Germline genetics, disease, and medication exposures influence longitudinal dynamics of clonal hematopoiesis of indeterminate potential

Authors:

T. Mack[1], Y. Pershad[1], A. Bick[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Background: Clonal Hematopoiesis of Indeterminate Potential (CHIP) is characterized by the acquisition of a leukemia-associated mutation in a hematopoietic stem cell that leads to the formation of a sub-clonal blood cell population. CHIP increases risk of blood cancer, cardiovascular disease, and mortality. Larger CHIP clones are associated with worse outcomes. Clonal dynamics over time are not well understood due to few longitudinal cohorts.
Methods: Here, we performed targeted CHIP sequencing of serial blood samples in 3,000 individuals from the Vanderbilt BioVU biobank, identifying 711 individuals with 892 CHIP mutations, which is the largest longitudinal CHIP cohort to date. We quantified clonal growth rates and then examined the influence of genetic, environmental, and phenotypic factors.
Results: Mutations in the SRSF2/SF3B1 genes showed the fastest mean annual growth rate (25% per year), while DNMT3A mutations showed the slowest mean rate of growth (6%).While most clones (78%) grew over time, some remained stagnant or shrank. In individuals with two CHIP mutations, we modeled clonal trajectories of each clone and determined that half of the mutations co-occurred in the same cells and half were in distinct cells. Mutations in distinct

subsets grew faster.
We next sought to identify determinants of clonal expansion rate. We identified a common germline SNP rs1800057 in ATM increases growth rate across CHIP driver genes. Consistent with prior studies, rs2887399 in TCL1A reduces clonal expansion rate in individuals with TET2 CHIP. Long term exposure to colchicine, methylprednisolone, and denosumab was nominally associated with reduced clone growth rates. CHIP clonal expansion rate was not associated with smoking status, BMI, and pre-existing atherosclerosis or autoimmune disease. We investigated phenotypic consequences of clonal expansion rate. Faster clonal growth was associated with shorter times to abnormal myeloid cell parameters and an increased risk of myeloproliferative disease. To understand how myeloid malignancy risk changes over time, we calculated a CH risk score for the same individual over time. We find that frequent resequencing to monitor molecular evolution is less informative than serial monitoring of blood counts.
Conclusions: We demonstrate that faster growing CHIP clones have worse outcomes and that serial monitoring of blood counts are sufficient for clinical management of CHIP patients. Our approach highlights how multiple axes of risk including germline genetics, medications, driver gene, number of mutations, and clonal trajectory can be combined to provide estimates of disease risk in CHIP patients.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5084W: Girl case report: Nevus Spilus Syndrome coexisting with Ehlers-Danlos Syndrome and psychiatric disorders

Authors:

L. Sosa Amador[1], J. Reyes Márquez[2], M. Aceves Villalvazo[2], R. Hernández Ramírez[3], A. Rincón Sánchez[4,1], R. E'Vega Hernández[2], V. Jiménez Rodriguez[1], J. Ortega Águila[1], N. Dávalos Rodríguez[2,5,1]; [1]Centro Univ.rio de Ciencias de la Salud, Univ. de Guadalajara, Guadalajara, Mexico, [2]Hosp. Regional Dr. Valentín Gómez Farías, ISSSTE, Guadalajara, Mexico, [3]Centro Médico Natl. de Occidente, IMSS, Guadalajara, Mexico, [4]Inst. de Biología Molecular en Med. y Terapia Génica (IBMMyTG), Guadalajara, Mexico, [5]Inst. de Genética Humana "Dr. Enrique Corona Rivera" (IGH), Guadalajara, Mexico

Abstract:

Introduction:Nevus spilus syndrome (NSS) presents as a café-au-lait spots, superimposed by multiple smaller melanocytic lesions in association with neurosensory dysfunction. Described by Happle as an entity in 2002 and considered rare with higher prevalence in males. The case also presents Ehlers-Danlos syndrome (EDS) and psychiatric disorders, unrelated pathologies with no

concurrency previously described. Clinical report:The proposita 9 years-old 3rd. pregnancy product. Gestational diabetes during pregnancy, 35 weeks of gestation, polyhydramnios; with psychomotor delayed, learning disabilities, early seizures. She was treated by pediatric psychiatry for attention deficit disorder and depression. Was referred for hyperpigmented spot in the right inguinal region with multiple small nevi on the inside. Additionally presented significant interaction problems, tendency to injure her nails and fingers, impulsivity and poor social adaptation. Patient has no control over self-injurious behaviors such as onychophagia, presenting “a very high resistance to pain” or low sensitivity to it. Physical examination: somatometry above the 97th percentile; Obesity; Velvety soft skin with significant hxperelasticity and hyperlaxity; In the right inguinal area, a café-au-lait spot measuring of 12 cm x 10 cm, with hyperpigmented nevi of different size and tone inside, a "cigarette paper" type scarring. Bilateral hallux nails are absent due to self-excision. Pediatric psychiatry diagnosed intense emotional disturbances with depressive presentations, irritability, anxiety with distress and somatic disturbances such as headaches, emotional lability and severe self-injurious behavior problems. Treatment now focused on control of emotions and behavior with sertraline, aripiprazole and imipramine, with limited improvement. Conclusions:Nevus spilus has been associated with neurosensory abnormalities described by Torchia and Happle as a syndrome presenting alterations in pain perception. Kennedy et al. reported a systematic review that psychiatric disorders in EDS may be related with high prevalence rates. Moreover, the present report of NSS and EDS exhibits with high pain threshold and Psychiatric disorders. The relevance in this case is that there are no previous reports in the literature with coexistence of these pathologies in the same patient: Ehlers-Danlos Syndrome, Nevus Spilus Syndrome and Psychiatric disorders. Considering the low incidence of the pathology, we report the simultaneous disorders presentation in a girl, that could be part of a wide expression of the same entity.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5085W: Global microRNA Profiling in Human Urinary Extracellular Vesicles: Identifying Novel Biomarkers and Pathways for Diabetic Nephropathy

Authors:

H. Ali[1,2], M. Abu-Farha[1], M. Malik[3], i. alkhairi[1], P. Cherian[1], R. Nizam[3], T. Thanaraj[1], J. Abubaker[1], F. Al-Mulla[4]; [1]Dasman Diabetes Inst., Dasman, Kuwait, [2]Kuwait Univ., Kuwait, Kuwait, [3]Dasman Diabetes Inst., Kuwait, Kuwait, [4]Dasman Diabetes Inst., Kuwait City, Kuwait

Abstract:

Diabetic nephropathy (DN) is a significant chronic kidney disease and a primary cause of end-stage renal disease (ESRD). Small RNAs have emerged as promising diagnostic markers and therapeutic targets. Identifying dysregulated microRNAs (miRNAs) can help discover disease biomarkers and investigate downstream interactions, providing insights into the molecular pathophysiology of DN. This study analyzed small RNAs in human urinary extracellular vesicles (ECVs) from DN patients using small RNA next-generation sequencing (NGS). In this cross-sectional study, urine samples were collected from 88 participants, divided into three groups: Type 2 Diabetes (T2D) with DN (T2D+DN, n=20), T2D without DN (T2D-DN, n=40), and healthy individuals (n=28). The study focused on isolating urinary ECVs to extract and sequence small RNAs. Differentially expressed small RNAs (DE-sRNAs) were identified, followed by functional enrichment analysis. The study identified a distinct subset of 13 miRNAs and 10 Piwi-interacting RNAs (piRNAs) that were significantly dysregulated in the urinary ECVs of the DN group compared to other groups. Notably, miR-151a-3p and miR-182-5p showed unique expression patterns, being downregulated in the T2D-DN group and upregulated in the T2D+DN group, thus effectively distinguishing between the two groups. Eight driver genes were identified: *PTEN, SMAD2, SMAD4, VEGFA, CCND2, CDK6, LIN28B,* and *CHD1*. Our findings provide valuable insights into the pathogenesis of DN, uncovering novel biomarkers and identifying potential therapeutic targets that could aid in managing and potentially slowing the progression of the disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5086W: GWAS of neuropsychiatric symptoms in MCI/ADRD

Authors:

S. Vattathil[1], F. Blostein[2], T. Miller-Fleming[2], L. Davis[2], T. Wingo[1], A. Wingo[1]; [1]Univ. of California, Davis, Sacramento, CA, [2]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

About 85% of people with mild cognitive impairment (MCI) or Alzheimer's disease or related dementia (ADRD) experience one or more neuropsychiatric symptoms (NPS). ADRD-NPS include depression, anxiety, apathy, irritability, disinhibition, agitation, delusions, hallucinations, and sleep disturbances, and are associated with faster disease progression, more functional impairment, and earlier institutionalization. Better insight into the mechanisms of ADRD-NPS would aid in the design of effective therapeutics. To this end, we examined whether inherited genetic variation contributes to ADRD-NPS.First, we performed GWAS for each of nine ADRD-

NPS domains in 12,728 Alzheimer's Disease Research Center (ADRC) participants of European ancestries who had either MCI (23%) or ADRD (77%). ADRD-NPS were assessed with the Neuropsychiatric Inventory Questionnaire (NPI-Q), and the number of affected individuals for each domain ranged from 2,027 to 7,895. The GWAS model was adjusted for sex, genotyping batch, and 10 genetic PCs. Genome-wide significant SNPs were found in a 41 Kb region of chr19q13.32 for anxiety, agitation, apathy, delusions, and hallucinations. This region overlaps the *APOE* gene. An additional significant signal was found in *ADAMTSL1* for agitation. Next, replication was performed in 2,456 individuals with MCI or dementia from Vanderbilt University Medical Center's electronic health record-linked biobank, BioVU. While NPI-Q assessments were not available, we tested SNP association with NPS-ADRD status defined using ICD codes, and with outcomes based on prescribed antipsychotic or antidepressant medications. Two of three index SNPs in the *APOE* locus replicated in BioVU participants using the ICD code- and antipsychotic-based outcomes. The *APOE* ε4 allele is a known risk factor for cognitive impairment. Thus, people with more copies of *APOE* ε4 may be expected to have longer duration of cognitive impairment and higher ADRD-NPS burden. Given this, we performed conditional analyses and mediation tests to characterize the relationship of the chromosome 19 GWAS signal with *APOE* ε4 count and cognitive impairment. These results suggest that the association between the *APOE* locus and ADRD-NPS is not fully mediated by degree of cognitive impairment except in the case of apathy. In sum, *APOE* variation may play a role in ADRD-NPS independent of its impact on cognitive impairment, and this relationship warrants further investigation.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5087W: Harnessing Polygenic Scores to Personalize Smoking Cessation

Authors:

A. Chaloemtoem, Y. Chang, L. Fox, L-S. Chen, L. Bierut; Washington Univ. in St. Louis, St. Louis, MO

Abstract:

Background: Genetic factors significantly influence smoking cessation outcomes. Genome-Wide Association Studies (GWAS) on smoking behaviors have allowed for the creation of polygenic scores (PGS) to approximate an individual's genetic liability. Here, we assess the predictive power of PGS for smoking cessation in determining the timing of smoking cessation in a large population using the UK Biobank dataset. **Methods:** We analyzed data from 170,000 UK Biobank

participants with a history of smoking. PGS for persistent smoking (failed smoking cessation) was constructed using GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN2) summary statistics. The association between the PGS and smoking cessation was examined using logistic regression, adjusting for relevant covariates. Survival analysis was used to assess the probability of quitting smoking and the median age of quitting across PGS groups, where the median age as a clinically interpretable proxy for smoking cessation prognosis. **Results:** The PGS for persistent smoking showed a significant association with quit outcomes, with higher scores linked to reduced cessation success and later age of quitting. Individuals in higher PGS groups demonstrated significantly lower cessation success rates and were more likely to quit smoking at an older age. There was a notable difference in median cessation age between the top and bottom 10% PGS groups, indicating that those with higher genetic risk are less likely to quit early. **Conclusion:** Individuals with higher PGS are less likely to quit smoking successfully and quit at a later age and, therefore, will benefit more from targeted smoking cessation treatments. One goal of precision medicine is to guide and support individuals in making lifestyle changes and treatment decisions based on their individual genetic and environmental risk profiles. These findings underscore the clinical utility of incorporating genetic information into smoking cessation strategies, promoting more personalized and effective interventions, and potentially motivating behavioral change. This research contributes to the translational effort in precision medicine by validating the use of PGS to identify those at the highest risk for later smoking cessation.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5088W: Healthcare provider responses to high polygenic risk disease reports: Return of results experience at the University of Alabama at Birmingham

Authors:

J. Cortopassi, J. Booth, J. Cimino, B. Davis, J. Denton, B. Etheridge, B. Goff, R. Johnson, B. Korf, K. Rodriguez, E. Rothrock, R. Speer, N. Limdi; Univ. of Alabama at Birmingham, Birmingham, AL

Abstract:

Electronic Medical Records and Genomics (eMERGE) IV is a cohort study (25,616 participants enrolled, with 48% from underrepresented races and ethnicities) aiming to inform genomic risk of common chronic conditions for adult and pediatric participants and their healthcare providers.
eMERGE IV integrates genomic, family history, clinical, and lifestyle/behavioral data into a

comprehensive genome informed risk assessment (GIRA) report that is provided to participants and their healthcare provider. Additionally, participants with GIRA reports that indicate high genomic risk (i.e. high polygenic risk score, single-gene variants, and/or high integrated risk score for breast cancer) for any of the chronic conditions receive a one-to-one return from study staff.
Outcomes evaluated will include additional laboratory monitoring, earlier diagnostic screening, lifestyle counseling, new diagnoses, specialist referrals, and medication initiation or intensification after the GIRA has been returned to all participants and their provider. Analysis will assess differences in outcomes (at 6 months and 12 months) in participants with high genomic risk GIRA vs. not high risk across phenotypes and within each phenotype. While these data form the foundation for eMERGE outcome evaluations, they do not capture provider acknowledgement and planned action outside the follow-up timeframe.
To date, the University of Alabama at Birmingham (UAB) (one of ten sites in the eMERGE Network) has enrolled 2,957 participants and returned 2,070 GIRAs, of which 579 indicate high genomic or integrated risk. Herein, we present how provider acknowledgement and planned action can be captured to provide an additional facet for outcome evaluation.
At UAB, 36 providers have received at least one GIRA. We observed provider intention and planned action after GIRAs with high genomic risk were returned to the EHR. This was captured by assessing GIRA acknowledgement, addendum to the last clinic encounter note, physician-to-patient portal messages, provider review of prior lab or screening test results, and clinic appointment request. Provider intention and planned action varied across phenotypes.
Here, we describe the frequency and modalities via which providers utilized genomic information from GIRA reports to plan and implement actions recommended in the GIRA. We demonstrate that capturing provider intention and planned action should be included as a metric for assessing the clinical utility of high genomic risk reports, particularly high polygenic risk, in clinical care.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5089W: Hippocampal single nucleus transcriptomics of environmentally impoverished mice to reveal the addictome

Authors:

A. Nato[1], V. M. Borges[1], M. Leonardo[1], S. Brunty[1], J. Fan[1], A. Kastigar[1], J. F. Comstock[2], S. A. Towers[2], A. Walden[1], J. Lipovich[1], B. Gourley[1], J. Huffman[1], D. B. Lester[2], T. Salisbury[1], D. A. Primerano[1], J. Denvir[1], P. E. Dickson[1]; [1]Dept. of BioMed. Sci., Joan C. Edwards Sch. of Med., Marshall Univ., Huntington, WV, [2]Dept. of Psychology, Univ. of Memphis, Memphis, TN

Abstract:

Addiction, the irrepressible use of substances despite negative effects, is a complex disorder influenced by genetic and environmental factors. It is a major public health issue with ~46.8 million (16.7%) Americans (≥12 years old) fighting a substance use disorder in 2022. Genetics accounts for ~40-60% of an individual's risk of the disorder. We used environmental impoverishment (relative to enrichment) using differential housing as a model of chronic stress for various addiction phenotypes. Several genes may be associated with the susceptibility to effects of environmental impoverishment on addiction phenotypes. Identification of candidate genes may improve our understanding of the etiology of addiction, which will ultimately reduce the morbidity and mortality caused by addiction. Here, female C57BL/6J mice were housed in environmental enrichment vs impoverishment for 10 weeks. After differential housing, mouse hippocampal nuclei were isolated. 10X Genomics library preparation was performed followed by paired-end read sequencing with a P2 flow cell using NextSeq2000. We employed Seurat to perform single nuclei RNA-Seq analysis. UMAP dimension reduction identified 33 clusters. We mapped these clusters using the Zeisel brain dataset for cell identification, excluded heterogeneous clusters, which resulted to 5 broadly-based cell-type clusters, namely: ependymal astrocytes, interneurons, oligodendrocytes, pyramidal CA1, and pyramidal SS. Differential expression analysis was performed within each of these clusters. We identified 858 differentially expressed genes (DEGs) ($p\text{-adj} < 0.05$ and $|\log2FC| \geq 0.1$) where top genes include: Wdr17, Maml2, Ttr (ependymal astrocytes), Ttr, Pde10a, mt-Nd2 (interneurons), Ttr, Pde10a, Phactr1 (pyramidal SS) and Ttr, Camk2a, Gm44151 (pyramidal CA1). No significant gene was found in oligodendrocytes using these criteria. We performed gene set enrichment analysis on the DEGs for each cell type to identify significant pathways for each cluster. Candidate genes and pathways may provide insights regarding underlying mechanisms contributing to addiction risk.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5090W: HLA-focused genetic risk prediction in populations of diverse ancestry with type 1 diabetes

Authors:

D. A. Michalek[1], C. Tern[1], C. C. Robertson[1,2,3], W-M. Chen[1,4], S. Onengut-Gumuscu[1,4], S. S. Rich[1,4]; [1]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [2]Dept. of Computational Med. and Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [3]NHGRI/NIH, Bethesda, MD, [4]Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA

Abstract:

Background: Type 1 diabetes (T1D) is an autoimmune disease resulting from the destruction of the insulin-producing β-cells in the pancreas. Genetic factors contribute to ~50% of the total risk of T1D, with variation in HLA genes accounting for up to 50% of T1D genetic susceptibility. Most research and risk prediction has been conducted in European (EUR) ancestry populations.
Objective: We performed dense genotyping with imputation and estimated the HLA and SNP associations of T1D across populations to evaluate the HLA contribution to T1D GRS in populations of European (EUR), African American (AFA), admixed (AMR) and Finnish (FIN) ancestry.
Methods: Samples (N = 41,366) were genotyped on the custom fine-mapping ImmunoChip array, and each subject was assigned using k-means clustering and 1000 Genomes Reference Panel, grouped into EUR (N = 33,601), AFA (N = 3,877), AMR (N = 1,084), and FIN (N = 2,804) populations. SNP and two-field classical HLA class I and class II alleles within the MHC region were imputed using HLA-TAPAS. HLA analyses were performed using logistic regression models (PLINK v.1.9) while SNP- and HLA-allele-based genetic risk score (GRS) prediction used KING (v.2.3.2).
Results: In AFA and AMR populations, the most strongly associated T1D HLA risk allele was *DQA1*03:01* ($OR_{AFA} = 5.45$, $P_{AFA} = 9.3 \times 10^{-116}$; $OR_{AMR} = 2.91$, $P_{AMR} = 2.4 \times 10^{-21}$), differing from the known EUR ($OR_{EUR} = 5.33$, $P_{EUR} = 5.1 \times 10^{-1145}$) and FIN ($OR_{FIN} = 3.91$, $P_{FIN} = 8.8 \times 10^{-76}$) T1D-associated *DQB1**03:02 risk allele. The most associated HLA class II haplotype in AFA and AMR was *DRB1**03:01-*DQA1**05:01-*DQB1**02:01, in contrast to the EUR and FIN group, where *DRB1**04:01-*DQA1**03:01-*DQB1**03:02 was the most associated haplotype. SNP-based and HLA allele-based GRS had similar performance across ancestry populations. The prediction power of the T1D GRS derived from SNPs within the HLA region yielded $AUC_{AFA} = 0.86$, $AUC_{AMR} = 0.78$, $AUC_{EUR} = 0.87$, $AUC_{FIN} = 0.82$.
Conclusions: Although HLA alleles and haplotypes have been “gold standard” predictors for T1D risk, these results suggest that (1) SNPs are equivalent predictors across diverse ancestry groups and (2) T1D SNP-based, HLA-focused, ancestry-driven GRS exhibit strong performance across diverse ancestry populations.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5091W: Hypothesis-free Mendelian Randomisation Identifies New Metabolites and Athletes Metabolites Linked to Risk of ALS

Authors:

E. Alhathli[1], T. Julian[2], T. Moll[3], Project MinE sequencing consortium, D. Wang[4], P. Shaw[3], J. Cooper-Knock[5]; [1]The Univ. of Sheffield, Sheffield, United Kingdom, [2]The Univ. of Manchester, Manchester, United Kingdom, [3]Univ. of Sheffield, Sheffield, United Kingdom, [4]Imperial Coll. London, London, United Kingdom, [5]Univ. Of Sheffield, Sheffield, United Kingdom

Abstract:

ALS, a rapidly developing and invariably fatal neurodegenerative disorder, is characterised by the progressive loss of motor neurons. The vast majority of ALS cases are sporadic, caused by a combination of genetic and environmental factors. With two-sample Mendelian randomisation (MR), causal inference can be made between various exposures and disease risks, such as serum concentrations of the entire set of metabolites. Summary statistics from genome-wide association studies (GWAS) of 575 metabolites were compared with those from a GWAS of amyotrophic lateral sclerosis (ALS) consisting of 29,612 ALS patients and 122,656 controls. Unbiased MR using the inverse variance weighted (IVW) estimate and weighted median causally associated five metabolites with risk for ALS after stringent Bonferroni multiple testing correction. Two hits are in the carnitine synthesis pathway - carnitine has been previously linked to the severity of ALS via a role in energy metabolism within motor neurons; this is the subject of an ongoing clinical trial (ALSUntangled). CSF acetycanitine from a GWAS of 291 samples was also found to increase the risk in ALS as opposed to the protective affect in blood samples. Strenuous physical exercise has been shown to be a causative factor for ALS. A GWAS of 751 athletes metabolites from 490 elite athletes were compared with the same ALS GWAS using MR. Shared metabolites between the athletes and non-atheletes metabolites were analysed. Acetylcarnitine was significantly associated with reduced ALS risk in both GWAS. Acetylcarnitine was also significantly different between ALS patients and control samples in both 134 plasma and 110 CSF samples from Metabolon. The enrichment of all the significant metabolites with more than 10 instrument SNPs identified a pathway enrichment of aminoacyl-tRNA biosynthesis pathway.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5092W: Identification of circular RNA in Alzheimer's Disease

Authors:

B. Phillips, J. Sanford, D. Dikec, J. Budde, L. Ibanez, C. Cruchaga; Washington Univ. in St. Louis, St. Louis, MO

Abstract:

Introduction: Detection of Alzheimer's disease (AD) prior to the development of clinical symptoms is critical in delaying irreversible neurodegeneration. Circular RNAs (circRNAs) are covalently closed non-coding RNAs enriched in the brain and are differentially expressed in the brain tissue and plasma of AD patients. Based on circRNA stability, specificity, and tissue abundance, these transcripts are promising biomarkers for the diagnosis of AD. We previously found several differentially expressed circRNAs, such as *circHOMER1*, correlated with AD severity measurements in cortex samples. With eight brain regions across three cohorts and the current largest transcriptomic whole blood dataset, circRNAs were analyzed based on AD predictive ability and pathway enrichment of co-expressed linear mRNA genes.
Methods: We generated bulk RNA-seq data using cortex tissue (n=489) and whole blood (n=1,234) from the Knight ADRC. The brain replication cohorts included additional cortex regions from MSBB (n=698) and ROSMAP (n=864). Multi-tissue mixed models of AD phenotypes (CDR score, Braak tau score, clinically determined status) in brain and cross-sectional models of clinical status (405 AD, 829 CO) in blood were performed using the circRNAs identified by both DCC and CIRI2.
Results: Correlation with AD severity measurements had 129 cortex circRNA transcripts pass meta-FDR correction, including top circRNA *circHOMER1* (Braak Meta P=1.54x10^{-09}) and *circMAN2A1* (CDR Meta P=2.54x10^{-09}). A total of 34 blood circRNA passed status FDR correction using DCC and CIRI2 counts, including top circRNA *circDNAJC6* (P=3.16x10^{-09}) also differentially expressed in brain (CDR Meta P=3.38x10^{-04}). *CircGSE1* was expressed in the co-expression parietal module c1_29 (CDR P=4.16x10^{-05}) and included linear genes enriched for AD pathways (KEGG AD, 16/152 genes, adjusted P=1.26x10^{-06}) such as *APP*. Likewise, the top 10 CDR circRNAs in frontal cortex (BM44) MSBB samples had higher predictive ability in the combined circRNA + baseline model (AUC=0.954) compared to the circRNA (AUC=0.924) and baseline demographics (AUC=0.832) model.
Conclusion: The identified circRNAs co-expressed with linear genes enriched for pathways biologically relevant to AD. Identifying circRNAs with AD predictive ability across tissue types will further AD diagnostic research and lead to a better understanding of how AD-specific circRNA expression changes throughout the brain and blood.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5093W: Identification of common genetic risk loci associated with ischemic stroke in Taiwan

Authors:

Y-C. Wang[1,2], K-M. Liu[3], N-F. Chi[2], C-H. Chen[3], L-S. Lu[3], T-H. Lee[4], J-Y. Wu[3], I-H. Lee[2,1]; [1]Inst. of Brain Sci., Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, [2]Div. of Cerebrovascular Diseases, Neurological Inst., Taipei Veterans Gen. Hosp., Taipei, Taiwan, [3]Academia Sinica, Taipei, Taiwan, [4]Stroke Ctr. and Stroke Section, Dept. of Neurology, Chang Gung Mem. Hosp., Coll. of Med., Chang Gung Univ., Taoyuan, Taiwan

Abstract:

Genome-wide association studies (GWAS) have identified several common genetic risk variants associated with ischemic stroke in European ancestry populations. However, genetic susceptibility to ischemic stroke in Asian populations remains unclear, especially in earlier-onset ischemic stroke. We identified genetic risk loci associated with ischemic stroke and earlier-onset ischemic stroke under age of 60 in Han Taiwanese by utilizing the clinical and genetic data of whole-genome single nucleotide polymorphism (SNP) arrays from the Taiwan Precision Medicine Initiative (TPMI). We conducted the GWAS in 21,544 ischemic stroke patients and 267,198 stroke-free controls for a total of 447,496 variants, including 8,510 earlier-onset (< 60 years) cases and 185,302 matched controls after adjustment for age, gender, types of gene chip and the first ten principal components of the population. We found robust associations of common variants at four genetic loci near the *OR7A17*, at *OR7A10*, *BRD4* and *FGF5* genes in all ischemic stroke patients. Interestingly, only one of them near the *OR7A17* gene was not only associated with earlier-onset ischemic stroke, but also associated with late-onset ischemic stroke in those over 60 year-old. Apart from one reported risk loci *rs16998073* for stroke in the *FGF5* gene, the others located in chromosome 19 are novel loci. The results help to understand Asian population-specific genetic risk for ischemic stroke and provide insights into genetic susceptibility for earlier-onset ischemic stroke. Further studies are warranted.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5094W: Identification of Genes Associated with Alzheimer's Disease Using Novel Region Specific Microglia Cell Models

Authors:

R. Sale[1], P. Lin[1], G. A. Kaas[1], T. J. Hohman[1,2], E. R. Gamazon[1,3]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Memory and Alzheimer's Ctr., Nashville, TN, [3]Clare Hall, Cambridge Univ., Cambridge, United Kingdom

Abstract:

Transcriptomic prediction methods, such as PrediXcan, have enabled the detection of genes associated with complex traits such as Alzheimer's Disease (AD). Identifying gene associations can aid with drug development as these genes can be prioritized for downstream validation. PrediXcan relies on models built using bulk RNA sequencing data from 49 major tissue types generated through the Genotype-Tissue Expression Project (GTEx). While these models have enabled researchers to identify several tissue-specific gene associations so far, they are limited in their capacity to capture risk genes in highly heterogenous and potentially disease relevant cell types. Recent studies have shown that microglia, the resident immune cells of the brain, are highly relevant to AD. The ability to build microglia-specific transcriptomic prediction models could lead to the identification of microglia-specific and potentially disease relevant gene associations for AD. Here we leverage data from the microglia cell atlas (MiGA), which contains postmortem DNA and paired RNA sequencing from 90 individuals across four distinct brain regions. Using the MiGA, we built transcriptomic prediction models in microglia for each of the four sampled brain regions and tested for AD-genes. We then compared our gene associations in the microglial models with those obtained using PrediXcan Joint Tissue Imputation (JTI) models to determine which may be unique to the microglial models. We anticipate microglia gene associations obtained from this study will aid in understanding the underlying biology of AD as well as prioritize gene candidates for therapeutic development.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5095W: Identification of multiple new loci associated with late-onset Alzheimer's disease in Japanese

Authors:

R. Mitsumori[1], Y. Asanomi[1], A. Miyashita[2], T. Morizono[1], N. Hara[2], D. Shigemizu[1,3], T. Ikeuchi[2], S. Niida[4], K. Ozaki[1,3,5]; [1]Med. Genome Ctr., Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [2]Dept. of Molecular Genetics, Brain Res. Inst., Niigata Univ., Niigata, Niigata, Japan, [3]Dept. of Cardiovascular Med., Hiroshima Univ. Graduate Sch. of BioMed. and Hlth.Sci., Hiroshima, Hiroshima, Japan, [4]Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [5]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan

Abstract:

Late-onset Alzheimer's disease (LOAD), the most common form of dementia, is a highly heritable neurological disease. Identification of genetic etiologies that confer risk of LOAD contributes to elucidate the pathogenesis. Although meta-analysis of LOAD genome-wide association studies (GWAS) in Caucasian discovered many candidate risk loci, the genetic

architectures differ from among ethnicity, large-scale GWAS with another population, particularly Asians, remains incomplete. Here we acquired the whole genome genotyping data obtained using Asia screening array with 21,782 Japanese subjects, including 5,066 LOAD cases, recruited at the National Center for Geriatrics and Gerontology (NCGG) biobank, Niigata University and Biobank Japan, and conducted large-scale LOAD GWAS in Japanese. We found associations for multiple new risk variants and LOAD with genome wide significance ($P < 5.0\times10^{-8}$) and suggestive significance ($P < 1.0\times10^{-6}$) in the GWAS. We also confirmed the associations for *APOE*, *SORL1* and several other known loci and LOAD in Japanese. We further performed replication analysis with additional Japanese subjects comprising approximately 1,400 LOAD and 4,000 cognitive normal subjects, and confirmed the associations with statistical significance at the multiple loci. We are currently conducting the trans-ethnic meta-analysis with UK biobank cohort, functional annotation for the GWAS as well as construction of polygenic risk scores. Our findings provide new genetic components and biological and physiological insight into LOAD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5096W: Identifying Genetic Determinants for Multi-Ancestry Portability of LDL Polygenic Risk Scores

Authors:

K. Iyer[1], A. Wang[1], H. Alkhairo[2,1], N. Risch[3,4], C. Tcheandjieu[5]; [1]J David Gladstone Inst.s, San Francisco, CA, [2]UCSF, San Francisco, CA, [3]Univ California San Francisco, San Francisco, CA, [4]Inst. for Human Genetics, Univ. of California, San Francisco, CA, [5]Gladstone Inst., San Francisco, CA

Abstract:

GWAS have identified numerous SNPs associated with CAD risk, including key biomarkers like LDL. Aggregating LDL-associated genetic variants into PRS and integrating them into clinical risk assessment tools can enhance the predictive capability for dyslipidemia and CAD. However, poor multi-ancestry PRS portability poses a significant challenge, primarily due to the Eurocentric focus of genetic studies and insufficient understanding of ancestry-specific and shared causal variants. To address this knowledge gap, we used a cross-ancestry approach to dissect LDL genomic loci.
We utilized LOGODetect to identify genomic regions exhibiting local genetic correlation across 5 ancestries (EUR, AFR, EAS, AMR, and SAS), leveraging their LDL summary statistics. The algorithm effectively identified 17 regions with concordant associations shared among 4+ ancestries, including well-known LDL-related genes, such as *PCSK9*, *LDLR*, *APOB/C/E*, *ANGPTL3*, *HMGCR*, and *FADS1/2*. We then examined key factors

affecting PRS portability, including SNP effect sizes, MAF, and LD differences among ancestries. Across these shared regions, GWAS significant SNPs showed comparable effect size trends and MAFs across ancestries. Using 1000 Genomes (TGP), we analyzed the LD structure of these regions; while the mean R^2 within each region was relatively lower in the AFR superpopulation, the overall trend remained consistent across ancestries.
We further identified ancestry-specific LDL-associated variants not shared among populations by analyzing the set difference between LOGODetect results and summary statistics. We found 64636 EUR-, 2323 AFR-, 952 EAS-, 137 HIS-, and 13 SAS-specific LDL-associated GWAS significant SNPs, with the SNP count correlating directly with sample size. These SNPs were analyzed across all TGP superpopulations to calculate LD and MAF. Ancestry-specific SNPs exhibited the highest MAF within their respective TGP superpopulations, with several SNPs missing in other populations. LD patterns also varied significantly, with pronounced differences in the AFR superpopulation.
Having concretely assessed the factors contributing to the low transferability of LDL PRS, our next steps involve developing a robust PRS using the identified shared and distinct SNPs for each ancestry.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5097W: Identifying rare variants for cognitive preservation on chromosome 2 in the Midwestern Amish

Authors:

Y. Liu[1], A. Lynn[2], Y. Song[2], L. Main[2], W. Wang[2], K. Miskimen[2], S. Fuzzell[2], S. Hochstetler[2], R. Laux[2], L. Caywood[3], J. Clouse[3], S. Herington[4], P. Wang[2], A. Gulyayev[3], D. Dorfsman[3], N. Moore[5], M. Prough[3], A. Zaman[3], L. Adams[3], P. Whitehead[3], P. Ogrocki[2], A. Lerner[2], J. Vance[3], M. Cuccaro[6], W. Scott[3], M. Pericak-Vance[7], J. Haines[8]; [1]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Case Western Reserve Univ., Cleveland, OH, [3]Univ. of Miami, Miami, FL, [4]Univ. of Miami Miller Sch. of Med. Hussman Inst. for Human Genomics, Miami, FL, [5]Case Western Reserve Univ., Cleveland Heights, OH, [6]John P. Hussman Inst. for Human Genomics, Miami, FL, [7]Univ. of Miami Miller Sch. of Med, Miami, FL, [8]Case Western Reserve Univ, Cleveland, OH

Abstract:

Cognitive preservation is a phenotype observed in older individuals who are at high risk of cognitive decline, mostly due to aging and genetics, but remain cognitively unimpaired (CU). Identifying genetic variants associated with cognitive preservation may offer new insights into treatment and prevention for Alzheimer disease and other related dementias (ADRD). The

Midwestern Amish are a unique population for identifying protective rare variants (RVs) due to their large families and reduced environmental variation. Previous linkage analysis in a large sub-pedigree of Amish individuals ('chr2-pedigree') identified a novel locus, rs1402906 (HLOD = 4.87) on chromosome 2 near *LRRTM1* and *CTNNA2,* as associated with cognitive preservation. This study aims to fine map this region by examining RVs around this locus. Amish subjects (age > 65 yrs) who were CU and had a first degree relative with dementia were ascertained for this study. Cognitive status was determined using multiple neurocognitive assessments and confirmed by clinical adjudication as CU or cognitively impaired (CI). After extensive quality control, 1,011 whole-genome sequenced samples (670 CU, 341 CI) were available for RV detection, including 16 CU subjects from the chr2-pedigree. RVs with a minor allele count > 5 in the chr2-pedigree were selected for further investigation. A total of 26 RVs near rs1402906 were present in just 12 of the Amish subjects, all of whom were from an extended version of the chr2-pedigree (7 CU, 5 CI). None of the more distantly related members of this Amish population carried these RVs. The 7 CU individuals were 78-92 years old (mean=84.9±4.8), and the 5 CI individuals were 83-94 years old (mean=87.4±3.8) at their latest assessments. These CI individuals may demonstrate delayed age of onset of cognitive decline compared to all Amish (mean = 84.0 ± 5.6). Among the 26 RVs, 5 overlapped with *LRRTM1* and 13 overlapped with *CTNNA2*. All RVs were annotated as intronic variants, with some also being non-coding transcript variants. High methylation activity in brain regions were seen in chr2:77225201 (value > 0.9) and chr2:77555241 (value > 0.8) based on whole genome bisulfite sequencing DNA methylation data from the GTEx database. This chr2-linked Amish pedigree, which previously revealed a significant signal for cognitive preservation on chromosome 2, provides additional evidence that genetic variants in this region may play a crucial role in preserving cognitive function, potentially through mechanisms involving methylation regulation.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5098W: Impact of *HLA* diversity in classifying diabetes subtypes in Indians using type 1 diabetes genetic risk score

Authors:

A. Sankareswaran S[1,2], P. Kunte[3,4], D. Fraser[5], M. N. Weedon[5], C. S. Yajnik[3], R. A. Oram[5], G. R. Chandak[1,2]; [1]CSIR - Ctr. for Cellular and Molecular Biology, Hyderabad, India, [2]Academy of Scientific and Innovative Res., Ghaziabad, India, [3]King Edward Mem. Hosp. and Res. Ctr., Pune, India, [4]Western Sydney Univ., Penrith, Australia, [5]Univ. of Exeter, Exeter, United Kingdom

Abstract:

An accurate diagnosis of Type 1 (T1D) and Type 2 diabetes (T2D) is critical to ensure appropriate treatment and avoid late-onset complications. Lately few studies have developed T1D genetic risk scores (GRS) that can discriminate between T1D and non-T1D in the Europeans. We recently replicated the discriminative ability of a 9-SNP GRS in Indians, albeit with lower power. In this study, we generated the latest 67-SNP GRS (GRS2) from the genome-wide genotype data on a well-characterized cohort comprising 593 T1D, 1116 T2D and 317 normal individuals and assessed its ability in Indians using area under curve (AUC) of receiver operator characteristic (ROC) curve. The GRS2 was highly discriminative of T1D and non-T1D in Indians but still lower compared to the Europeans (AUC: 0.86 vs 0.93; P &lt 0.001) which may be attributed to the significant variations among the risk-allele frequency at associated variants. Since *HLA* is a major component of T1D GRS2, we explored its diversity in Indians, especially on *HLA DQ* haplotypes, which were imputed using HIBAG R package and association analysis (logistic regression with T1D status as the outcome) was conducted. The association was directionally consistent but differed in the magnitude and frequencies between Indians and Europeans. Notably the frequency of the protective *DQ* haplotype 01:03~06:01 was higher in Indians (20%) compared to Europeans (1%). Conversely, the most frequent European haplotype 01:02~06:02 had lower frequency in Indians (15% vs 1.5%). The *DQ* haplotype 05:01~02:01 had twice the effect size in Indians (2.05 [1.67-2.48], P = 1.37 x 10^{-23}) than in Europeans (1.16 [1.10-1.25], P &lt 2.2x10^{-16}) whereas an inverse relationship was noted for 03:01~03:02 (Indians = 0.92 [0.56 - 1.30], P = 9.6 x 10^{-7}; Europeans = 1.83 [1.76-1.90], P &lt 2.2 x 10^{-16}). Further evidence on the importance of *HLA* heterogeneity came from the observation that T1D GRS had better discriminative ability in antibody positive T1D (AA Pos) than antibody negative (AA Neg) patients (AUC: 0.86 vs 0.75; P = 4.6 x 10-4). The GRS2 with only HLA variants had a higher classification power for AA Pos compared to AA Neg (AUC: 0.84 vs 0.75; P &lt 0.001) whereas non HLA had similar AUC (0.64 vs 0.63, P = 0.59). In summary, wewewe confirm the utility of GRS2 in diabetes type classification in Indians and the lower discriminative power compared to the Europeans could be attributed to significant differences in the effect size and frequencies of *HLA DQ* haplotypes between them. Our study highlights the importance of ancestry in genetic association studies and the need to better capture *HLA* diversity for GRS based robust diagnosis of diabetes in Indians.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5099W: Impact of lifestyle modifications on individuals with high polygenic scores for obesity: insights from the TMM CommCohort Study

Authors:

Y. Sutoh, T. Hachiya, Y. Otsuka-Yamasaki, S. Komaki, S. Minabe, H. Ohmomo, M. Sasaki, A. Shimizu; Iwate Med. Univ., Yahaba, Japan

Abstract:

Obesity and overweight significantly increase the susceptibility to lifestyle-associated diseases. The efficacy of lifestyle changes in mitigating the risk of obesity among individuals with substantial genetic predispositions remains unclear. This study aimed to evaluate the impact of lifestyle modifications on outcomes in individuals with high polygenic scores (PGS). Four datasets from the Tohoku Medical Megabank Community-based Cohort (TMM CommCohort) were used. One subset (n = 9,958) facilitated the selection of an optimal model for computing the PGS, whereas the others (n = 69,341) supported a meta-analysis for validation and risk assessment. Model development involved estimating weights based on GWAS summaries of 158,284 individuals from BioBank Japan. We applied two modeling approaches; the P+T and LDpred methods, generating 24 and seven models under various hyperparameters, respectively. Among these, the model using LDpred at ρ = 0.03 yielded the highest adjusted area under the curve (0.63) for predicting obesity, which was used for further analysis. Validation results revealed that the odds ratios (OR) for obesity risk in the intermediate (11th-90th percentiles) and high PGS categories (91st-100th) were 2.27 (95% CI; 2.12-2.44) and 4.83 (4.45-5.25), respectively, relative to the low PGS category (1st-10th). To confirm the robustness of our findings, alternative metrics, such as waist circumference, body fat percentage, and visceral fat area, were evaluated and showed moderate associations (Spearman's ρ; 0.131-0.201) compared to body mass index (Spearman's ρ; 0.131-0.254). Trend analysis indicated that increased leisure activities significantly lowered the risk of obesity in all genetic risk categories, even in the high PGS group (OR; 0.9 [0.87-0.94]). Conversely, sodium intake was positively correlated with obesity in all genetic risk strata, yielding an OR of 1.24 [1.17-1.31] in the high PGS category. These findings demonstrate the effectiveness of lifestyle interventions in individuals with high genetic predisposition to obesity and offer insights for incorporating PGS into preventive healthcare strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5100W: Inferred course of hospitalized COVID-19 progression from longitudinal multi-omics data associates with disease severity, patient survival, dynamic molecular changes, and genetic variation.

Authors:

R. Thompson, G. E. Hoffman, P. O'Reilly, D. M. Del Valle, The Mount Sinai COVID-19 Biobank Team, S. Kim-schulze, R. Sebra, M. Merad, S. Gnjatic, S. Eric, A. W. Charney, N. D. Beckmann; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Modeling the natural history of COVID-19 from infection to resolution is an essential prerequisite to understanding COVID-19 and optimizing interventions at actionable points in the disease course. Densely sampled longitudinal cohorts are ideal for this but rarely (if ever) achievable in practice given limited tissue access, high costs, and ethical issues. Cross-sectional studies, which by design lack temporal information, are thus the norm. A growing number of COVID-19 research cohorts employ sparse longitudinal sampling (i.e., many individuals and few samples per individual) with high-dimensional data being generated. The short sampling periods across individual disease courses in these cohorts are typically aligned based on time of symptom onset, assuming identical rates of progression. This is suboptimal because time of symptom onset and rate of progression are both highly variable. There is thus an urgent need for new methods to analyze these increasingly common sparse longitudinal data sets. We devised a new machine learning method that leverages temporal information within each subject to infer a unified progression axis from sparse longitudinal data, along with an unbiased estimate of the accuracy of this axis. We validated our method in simulated and real data with known timelines. Using a large cohort of COVID-19 patients with sparse longitudinal whole blood RNA sequencing (RNA-seq), we inferred an axis of COVID-19 progression. The axis correlates with real time within subjects (cross-validated Kendall's τ;=0.49); is validated by the axis inferred from Olink data of the same samples (R^2 = 0.51, p = 7E-135), replicates in an external data set (p = 2.1E-7), correlates with patient-reported time of symptom onset (internal holdout: p = 6E-5; external validation: p = 1E-12); and is specific to COVID-19 patients (Wilcoxon test vs. uninfected controls, p = 2E-3). Comparing to clinical measures, the axis correlates with increasing total monocytes (p = 1E-10) and associates nonlinearly with severity, ICU admission, and SOFA score, each of which reaches a peak in the middle of the disease course. Furthermore, rate of progression along the timeline correlates positively with survival (Wilcoxon test p = 8E-06) and associates with 8 genome-wide significant loci (p < 5E-8). 10,171 genes' RNA expression associate either linearly or non-linearly with the Olink-inferred timeline (FDR ≤ 0.05). In conclusion, our inferred and validated time axis of COVID-19 progression associates with severity, survival, and genetics, demonstrating the value of our method for reconstructing molecular trajectories of disease and investigating the biology of disease progression.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5101W: Integrating Brain Cell-Type Specific Omics Data to Identify Causal Genes and Cell Types for Parkinson's Disease Associated Loci

Authors:

J. Choi[1], K. Fletez-Brant[2], T. Bhangale[2]; [1]Roche, Toronto, ON, Canada, [2]Genentech, South San Francisco, CA

Abstract:

Genome-wide association studies for Parkinson's Disease (PD) have identified over 90 loci [1], providing insights into the underlying pathways. However, linking causal variants to specific effector genes and cell types remains a challenge for most loci.
To address this challenge, we employed three new approaches. First, we analyzed cell type specific brain eQTL data [2], which included 424 samples across 7 cell types. Second, we utilized SharePro [3], a novel co-localization method that delivers more powerful results through simultaneous fine-mapping of traits being compared. Third, we conducted cell type specific SNP-to-gene mapping using SCENT [4] applied to SEA-AD [5], involving 28 samples across 18 brain cell types using 10x multiome data.
For the 90 PD associated loci [1], the widely used, coloc.susie [6] identified 11 loci altering the expression of 16 unique genes, whereas SharePro identified 30 loci with 43 unique genes. Within these 43 genes, signals were distributed across seven cell types: 10 in astrocyte (AST), 47 in excitatory neuron (EXC), 19 in inhibitory neuron (INH), 4 in microglia (MIC), 9 in oligodendrocyte (OLI), and 3 in Oligodendrocyte precursor cell (OPC). Of all co-localization signals identified from SharePro (total of 92 signals, across all cell types), 50 were up-regulations and 42 were down-regulations. Additionally, signals for 12 of the 43 genes were replicated in at least two different cell types. Examples of our findings include: *PGS1* downregulated in OLI, EXC, and INH; *LRRK2* upregulated in MIC and INH; and *GPNMB* downregulated in MIC and OPCs.
Using SCENT, we identified 2,977 significant gene-peak pairs across cell types, within a 1Mb region around the transcription start sites (TSS) of 43 genes identified by SharePro. Among these signals, 17 pairs showed overlap with credible sets identified by SharePro across all cell types. For 14 of the 17 pairs, the cell types identified by SharePro and SCENT were concordant.
In addition, we performed fine-mapping using SuSiE for the 90 PD loci [1] to derive credible sets. Among these, 12 of the 90 loci had at least 50% of their credible set variants in a significant peak-to-gene mapping identified by SCENT, enabling us to identify the likely effector gene and cell-types.
Overall, in addition to neurons, our work identifies AST, OLI, and OPC as cell types of interest in PD. Our application of novel methods (SharePro and SCENT) to large-scale cell type specific

brain single cell datasets, uncovers causal biology for a number of PD associated loci providing new mechanistic insights.

References (PMIDs):
[1] 31701892
[2] 38514782
[3] 38688586
[4] 38594305
[5] 37292694
[6] 34587156

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5102W: Integrating Multi-Omic Data to Unveil Molecular and Cellular Pathways Mediated by Common Polygenic Variation Using Deep Neural Networks

Authors:

S. Paylakhi[1], R. Wedow[2], M. Tegtmeyer[1]; [1]Purdue Univ., West Lafayette, IN, [2]Purdue Univ./Indiana Univ. Sch. of Med., Indianapolis, IN

Abstract:

The influence of genetic risk factors on biological function is pivotal to understanding the molecular and cellular mechanisms which underlie complex traits. Many studies have examined the function of individual risk genes and variants, yet most disease risk is the result of the summary effects of many hundreds or thousands of variants. To advance beyond single variant translational approaches, this study investigates how summary effects of genetic risk (via Polygenic Risk Scores (PRS)) for complex traits impacts biological function at the molecular and cellular level. Using a deep neural network (DNN), we integrate multi-omic datasets, including PRS from genome-wide association studies (GWAS), RNA-seq gene expression profiles, and high-dimensional cellular morphology profiles from 300 individuals. By leveraging feature importance analysis within our DNN framework, we pinpoint critical intersections across our diverse data modalities. This approach reveals potential cell biological pathways and processes mediated by complex trait risk. Our findings demonstrate complex, nonlinear relationships between PRS and cellular function, providing insights into molecular mechanisms driven by genetic risk factors. The robustness and generalizability of the DNN model are validated through independent datasets, underscoring its effectiveness in elucidating these relationships. This study highlights the utility of DNN modeling in exploring the biological impact of genetic risk scores, advancing

our understanding of the molecular and cellular underpinnings of complex traits. The insights gained pave the way for hypothesis generation and experimental validation, contributing to precision medicine and targeted therapeutic development.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5103W: Integrative Analysis of Allele Specific Expression and miRNA Profiling Reveals Underlying Disease Mechanism in Age-Related Macular Degeneration ⭐

Authors:

T. Shwani[1], J. Slone[1], L. A. Owen[2], I. K. Kim[3], A. L. Reynolds[1], J. H. Lillvis[1], D. Stambolian[4], L. A. Farrer[5], M. M. Deangelis[1]; [1]Univ. at Buffalo, Buffalo, NY, [2]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [3]Mass Eye & Ear Infirmary, Boston, MA, [4]Univ. of Pennsylvania, Monroeville, NJ, [5]Boston Univ Sch Med, Boston, MA

Abstract:

Age-related macular degeneration (AMD) is a complex, progressive neurodegenerative disease leading to irreversible vision loss. It's comprised of early, intermediate and late stages: geographic atrophy (dry form) and neovascular (wet form). Currently, AMD affects nearly 20 million people in the United States and is projected to impact 288 million globally by 2040. Dry AMD comprises 85-90% of these cases. However, recent FDA-approved (C3 and C5) inhibitors for geographic atrophy (GA) are not curative and do not reverse blindness, underscoring the need for novel biomarkers and therapeutics. Previous studies and our recent paper suggest allele-specific expression (ASE) and miRNA regulation may underlie AMD disease pathophysiology. ASE analysis offers insights into gene regulation and expression imbalances, while miRNAs, can provide critical information on post-transcriptional regulation. Both ASE and miRNAs have been implicated in other neurodegenerative diseases, such as Alzheimer's disease. Therefore, integrating miRNA profiling with ASE analysis can provide a comprehensive understanding of AMD pathophysiology.To explore this, we dissected the macular region of the RPE/choroid and, separately, the neural retina, as this is the disease-affected tissue. Donor eyes were deeply phenotyped (<6 hours postmortem) and obtained from 14 patients over 70 years of age, including 6 with intermediate AMD (iAMD) and 8 age-matched controls. Our novel, agnostic genetics workflow merged whole-exome sequencing, SNP genotyping (Illumina's HumanOmni2.5-8 BeadChip v1.2), and RNA sequencing data to construct personalized genomic alignments. This enabled non-paired differential ASE analysis using a custom DESeq2 design. Gene prioritization scores, incorporating imprinted genes, eQTL, and GTEx data, guided ASE

candidate selection. Ingenuity Pathway Analysis elucidated associated pathways.Prioritized genes were EXOC1, FMN1, TF (from the macular retina) and CAT, FPR1, GAA, SPARCL1 (from the macular RPE/choroid). Drop-seq PCR was used to validate prioritized genes. Validation was only seen in the macular RPE/choroid and these validated ASE genes were compared to our sequenced miRNA dataset (n=60 eyes). We identified significant differentially expressed miRNAs targeting FPR1 (8 targeting miRNAs) and SPARCL1 (1 miRNA).This study is the first to report non-paired differential ASE in well-phenotyped AMD-affected tissues, providing critical insights into tissue-specific changes and regulatory mechanisms in iAMD. Integrated with miRNA analysis, it offers potential avenues for targeted therapies and interventions at an earlier stage of disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5104W: Interaction effects of genetic variants and white matter hyperintensities on Alzheimer's disease biomarkers

Authors:

Y. Wong[1,2,3], C-Y. Wu[1,2,3], D. K. Mori-Fegan[1,2,3], S. Noor[1,2,3], L. Y. Xiong[1,2,3], S. S. Mirza[1,2,3], M. Masellis[1,2,3,4,5], E. Rogaeva[1,4], Y. Amemiya[2,3], A. Seth[1,2,3], J. Ramirez[1,2,3], C. J. Scott[1,2,3,5], F. Gao[2,3,5], S. Symons[1,2,3], K. Zukotynski[2,3,6,7], A. Bhan[1,2,3], R. H. Swartz[1,2,3,8], M. Goubran[1,2,3], J. Keith[1,2,3], S. E. Black[1,2,3,5,9], M. J. Chenoweth[1,10], W. Swardfager[1,2,3,11]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Sunnybrook Res. Inst., Toronto, ON, Canada, [3]Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [4]Tanz Ctr. for Res. in Neurodegenerative Disease, Toronto, ON, Canada, [5]Heart and Stroke Fndn. Canadian Partnership for Stroke Recovery, Toronto, ON, Canada, [6]McMaster Univ., Hamilton, ON, Canada, [7]Univ. of British Columbia, Vancouver, BC, Canada, [8]ICES, Toronto, ON, Canada, [9]Toronto Dementia Res. Alliance, Toronto, ON, Canada, [10]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [11]KITE UHN Toronto Rehabilation Inst., Toronto, ON, Canada

Abstract:

Alzheimer's disease (AD) is the most common cause of dementia, characterized by beta-amyloid and tau protein deposition, neurodegeneration, and progressive cognitive decline. White matter hyperintensities (WMH) are confluent patches in the brain's periventricular white matter that appear as hyperintense regions on MRI and are associated with cognitive and motor decline. WMH originate from venous collagenosis in the brain and are commonly observed in AD patients. However, the relationship between WMH and AD remains unclear. We aim to identify genetic variants that may modify the relationship between WMH and AD biomarkers.
We included individuals with AD, mild cognitive impairment, and normal cognition from the Alzheimer's Disease Neuroimaging Initiative (ADNI, n=1012). WMH were measured from FLAIR

MRI scans. CSF-Aβ42 concentrations were measured using immunoassays. Hippocampal volumes (HP) were obtained using FreeSurfer. Genome-wide genotype data underwent standard quality control, and individuals of European ancestry confirmed via principal component (PC) analysis were selected. Genetic imputation was conducted on the TOPMed Imputation server. Interaction effects between genetic variants and WMH on AD biomarkers were assessed using linear regression models, adjusting for age, sex, diagnosis, MMSE, APOE-ε4 status, head size, and four ancestral PCs in PLINK2. Linkage disequilibrium-based clumping was performed. Significant variant-WMH interactions discovered in ADNI were validated in participants from the Sunnybrook Dementia Study (SDS, n=433) and UK Biobank (UKB, n=16720).
A 22-variant intergenic locus on chromosome 18 (top hit variant: rs72899960 T>A, β=0.19 pg/ml (logged), SE=0.03, p=6.28x10-9, MAF=11.1%, Imputation R2>0.99%, n=848) significantly interacted with WMH to predict Aβ42, with the nearest gene being a novel long noncoding RNA (lncRNA), ENSG00000286844 (-360KB). A 9-variant intronic locus on chromosome 11 in the lncRNA MIR4300HG gene (top hit variant: rs3912008 C>T, β=0.25 cm3, SE=0.044, p=1.47x10-8, MAF=22.6%, Imputation R2>99%, n=987) significantly interacted with WMH to predict HP. Both variants additively moderated the relationships between AD markers and WMH, which were validated in independent cohorts for WMH-plasma-Aβ42 relationship in UKB and WMH-hippocampal volume in SDS. Potential regulatory effects of the identified variants on gene and protein expressions are now being investigated in the Religious Orders Study/Memory and Aging Project (n=1709).
This interaction study may help identify the contributions of specific genes and pathways in the relationships between WMH and AD biomarkers.

---PAGEBREAK---

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5105W: International meta-analysis of genome-wide association study on adolescent idiopathic scoliosis

Authors:

T. Iwami[1,2], N. Otomo[3,2,1], Y. Ming[4], Y. Yonezawa[2], K. Takeda[2], J. P. Cheung[5], J. Rios[3,6,7], S. Ikegawa[1], M. Matsumoto[2], M. Nakamura[2], U. Thorsteinsdottir[8,9], Y. Song[4], C. Wise[3,6,7], K. Watanabe[2], C. Terao[1]; [1]Lab. for Statistical Analysis, Ctr. for Integrative Med. Sci., RIKEN, Yokohama, Japan, [2]Dept. of Orthopedic Surgery, Keio Univ. Sch. of Med., Tokyo, Japan, [3]Ctr. for Translational Res., Scottish Rite for Children, Dallas, TX, [4]Sch. of BioMed. Sci., The Univ. of Hong Kong, Hong Kong, China, [5]Dept. of Orthopaedics and Traumatology LKS Faculty of Med., The

Univ. of Hong Kong, Hong Kong SAR, China, [6]Dept. of Orthopaedic Surgery and Pediatric, Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [7]McDermott Ctr. for Human Growth and Dev. Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [8]deCODE genetics, Reykjavik, Iceland, [9]Faculty of Med., Sch. of Hlth.Sci., Univ. of Iceland, Reykjavik, Iceland

Abstract:

Background: We have conducted genome-wide association studies (GWASs) for adolescent idiopathic scoliosis (AIS) and identified 20 susceptibility loci; however, our understanding of genetic background of AIS remains insufficient. Therefore, we conducted an international meta-analysis of AIS GWAS to further elucidate susceptibility loci, and provide a more detailed understanding of the genetic background of the disease. **Methods:**First, we conducted a meta-analysis of our three Japanese AIS GWASs, including 5,427 cases and 155,983 controls. We then performed a meta-analysis with GWASs from cohorts in Hong Kong (977 cases and 2,085 controls), the US (1,503 cases and 20,153 controls), and Iceland (146 cases and 255,191 controls). With these results, we performed an analysis of the genetic correlation between East Asians and Europeans and inferred susceptible tissues and cell types through a heritability enrichment analysis using publicly available epigenetic databases. Additionally, we developed a prediction model for AIS using polygenic risk scores (PRSs) based on the clumping and thresholding method. **Results:**We identified 22 novel susceptibility loci, some of which were associated with genes highly expressed in musculoskeletal tissues such as cartilage, muscle, and nerve. Also, we observed a strong genetic correlation (ρ:0.87) between East Asians and Europeans. The genetic background of AIS was strongly enriched in the regulatory regions of rhabdomyosarcoma (7.1×), chondrocytes (9.5×), and myoblasts (7.4×) from public databases. Additionally, by combining the several regions with strong associations, 72% of the AIS genetic background could be explained by 9.4% of the entire genome. In a PRS analysis, we confirmed that the predictability of the model highly surpassed that of the previous models. We demonstrated that as more populations were included, the AUC, a measure of predictability, improved to 0.708 (Japanese), 0.718 (East Asians), and 0.723 (all populations). Moreover, a model that prioritized SNPs within the strongly enriched regions mentioned above exhibited the highest predictability (AUC:0.733) among all the models. **Conclusion:**We conducted the international meta-analysis of GWAS using more than 500,000 subjects and identified 22 novel AIS susceptibility loci. The genetic background of AIS was shared across different ancestries, 72% of which could be explained by the transcription regulatory regions of specific cell types. The prediction model that implemented PRS based on global populations with these regions showed the highest predictability, potentially becoming an effective tool in clinical settings.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5106W: Investigating CYP26B1 as novel obesity locus in indigenous Americans living in Arizona

Authors:

F. Khan, K. Freeland, R. L. Hanson, Y. L. Muller, L. J. Baier; NIDDK, NIH, Phoenix, AZ

Abstract:

To identify genetic factors that contribute to ethnic disparity in rates of obesity, we performed an association analysis between 7.9 M imputed variants in 6789 longitudinally studied indigenous individuals living in Arizona and their maximum BMI recorded during any exam at age ≥15 years. After adjusting for age, sex, birth year, and the first 5 genetic principal components, the strongest associations were rs114857151, rs187498396, and rs189879565 (r^2≥0.99; 2.7% increase in BMI per risk allele, P=4 x 10^{-7}) in *CYP26B1*, an unreported locus for obesity. The obesity risk alleles for rs114857151, rs187498396, and rs189879565 are highly enriched in the indigenous population (MAFs=0.22) as compared to the global gnomAD dataset (MAFs=0.028, 0.0032, and 0.0033, respectively). CYP26B1 is a member of the cytochrome P450 family and is involved in the catabolism of retinoic acid (RA), the active metabolite of vitamin A. Inhibition of CYP26B1 results in increased retinoid signaling, while overexpression results in RA clearance leading to decreased signaling. It has been suggested that decreased RA may influence the development of obesity. Most of biological activities of RA are mediated by retinoic acid receptors (RARs) including inhibition of adipogenesis and boosting of lipid oxidation.Expression of CYP26B1 is limited to a few tissues including adipose and brain. An expression quantitative trait loci (eQTL) analysis in adipose tissues biopsies (N=195) from indigenous participants showed that the obesity risk alleles are modestly associated with increased CYP26B1 expression (P=0.018). The associated variants likely affect either transcript 1 or 3 of *CYP26B1* and we are currently examining clones for each transcript in 3T3-L1 cells. To date in vitro functional studies of obesity risk haplotype have shown increased luciferase activity for transcript 3 promotor (P = 0.027) similar to what was observed for total CYP26B1 expression in the adipose biopsies. In differential expression experiments in OP9 cells we observed downregulation of *ADIPQ1* (41%), *ACADM* (34%) and *SCD1*(38%) in response to over expression of *CYP26B1* while knock down showed the opposite effect. In summary, we identified variants in *CYP26B1* that are highly enriched in indigenous population of Arizona and associated with BMI likely via a mechanism of decreased retinoid signaling. Our preliminary data suggests these SNPs cause differential expression of transcript 3 which affects expression of other adipocyte genes with known roles in fatty acid oxidation and lipogenesis. Our future studies involve further assessment of these variants on adipogenesis in the presence/absence of all-trans-retinoic acid.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5108W: Joint Fluid Single Cell Transcriptomics in Pediatric Septic Arthritis Discovers Signatures of Autoimmunity in Culture Negative Patients

Authors:

P. Raj[1], B. Zhang[1], Y. P. Lopez[1], C. Arana[1], N. Tareen[2], I. Raman[1], Y. Liu[1], P. Selvakumar[1], N. vanoers[1], S. morris[1], L. Hooper[1], L. Copley[1]; [1]UT Southwestern Med Ctr, Dallas, TX, [2]Children's Dallas, Dallas, TX

Abstract:

Objective: To demonstrate the feasibility and potential clinical value of multi-omics approach to guide clinical decision-making for children suspected to have presumed septic arthritis (SA) and acute hematogenous osteomyelitis (AHO). **Study design:** Twenty Children with AHO and SA diagnosis were longitudinally studied for changes in Knee Joint-fluid, blood, oral, gut microbiome profile. All standard of care culture and polymerase chain reaction (PCR) testing of joint fluid samples was performed. A multi-omics approach that combines metagenomics, cytokines, autoantibodies and single cell transcriptomics signatures was applied to enhance the diagnostic confidence for children with both culture- and PCR-negative diagnosis. **Results:** Metagenomics sequencing confirmed the presence of *Streptococcus pyogenes* and *Staphylococcus aureus* species in 50% samples, whereas remaining 50% Joint fluid cultures were negative for any pathogen. The clinical phenotypes i.e. Neutrophils count and CRP level were similarly elevated in both, infection negative and positive group. Many inflammatory cytokines including IL-4, IL-17A, MCP-1, IL-8, IL-2, IL-1b, IL-6, TNF-a, etc. were significantly elevated in the joint-fluid of children with bacteremia as compared to the infection negative subset. Interestingly, 50% children without any bacteremia exhibited strong signature of autoantibodies (IgGs) targeting several nuclear i.e. *dsDNA, histones, Jo-1, scl-70, Ro/SS-A, SmDs, CENP-A* and non-nuclear antigens i.e. *Collagens, Myosin, Laminin*. Furthermore, single cell transcriptomics in autoantibody positive subject confirmed presence of autoantibody producing B cells in this joint-fluid. The results from all three assays, including metagenomics (p=0.006), cytokine profile (p=0.009) and autoantibody profile (p=0.04) significantly differentiated between children with and without infection. Besides confirming the standard-of care positive results, the multi-omics approach guided the diagnosis of children with inflammatory arthritis, Juvenile idiopathic arthritis (JIA) and Post-Streptococcal Reactive Arthritis. Overall, this improved the diagnostic confidence for children with suspected septic arthritis and AHO especially in the

infection negative subset.**Conclusions:** This approach rapidly stratifies children with bacterial infections and inflammatory arthritis and unravels potential mechanisms of pathology in culture negative subset.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5109W: Joint modeling across tissues via tensor decomposition implicates GCM2 and ZNF73 regulons in genetic risk of Alzheimer's disease

Authors:

Y. Yan[1], R. Chen[1], H. Kang[1], Y. Tan[1], A. Tiwari[1], Z. Wen[2], X. Zhong[3], B. Li[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Emory Univ., Atlanta, GA, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Alzheimer's disease (AD) is the most prevalent form of dementia and the leading neurodegenerative disorder. However, our understanding of its etiology remains incomplete. Gene regulatory programs play a vital role in disease onset and progression. Regulon, defined as a network of a transcription factor (TF) and its regulated target genes, represents the core of the regulatory program. Mechanisms implicated in AD, such as inflammation and mitochondrial dysfunction, often result from the activation or dysfunction of regulons. Identifying regulons implicated in AD genetic risk is key to understanding the regulatory programs underlying AD pathogenesis and identifying potential therapeutic targets. In this study, we utilized GTEx expression data from multiple tissues and employed SCENIC, a computational method for gene regulatory network reconstruction, to infer regulons. In total, we obtained 651 regulons across 48 tissues in GTEx. We decomposed the regulon by tissue data via tensor decomposition and performed association analyses of the latent components with AD GWAS-identified genetic variants using sparse canonical correlation analysis (sCCA). Our analysis revealed that two regulons, involving TFs of *GCM2* and *ZNF732*, are significantly associated with AD genetic variants. Among them, *ZCWPW1* within the *GCM2* regulon is a known AD-risk gene, while *ZNF723* is located in the *cis*-region of an AD risk variant (rs3822020). We further investigated the associated regulons via expression quantitative trait loci (eQTL) analysis, and specifically we tested whether the AD risk variants act as *trans*-eQTL of the regulons via regulating the corresponding TFs *in cis*. While the eQTL analysis of the TFs did not yield significant results, a closer examination of genes in these regulons revealed that AD risk variant rs7181268 is a *trans*-eQTL of *SPINK2* in the *GCM2* regulon and rs199520 is a *cis*-eQTL of *LRRC37A2* in the *ZNF732* regulons. In summary, in this study, we identified two regulons

significantly associated with AD genetic risk and further revealed that the involvement of the associated regulons in AD genetics may be mediated through *cis*- or *trans*-eQTLs.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5110W: Large-scale analysis of sequencing data uncovers association between ultra-rare protein-truncating variants in *ITSN1* and Parkinson's disease ⭐

Authors:

T. Spargo[1], C. Sands-Falconer[2], V. Ravanmehr[2], J. Butts[2], R. B. De-Paula[2], B. Hollis[1], J. Mitchell[1], F. Hu[1], Q. Wang[3], J. Shulman[2], M. Messa[4], G. del Angel[5], K. Tan[1], S. Petrovski[1], I. Al-Ramahi[2], I. Tachmazidou[1], R. S. Dhindsa[2]; [1]AstraZeneca, Cambridge, United Kingdom, [2]Baylor Coll. of Med., Houston, TX, [3]AstraZeneca, Gaithersburg, MD, [4]AstraZeneca, Gothenburg, Sweden, [5]AstraZeneca, Boston, MA

Abstract:

Parkinson's disease (PD) is a progressive neurodegenerative disorder that typically emerges in mid-late adulthood and is characterised by tremor, rigidity, slow movement, and difficulty balancing. Both common and rare genetic variation are implicated in the disease, with effects ranging between monogenic and complex. However, known variants only explain the occurrence of PD for a fraction of people affected. With greater availability of large-scale Biobank cohorts with sequencing data, a genome-wide investigation of rare-variant associations to PD has recently become feasible. This could provide vital insights into disease-implicated mechanisms and valuably inform development of treatments for PD, which are currently limited to symptom management.

We performed a large sequencing-based study of gene- and variant-level associations to PD susceptibility in the UK Biobank, focussing upon a European-ancestry subset and then in a pan-ancestry analysis. Our analysis replicated known associations to variation in genes such as *GBA*, *MAPT*, and *LRRK2*. The gene-level analysis of Europeans revealed a putative association to ultra-rare protein truncating variants in the *ITSN1* gene (Fisher's Exact Test: $p=6.05\times10^{-7}$), with an effect size greater than genes/variants previously implicated in PD (Odds ratio [95% confidence interval] = 10.53 [5.2, 21.34]).

Follow-up investigations substantiated the discovery: (1) The signal replicated in an analysis of independent data from AMP-PD and All of Us ($p=5.41\times10^{-3}$); (2) Functional analysis showed decreased foot speed and more frequent stumbling in drosophila after knockout of *DAP160*, the fly homolog of *ITSN1*; (3) further experiments in flies expressing human alpha-synuclein showed

an interaction between expression of *DAP160* and alpha-synuclein-induced eye degeneration. Since *ITSN1* has received little focus in human biological studies its role is not fully known. The ITSN1 protein is thought to be involved in synaptic vesicle recycling, interacting directly with SYNJ1 which is implicated in PD. Moreover, it is highly expressed in the nervous system and previous studies have shown altered motor phenotypes in *ITSN1* knockout mice. Accordingly, various lines of evidence suggest that loss of function in *ITSN1* may lead to a neurodegenerative motor phenotype such as PD. Our analyses of PD risk demonstrate that the impact of this effect may be substantial. Since *ITSN1* has been linked to autism spectrum disorder previously, variable expressivity and penetrance may play a role. Further research into *ITSN1* and its associated biological pathways may yield new insights into the aetiology of PD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5112W: Large-scale Tourette Syndrome whole-exome sequencing analysis reveals a significant contribution of de novo mutations different from Autism Spectrum Disorder.

Authors:

L. Zhan; UCLA, Los Angeles, CA

Abstract:

Tourette syndrome (TS) is an early-onset neurodevelopmental disorder (NDD) characterized by vocal and motor tics. TS is highly heritable (60-80%) and has a complex genetic architecture with both rare and common variants contributing to the genetic etiology. However, the contribution of de novo mutations (DNMs) has not been widely studied due to the lack of large-scale high-quality family-structured sequencing datasets and insufficient statistical power. In this study, we generated whole-exome sequencing (WES) data for over 1,300 TS trio families and jointly called the data with a selected subset of over 6,600 families from SSC and SPARK datasets containing at least one Autism Spectrum Disorder (ASD) proband and an unaffected sibling. This produced a high-quality unified dataset with 30,803 individuals and 4,425,974 variants. Our principal component analysis showed a diverse ancestry background with all major population groups present in the ASD cohorts, while only European and Admixed American ancestries were present in the TS cohort. Analyzing samples from all three cohorts demonstrated a significant exome-wide enrichment of DNMs, in the category of protein-truncating variants (PTVs) and damaging missense variants (Dmis), in TS probands compared to healthy controls, while confirming a significant enrichment of DNMs in ASD probands compared to healthy controls. The enrichment was further increased when we focused on genes with high pLI scores,

indicating an impact of DNMs on the core body functionalities. To compare the genetic architecture of TS and ASD, we obtained genesets known to be associated with ASD and DD and confirmed a significant enrichment of DNMs in ASD probands compared to healthy controls, while TS probands only demonstrated a trend of enrichment in DD genes. Interestingly, when we looked at non-ASD genes, both TS probands and ASD probands from the SPARK cohort demonstrated significant enrichments of DNMs, suggesting a different genetic architecture between TS and ASD, while also recapitulating the potential phenotypic heterogeneity within the SPARK cohort. Lastly, to identify TS risk genes, we performed transmitted-and-de-novo-analysis (TADA) and identified three significant genes with FDR < 0.05. In brief, our study includes one of the largest TS-ASD WES family datasets that will facilitate future genetic and clinical studies on TS and comorbidities with improved power. Moreover, our work provides evidence of the contribution of de novo mutations to TS etiology, shows a different genetic architecture between TS and ASD, and pinpointed three TS risk genes.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5113W: Leveraging trisomy 21's unique characteristics to unveil the genetic basis of common complex diseases

Authors:

M. Subirana Granés[1], C. Greene[1], J. Costello[2], M. Pividori[1]; [1]Dept. of BioMed. Informatics, Univ. of Colorado Sch. of Med., Aurora, CO, [2]Dept. of Pharmacology, Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Significant effort has been dedicated to uncovering the genetic determinants of complex traits, yet single-variant/gene methods have often fallen short in capturing the intricate regulatory networks influencing phenotypes. This limitation is especially relevant for individuals with trisomy 21 (T21), the genetic cause of Down syndrome (DS), who experience a range of co-occurring conditions such as Alzheimer's disease, autoimmune disorders, and various neurological issues. The additional chromosome 21 introduces a unique layer of complexity, presenting an opportunity to understand both the mechanisms underlying these co-occurring conditions in T21 and the molecular basis of common complex traits in the general population. Based on the omnigenic model of complex traits, we hypothesize that T21-induced gene expression creates unique, variable, and cascading effects in gene regulatory networks, contributing to DS co-occurring conditions through modified mechanisms compared to the disomic population (D21). To explore these interactions, we projected expression profiles from

DS individuals in the Human Trisome Project (HTP, n=400) onto machine learning-derived gene expression modules (MLDMs), constructed using a low-dimensional representation guided by prior biological knowledge to align with pathway activities. We also integrated gene-trait associations from publicly available genome-wide and transcriptome-wide association studies (GWAS and TWAS) for co-occurring conditions in T21 and D21, enabling a comprehensive joint analysis.
By projecting HTP expression data into this low-dimensional space and contrasting groups (T21 vs. D21), we identified 24 DS differentially expressed MLDMs (FDR < 0.05). Among these, nine MLDMs were significantly enriched with chromosome 21 genes (FDR < 0.05), suggesting regulatory effects from this chromosome. These MLDMs were linked to pathways implicated in DS, such as immune cell states, cytokines, and inflammation (FDR < 0.05), and were associated with DS-specific phenotypes including autoimmune diseases (FDR < 0.05), type 1 diabetes (FDR < 1e-5), and Alzheimer's disease (FDR < 1e-3).
Our MLDM-based approach identified differentially expressed MLDMs that were associated with co-occurring conditions in DS when integrated with GWAS/TWAS, providing new insights into DS-related molecular processes. These findings have potential implications for the prognosis, treatment, and diagnosis of DS, and may also extend to the general population, enhancing our understanding of the molecular basis of common complex traits.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5115W: Mapping AD Progression: Multi-region spatial transcriptomics to study susceptibility to amyloid beta (Aβ) plaque in Alzheimer's Disease (AD)

Authors:

O. Bayaraa[1], E. DeBose-Scarlett[2], E. Hocke[3], V. Jain[3], D. A. Cruz[4], S-H. J. Wang[5], S. G. Gregory[3,6]; [1]Duke Univ. Program in Genetics and Genomics, Durham, NC, [2]Duke Dept. of Molecular Genetics and Microbiol., Durham, NC, [3]Duke Molecular Physiology Inst., Durham, NC, [4]Duke Dept. of Psychiatry and Behavioral Sci., Durham, NC, [5]Duke Dept. of Pathology, Durham, NC, [6]Duke Dept. of Neurosurgery, Durham, NC

Abstract:

Alzheimer’s disease (AD) is the leading cause of dementia affecting 55 million people worldwide. The pathological hallmarks of AD, beta-amyloid (Aβ) plaques and neurofibrillary tangles (NFT), follow distinct stereotypical patterns of progression and trigger a multicellular response that ultimately leads to neuronal loss and cognitive decline. However, there is a gap in

our understanding of the molecular mechanisms that underly the differential susceptibility of specific brain regions and their cell types to AD pathology. We hypothesized that brain region and cell type specific transcriptional responses within Aβ microenvironments and more broadly within the grey matter lead to variation in AD susceptibility. To test this, we used spatial transcriptomic profiling to characterize gene expression changes during disease progression. We uniquely generated matched data from four brain regions that progressively develop plaques and tangles at different stages from two individuals with AD. By integrating spatial transcriptomic and Aβ immunofluorescence staining data from adjacent sections from the entorhinal, occipito-temporal, dorsolateral prefrontal and striate cortices, we were able to identify brain region and cell type specific transcriptional changes that potentially contribute to AD susceptibility.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5116W: Mechanistic insights from liver fat GWAS loci

Authors:

H. Somineni, S. Mukherjee, S. Satapati, D. Lloyd, C. O'Dushlaine; Insitro, South San Francisco, CA

Abstract:

Metabolic dysfunction-associated steatotic liver disease (MASLD) is a complex condition with diverse underlying mechanisms. Our recent genome-wide association studies (GWAS), leveraging machine learning-derived liver fat estimates, identified 321 loci, significantly expanding the genetic landscape of MASLD. In this study, we aimed to gain mechanistic insights into MASLD by stratifying these 321 liver fat-associated index variants based on their patterns of association with 37 relevant phenotypes and traits. We applied truncated singular value decomposition (tSVD) to systematically identify key components of genetic associations, followed by characterizing the contributions of variants and phenotypes to each component. This approach allowed us to decompose liver fat loci into various mechanistic components characterized by lipid metabolism, overall body composition, lipid and metabolic markers, liver function markers, blood cell composition, and inflammation. On the other hand, while variants in APOE were loaded onto the component characterized by lipid metabolism, CRP, HNF1A, APOE, and PDE4B - all known to regulate inflammation - were loaded onto the inflammation component. These findings suggest that stratifying GWAS signals into biologically meaningful genetic clusters based on their association patterns with relevant traits can facilitate mechanistic insight-based prioritization of targets for drug discovery.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5117W: Medical phenome-scale evaluation of the Polygenic Score Catalog in the BioMe Biobank in New York City

Authors:

K. Ferar, T. J. Mize, S. Suckiel, M. F. Murray, E. E. Kenny; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Background: Polygenic scores (PGS) derived from genome-wide association studies (GWAS) are increasingly used in biomedical research to predict susceptibility to common traits and diseases. Despite this trend, few studies have calculated PGS for a broad spectrum of traits and tested associations with clinical phenotypes from the electronic health record (EHR), especially in non-European cohorts. Phenome-wide association studies (PheWAS) can be used to examine ancestry-specific associations between PGS and a range of clinical phenotypes. Here, we perform ancestry-specific PheWAS for 3640 scores from the PGS Catalog to identify predictors of phenotypes in the EHR. Methods: We analyzed PGS and EHR phenotype associations using Mount Sinai's diverse BioMe Biobank, with 31705 individuals clustered into five ancestry groups: European (EUR; n=10294), African (AFR; n=9774), Hispanic/Latino (HIS; n=6420), South Asian (SAS; n=850), and East Asian (EAS; n=780). ICD-10 codes were mapped to 1856 phecodes using the PheWAS R package. PGS were calculated using TOPMed-imputed Regeneron GSA genotypes and PGS Catalog scoring files. For PGS with >90% SNP overlap between GWAS and available genotype data, logistic regression with Firth's bias correction was used to identify PGS-phecode associations within ancestry groups. Results: A total of 628 associations reached phenome-wide significance ($p<6.4e-6$, correcting for 156500 tests) across the five cohorts, 479 (76%) of which were identified in EUR. Among the most significant associations in EAS, PGS for cervical cancer and type 2 diabetes were associated with endocrine/metabolic and circulatory system disorders, respectively. In SAS, bladder cancer and Alzheimer's disease were associated with endocrine/metabolic disorders. In HIS, PGS for asthma and bladder cancer were associated with respiratory disorders. In AFR, PGS for mineral metabolism disorders and late-onset Alzheimer's were linked to hematopoietic and endocrine/metabolic disorders, respectively. In EUR, PGS for skin carcinoma and coagulation defects were associated with various cancers. Discussion: We identified many statistically significant associations between PGS and clinical phenotypes in the BioMe Biobank, most of which were unique to ancestry. In non-European cohorts, significant PGS were largely from European-only GWAS, and associations often lacked clinical plausibility (e.g., Alzheimer's and endocrine/metabolic disorders). To further examine

these associations, we will conduct PheWAS at additional SNP overlap thresholds and assess the predictive capacity of PGS with matched phenotypes, both within and across ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5119W: Mendelian Randomization reveals causal links between circulating metabolites and Type 1 Diabetes. ⭐

Authors:

B. Jumentier[1], D. A. Michalek[2], W-M. Chen[2], S. Onengut-Gumuscu[2], C. Polychronakos[3], S. Rich[2], D. Manousaki[1]; [1]Sainte Justine Hosp. Univ. of Montreal, Montreal, QC, Canada, [2]Univ. of Virginia, Charlottesville, VA, [3]McGill Univ. Hlth.Ctr., Montreal, QC, Canada

Abstract:

BACKGROUND Type 1 diabetes (T1D) is an autoimmune disease that causes destruction of the insulin-producing pancreatic beta cell, leading to profound alterations in carbohydrate, lipid, and protein metabolism if untreated. But if early metabolic alterations are involved in the pathogenesis of T1D remains to be elucidated. Here, we sought to detect if genetically predicted levels of circulating metabolites in the blood are causally linked to risk of T1D, using two-sample Mendelian Randomization (MR).

METHODS AND RESULTS For our MR analyses, we used SNP-instruments for circulating metabolites measured in seven GWAS (Kettunen et, Long et al, Lotta et al, Shin et al, Surendran et al, Chen et al, Schlosser et al). The corresponding effects of the SNP-instruments on T1D were derived from 5 GWAS (the European GWAS by Chiou et al, and two multi-ethnic GWAS by Michalek et al, and Sakaue et al). The MR effects were estimated using the Wald Ratio for metabolites with a single instrument, and the Inverse Weighted Variance (IVW) method for metabolites with multiple SNP-instruments. Sensitivity analyses for pleiotropy were performed using the Weighted median, Weighted mode and MR Egger methods. In order to confirm the direction of causality in our significant MR associations, we applied reverse MR.For selected candidate metabolites, we tested the presence of shared causal variants between these metabolites and T1D using genetic colocalization and explored their shared genetic architecture using LD score regression (LDSC).

Of the 1679 tested metabolite-T1D associations, 46 metabolites showed a Bonferroni-adjusted significant effect ($P < 2.97*10^{-5}$) on T1D. Within this set, 14 metabolites replicated ($P < 9*10^{-4}$) using independent T1D GWAS. Colocalization analyses identified 6 of these metabolites sharing a causal variant with T1D while 2 among these metabolites presented significant genetic correlation with T1D. The metabolite with the most important MR effect, vanillactate, showed a

strong negative association with T1D risk in 3 T1D GWAS and is an amino acid previously associated with T2D. The following metabolites also showed a protective effect on T1D: CMPF, myristoylcarnitine and N-acetyl-3-methylhistidine*. Finally, 1,6-anhydroglucose, a had a positive MR effect on T1D risk, and has been linked with T1D in a previous study.
CONCLUSION Combining evidence from MR and various follow-up analyses, we identified causal circulating metabolites for T1D, highlighting the role of specific metabolites, such as the amino acid vanillactate, in the pathogenesis of this disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5120W: Meta-analysis of European and East Asian GWAS Identifies Novel Susceptibility Loci in Females with Adolescent Idiopathic Scoliosis

Authors:

N. Otomo[1,2,3], I. Takuro[2,4], C. Tian[5,6], K. Anas[1,7], E. Elisabet[8], N. Masaya[2], M. Morio[2], I. Shiro[3], R. Jonathan[1,9,10], W. Kota[2], G. Paul[5,6], T. Chikashi[4], W. Carol[1,9,10]; [1]Scottish Rite for Children Ctr. for Translational Res. at UTSW, Dallas, TX, [2]Dept. of Orthopaedic Surgery, Keio Univ. Sch. of Med., Tokyo, Japan, [3]Lab. for Statistical and Translational Genetics, Ctr. for Integrative Med. Sci., RIKEN,, Yokohama, Japan, [4]Lab. for Statistical and Translational Genetics, Ctr. for Integrative Med. Sci., RIKEN,, Tokyo, Sweden, [5]Dept. of Surgical Sci., Uppsala Univ., Uppsala, Sweden, [6]Dept. of Orthopaedics and Hand surgery, Uppsala Univ. Hosp., Uppsala, Sweden, [7]Dept. of Pathology, Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [8]Sci. for Life Lab., Dept. of Gene Technology, KTH-Royal Inst. of Technology, Solna, Sweden, [9]Dept. of Orthopaedic Surgery and Pediatric, Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [10]McDermott Ctr. for Human Growth and Dev. Univ. of Texas Southwestern Med. Ctr., Dallas, TX

Abstract:

Adolescent idiopathic scoliosis (AIS) is a lateral spine deformity and the most common musculoskeletal disorder in children. AIS displays remarkable sexual dimorphism, with girls having more than five-fold risk of progressive disease compared to boys. The inheritance of AIS, which affects ~3% of the population, is polygenic. While genetic studies of AIS patients and families have identified many associated loci, most of the disease heritability and the mechanisms conveying female bias are undefined. In this study, we sought to identify novel sex-specific susceptibility loci by performing an extended genome-wide association study (GWAS) in the largest AIS cohort assembled to date. We recruited a U.S cohort of European ancestry (AIS cases: 1,157 female and 236 male; controls: 7,138 female and 5,334 male) and a Japanese

cohort of East Asian ancestry (AIS cases: 5,133 female, 339 male; controls: 83,505 female, 72,478 male). Genotyped and imputed single nucleotide polymorphisms (SNPs) were analyzed, and sex-specific GWASs were performed using a logistic regression model implemented in Plink2.0. GWAS summary statistics of both cohorts were combined by meta-analysis, and genome-wide significant loci were compared in female, male + female, and male groups. Significant loci were not identified in the male-only group likely due to power limitations. After female-only meta-analysis, we identified 17 genome-wide significant (p-value < 5×10-8) loci, including two that were novel. Of the two "female" loci, we found evidence of replication at one locus by meta-analysis with an independent Swedish cohort (1,429 AIS subjects and 3,873 controls) (rs6672559, OR 1.10, p=1.36×10-8). This signal occurs in introns of the FAT3 gene encoding FAT Atypical Cadherin 3. FAT3 is a membrane-bound receptor that modulates beta-catenin signaling in brain and cartilage growth plate. We additionally identified rare, co-segregating FAT3 nonsynonymous variants by genome sequencing in a cohort of AIS families. Finally, gene-based analysis of GWAS data identified 37 significant genes, of which 15 were not included in single SNP-based significant regions. Our data highlight new AIS candidate susceptibility genes and suggest that FAT3 mediates sexually dimorphic disease expression.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5121W: Metabolic pathways in aortic stenosis: A Mendelian randomization approach.

Authors:

T-Y. Yang[1,2,3], L. Dufresne[1], Y. Nagai[1], A. Small[4,5], G. Thanassoulis[1,6], J. Engert[1,2,6], Therapeutic targets for AoRtic stenosis using GEneTics (TARGET) Consortium; [1]Preventive and Genomic Cardiology, McGill Univ. Hlth.Ctr. and Res. Inst., Montreal, QC, Canada, [2]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [3]Kyoto-McGill Intl. Collaborative Sch. in Genomic Med., Kyoto Univ., Kyoto, Japan, [4]Div. of Cardiovascular Med., Dept. of Med., Brigham and Women s Hosp., Boston, MA, [5]Cardiovascular Res. Ctr. and Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [6]Div. of Experimental Med., McGill Univ., Montreal, QC, Canada

Abstract:

Background: Aortic stenosis (AS) is the most common valvular heart disease with a prevalence of 10%-15% among individuals over 75 years. Currently, aortic valve replacement is the only available treatment. Metabolite quantitative trait loci (mQTL) (genetic variation that influences metabolite levels) combined with Mendelian randomization (MR), offers a promising approach

to identify potential causal biomarkers for AS. **Methods:** We employed a two-sample MR approach to identify causal contributors to AS among plasma metabolites. mQTL results from the Canadian Longitudinal Study on Aging (CLSA) were used as instruments for metabolite exposures. Outcomes included AS, using GWAS results from the TARGET cohort and coronary artery disease (CAD), using a UK Biobank GWAS. **Results:** We obtained 975 mQTLs (783 single metabolites and 192 metabolite ratios for pairs that share enzymes or transporters) consisting of 1,580 independent SNPs from CLSA summary statistics. MR of these mQTLs with AS identified 29 metabolites and four metabolite ratios as significant after FDR correction. Notably, 22 (67%) of the significant mQTL associations were lipid pathway metabolites. We annotated mQTL associations based on the *FADS* locus, which is an established genome-wide significant risk factor for AS. We identified 14 significant metabolites/ratios with mQTLs located only within the *FADS* locus, eight with mQTL SNPs located within the *FADS* locus and other genes (FADS/Other), and 11 with only mQTL SNPs outside the *FADS* locus. Among the FADS/Other group results, we observed that phosphatidylethanolamine (PE, 1-(1-enyl-stearoyl)-2-linoleoyl-GPE) had a protective effect on AS (OR = 0.78, $p = 1.28 \times 10^{-3}$). Interestingly, one of the mQTLs for PE is ApoE isoform 2 (rs7412). To elucidate shared pathophysiological processes between AS and CAD, we also conducted MR using CAD GWAS results. MR of mQTLs with CAD identified sixteen significant metabolites/ratios after FDR correction. Only one result was significant in the MRs of both CAD and AS: the metabolite ratio (glucose/N-palmitoyl-sphinganine (d18:0/16:0)) (OR = 2.40, $p = 8.51 \times 10^{-22}$). **Conclusion:** Our MR of mQTLs and AS, identified 33 metabolites/metabolite ratios with a significant causal association to AS. Notably, the majority of significant mQTLs overlapped *FADS* loci, highlighting the central role that *FADS* loci play in regulating plasma levels of many metabolites that could be involved in AS etiology. We also indicate a potential novel role for PE and ApoE in AS etiology. In addition, we found a shared metabolite ratio in both AS and CAD, indicating a possible novel shared pathophysiological process.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5122W: Monogenic diabetes in Tunisia: Genetic testing reveals novel variants and an oligogenic signature enabling a new classification of diabetic patients and improved health care management ⭐

Authors:

N. Kheriji[1,2,3], H. Dallali[1], I. Gouiza[1,2,3,4], M. Hechmi[1], F. Mahjoub[2,5,6], M. Mrad[2,3,7], A. Krir[7], A. Bahlous[2,7], M. Ben Ahmed[2,8], H. Jamoussi[2,5,6], R. Kefi[1,2]; [1]Lab. of BioMed. Genomics and Oncogenetics, Inst. Pasteur de Tunis, Tunis, Tunisia, [2]Univ. of Tunis El Manar, Tunis,

Tunisia, [3]Faculty of Med. of Tunis, Tunis, Tunisia, [4]MitoLab Team, Unité MitoVasc, UMR CNRS 6015, Inst. Natl. de la santé et de la recherche médicale U1083, SFR ICAT, Univ. of Angers, Angers, France, [5]Res. Unit UR18ES01 on "Obesity", Tunis, Tunisia, [6]Natl. Inst. of Nutrition and Food Technology, Tunis, Tunisia, [7]Lab. of Clinical Biochemistry and Hormonology, Inst. Pasteur de Tunis, Tunis, Tunisia, [8]Lab. of Clinical Immunology, Inst. Pasteur de Tunis, Tunis, Tunisia

Abstract:

Monogenic Diabetes (MD) is an atypical form of diabetes, accounting for 3-6% of all cases. This prevalence is underestimated due to misdiagnosis. Genetic testing is therefore the most appropriate tool for accurate diagnosis. Research from Europe, Asia, and the US has revealed genetic, phenotypic, and clinical variability in MD, however molecular studies of MD are rare in Tunisia. The aim of this study was to unravel the genetic basis of MD among Tunisian patients using Whole Exome Sequencing (WES). We implemented exome sequencing in 24 Tunisian patients suspected with MD recruited from collaborating medical centers. The Sure Select Human All Exon V6 Kit panel was used for exome capture, followed by high-throughput sequencing via the NovaSeq6000 platform to generate 151-bp paired-end reads. The raw sequence data were meticulously processed and aligned against the human reference genome (GRCh38). Variants' pathogenicity was evaluated using several bioinformatics in-silico prediction tools as well as protein structure through advanced homology modelling and molecular dynamics. For proteins with no reference structure, the I-TASSER web server was used to generate a three-dimensional structure (3D). Also, the online ORVAL platform was used to predict the likelihood of combinations of pathogenic variations. Finally, Sanger sequencing was performed to confirm the pathogenic mutations predicted in the patients and verify familial segregation Our results revealed high clinical and genetic heterogeneity of MD in Tunisians. Among the 24 patients studied, five (20%) were diagnosed with MODY subtypes (*ABCC8, HNF1A, GCKR* and *PDX1*), 45% had syndromic forms of MD namely BBS (*BBS, TTC8*), Alström syndrome (*ALMS1*) and Wolfram syndrome (*WFS1*). The remainder had unclassified forms of MD. Thanks to our study, we were able to adjust treatment for five patients. Our study highlighted the importance of genetic screening for MD in under-represented populations. This study will help in the development of a MD-specific DNA chip that will enable rapid and more cost-effective genetic diagnosis. This molecular tool will help clinicians make the right therapeutic choice for better care of patients and their families.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5123W: Mosaic Loss of Chromosome X in Peripheral Blood Cells from Whole-Genome Sequencing and Alzheimer's Disease in the Amish

Authors:

Y. Song[1], R. Laux[1], K. Miskimen[1], A. Lynn[1], S. Fuzzell[1], S. Hochstetler[1], W. Wang[1], L. Main[1], P. Wang[1], Y. Liu[1], N. Moore[1], M. Prough[2], D. Dorfsman[2], L. Caywood[2], J. Clouse[2], S. Herington[2], A. Gulyayev[2], S. Slifer[2], L. Adams[2], P. Whitehead[2], J. Vance[2], M. Cuccaro[2], P. Ogrocki[1], A. Lerner[1], M. Pericak-Vance[2], W. Scott[2], W. Bush[1], J. Haines[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Miami, Miami, FL

Abstract:

Mosaic loss of chromosome X (mLOX) refers to acquired aneuploidy in a fraction of somatic cells. In aging women, this is the most commonly occurring clonal somatic alteration detected, yet little is known about its effect on different phenotypic consequences and the reported effects are controversial. We investigated mLOX estimated from whole genome sequencing (WGS) as a risk factor for Alzheimer's disease AD in the Midwestern Amish, a founder population with homogeneous lifestyle, reducing the effect of confounding environmental factors. The calling of mLOX was performed using the Mosaic Chromosomal Alterations (MoChA) pipeline, utilizing long-range phase information to search for imbalances between maternal and paternal allelic fraction in a cell population. A loss was called when the slope of coverage over allelic deviation was < -0.94. The cognitive status of each individual was assigned via consensus review of medical history and neuropsychological testing. The cognitively-impaired (CI) group combined AD, MCI and Unclear individuals and compared to the cognitively-unimpaired (CU) group. Individuals in 3 age groups (65-74, 75-84, >=85) were included. To evaluate the association between mLOX and risk of developing CI in follow-up, we fit a Cox proportional hazards model associating mLOX to years of disease-free follow-up within individuals for whom blood was sampled prior to CI diagnosis.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5124W: Multi-Ancestry Genome Wide Association Study (GWAS) in the *All of Us* Cohort for Primary Open Angle Glaucoma

Authors:

K. Tavakoli[1,2], B. Huang[1,2], Z. Xu[3], R. N. Weinreb[1,2], S. L. Baxter[1,2], T. Amariuta[1]; [1]Univ. of California San Diego, San Diego, CA, [2]Shiley Eye Inst., San Diego, CA, [3]Univ. of California, San Diego, La Jolla, CA

Abstract:

Primary open-angle glaucoma (POAG) is the main cause of irreversible vision impairment, affecting 58 million individuals globally. We aimed to identify genetic determinants of POAG critical to disproportionately affected minority populations by utilizing *All of Us,* a genotype/phenotype cohort matching the diversity of the US population. We performed genome-wide association studies (GWAS) across individuals of European (N=1,472 cases; 122,008 controls), African (N=813 cases; 48,015 controls), and Native American (N=322 cases; 39,431 controls) descent. In the European cohort, we identified 23 genome-wide significant (p < 5x10^-8) variants, consisting of 2 independent loci (odds ratios: 0.71-1.6), including the well-established *TMCO1* locus, which influences intraocular pressure and cellular stress response. We identified one new locus (chr6, 132.6Mb), 5kb upstream of *MOXD1,* a gene that is implicated in anemia, a risk factor for glaucoma (Firat et al. 2019 Anemia). In the African ancestry cohort, we identified 2 genome-wide significant variants (odds ratios = 1.60, 2.24), consisting of 1 independent locus, in the intron of *SLC16A7,* which acts as a molecular transporter in glial cells, which are ubiquitous in ocular tissues and may be a therapeutic target for glaucoma due to their role in neurodegeneration. Notably, this locus was not detected in the European cohort and variants demonstrated population-specific allele frequencies (minor allele frequency (MAF)<4.5e-04 in Europeans). In the Native American ancestry cohort, we identified 1 genome-wide significant variant in the intron of *SLC22A20P* (odds ratio = 2.49, MAF=0.06 in Europeans). While no GWAS loci were shared across ancestries, these differences may explain disparities in population-specific prevalence (2.5% in the European cohort; 4.2% in the African ancestry cohort; <1% in Native American ancestry) and highlight the need for personalized genetic risk assessment and treatment strategies. To understand the functional propensity of associated variants, we leveraged IMPACT functional annotations generated in retinal and other tissues. We found that the European cohort chr6 locus and the Native American cohort chr11 locus both have exceedingly high likelihoods of regulation by the KLF5 transcription factor, zinc finger protein that is expressed in the conjunctival epithelium of the anterior surface of the eye. Notably, KLF5 mutants have been associated with vision impairment. Overall, these findings stress the essentiality of leveraging diverse ancestries to better understand the population-specific nature of genetic risk for POAG.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5125W: Multi-ancestry Genome Wide Association Study Meta-analysis of Non-syndromic Orofacial Clefts

Authors:

Z. Jia[1,2], N. Mukhopadhyay[3], Z. Yang[4], A. Butali[5], B. Shi[1], J. C. Murray[6], L. Sun[7], M. L. Marazita[3,8], R. M. Freathy[2], R. N. Beaumont[2]; [1]State Key Lab. of Oral Diseases & Natl. Clinical Res. Ctr. for Oral Diseases & Dept. of cleft lip and palate, West China Hosp. of Stomatology, Sichuan Univ., Chengdu, China, [2]Dept. of Clinical and BioMed. Sci., Faculty of Hlth.and Life Sci., Univ. of Exeter, Exeter, United Kingdom, [3]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA, [4]The Sichuan Provincial Key Lab. for Human Disease Gene Study, Dept. of Clinical Lab., Sichuan Provincial People's Hosp., Sch. of Med., Univ. of Electronic Sci. and Technology of China, Chengdu, China, [5]Dept. of Oral Biology, Radiology, and Med., Univ. of Iowa, Iowa City, IA, [6]Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA, [7]Hlth.Sci. Ctr. & North China Univ. of Sci. and Technology Affiliated Hosp. & Inflammation and Immune Diseases Lab. & Sch. of Publ. Hlth., North China Univ. of Sci. and Technology, TangShan, China, [8]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Multi-ancestry Genome Wide Association Study Meta-analysis of Non-syndromic Orofacial Clefts

Zhonglin Jia[1.2,] *, Nandita Mukhopadhyay[3], Zhenglin Yang[4], Azeez Butali[5], Bing Shi[1], Jeffrey C. Murray[6,#], Liangdan Sun[7,*],Mary L. Marazita[3,8,#], Rachel M. Freathy[2,#], Robin N. Beaumont[2,*]
***Corresponding authors**
 # Senior authors

Non-syndromic orofacial clefts (NSOC) is one of the most common craniofacial birth defects, and results from both genetic and environmental factors. Genome-wide association studies have, to date, found about 75 susceptibility loci. However, these loci only explain a limited proportion of the heritability of NSOC, which include three major sub-phenotypes: non-syndromic cleft lip with palate (NSCLP), non-syndromic cleft lip only (NSCLO) and non-syndromic cleft palate only (NSCPO). Differences in genetic susceptibility among the sub-phenotypes are still poorly characterized. To identify additional susceptibility loci by performing a multi-ancestry genome wide association study (GWAS) meta-analysis on 44,094 individuals (9,381 cases, 28,510 controls and 6,203 individuals of case-parent trios) of East Asian, European, Latin and South American, and African ancestry. In addition, we aimed to compare genetic loci implicated in the different subtypes of NSOC. We confirmed 40 out of the 75 previously reported loci, and identified 14 novel loci: five loci associated with both NSOC (lowest p-value=5.11E-09) and NSCL/P (lowest p-value=5.17E-11); one locus *CALD1* associated with NSOC (rs4732060, p-value=8.65E-09), NSCL/P (rs10488465, p-value=5.17E-11) and NSCLP (rs10488465, p-value=6.78E-09), and another locus *ALX1* specifically associated with NSCLO only (p-value=2.04E-08). There were no novel loci associated with NSCPO. Six of the novel loci are located in regions containing genes associated with syndromic orofacial clefts *(NTRK1, HYAL2, ADAMTS9, CALD1, ALX1* and *SOX9*); Nine of the novel loci are located in regions containing

genes implicated in craniofacial development (*HMGCR*, *HYAL2, ADAMTS9, SHH, PRICKLE1, ALX1, SOX9, RUNX1, RBFOX2*). Genetic correlation and colocalization analyses revealed an overlap between signals associated with NSCLO, NSCPO and NSCLP, but there were also notable differences, emphasizing the complexity of common and distinct genetic processes affecting lip and palate development. Future research may focus on additional proxies for orofacial clefts and environmental influences, as well as the role of orofacial clefts for later-life health outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5126W: Multi-ancestry genome-wide association study for neuroticism

Authors:

M. Kaka, A. Topaloudi, G. Chen, P. Paschou; Purdue Univ., West Lafayette, IN

Abstract:

Neuroticism is a heritable personality trait characterized by negative emotionality and is genetically correlated with psychiatric disorders such as schizophrenia, anxiety, and depression. However, to date, studies of this trait have largely been Eurocentric. Given the differences in population genetic structure among populations, it is imperative to investigate whether non-European populations can already start benefiting from the results uncovered in large-scale European studies. Here, we report a multi-ancestry Genome-Wide Association Study (GWAS) for the trait, consisting of Africans, South, and East Asians from the UK Biobank, as well as summary statistics from the largest European Neuroticism GWAS to date. Using the meta-regression model in MR-MEGA, as well as the random and fixed effect models in METASOFT, we report a total of 93 genomic risk loci. Five of these were in novel regions in chromosomes 2, 4, 5, 10, 14, and mapped to genes *TENM3*, *CDH12*, and *LRFN5*. Leveraging the linkage disequilibrium (LD) diversity in non-European populations, we performed a multi-ancestry fine-mapping using MESuSiE. Here, an overall improvement in the credible set sizes was observed as compared to the single-ancestry fine-mapping conducted using SuSiE across all 93 regions. We then identified a total of 16 causal SNPs out of which 9 were ancestry-specific, and 7 were shared between pairs of ancestries. This suggests a basis of a shared genetic architecture for Neuroticism. Further, we prioritized all genes from this study by performing gene-based MAGMA, e-MAGMA, h-MAGMA, which are extensions of MAGMA but incorporate information from eQTL and Hi-C data respectively. We also performed tissue and cell-type enrichment analyses as well as Transcriptome-Wide Association Studies (TWAS). In all, this study

successfully reports the importance of increasing diversity in genetic studies to discover new insights into the etiology of mental health traits.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5127W: Multi-ancestry meta-analysis investigating the genetic susceptibility of intracranial aneurysms in high-risk populations

Authors:

J. Malizia[1,2], J. Pelletier[2], D. Spiegelman[3], G. Rouleau[3], C. Bhérer[1,2], S. Zhou[1,2,3]; [1]Dept. of Human Genetics, Faculty of Med. and Hlth., McGill Univ., Montreal, QC, Canada, [2]Canada Excellence Res. Chair Program in Genomic Med. and VP Dahdaleh Inst. of Genomic Med., McGill Univ., Montreal, QC, Canada, [3]Montreal Neurological Inst.-Hosp., Montreal, QC, Canada

Abstract:

Background: Intracranial aneurysms (IA) are a complex, multifactorial cerebrovascular disease characterized by a weakening of the endothelial layer leading to the formation of small dilations that pool with blood. If an IA ruptures, it frequently leads to a devastating form of stroke known as subarachnoid hemorrhage (SAH). Individuals with a family history of IAs or SAHs present the most severe disease manifestations as indicated by aneurysm multiplicity, size, age at rupture, and location within the cerebral vasculature, suggesting a strong genetic component is involved in IA development. This study investigates the genetic underpinnings of IAs using genotyping array data available from three high-risk populations that exhibit elevated IA incidence: the Finnish, Japanese, and a Canadian Inuit population. **Methods:** Genome-wide association study summary statistics for the Finnish and Japanese populations are available through FinnGen and Biobank Japan (BBJ), respectively. We used our previously published genotyping array data from the Inuit cohort to perform single-variant tests of association with SAIGE, integrating a genetic relationship matrix. Results from each were then meta-analyzed employing an inverse variance weighted approach. **Results:** Meta-analysis of the three cohorts comprised over 500,000 shared variants across populations, and 4172 IA cases and 355,210 controls. Among variants that attained genome-wide significance following Bonferroni correction, the strongest associations occurred in the 9p21 locus near *CDKN2B-AS1*, a gene linked to several cardiovascular conditions. Moreover, signals were detected in *EDNRA* and *ATP2B1* which have both been previously implicated in IA development in the Finnish and Japanese populations. Finally, we identified novel IA-associated loci in *TMEM170A* and *CFDP1*, genes involved in cardiovascular diseases, as well as moyamoya disease. The latter is a rare, chronic condition characterized by progressive narrowing of the cerebral arteries that often leads to the onset of a stroke. **Conclusions:** These

findings help to begin elucidating the complex genetic architecture of IAs in three distinct populations at an elevated risk for the disease. We recapitulated results from previous analyses of the Finnish and Japanese populations and found novel associations in two genes related to the cerebral vasculature and linked with moyamoya disease which shares a similar endpoint as ruptured IAs. Further analyses are needed to validate these variants in other cohorts of diverse ancestries and expand insights into how these loci impact IA development.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5128W: Multimodal Deep Learning for Disease Forecasting using UK Biobank ⭐

Authors:

D. Pandey, L. Xu, E. Kun, J. Wang, V. Narasimhan; The Univ. of Texas at Austin, Austin, TX

Abstract:

Predicting risk of disease incidence well before their diagnosis has been a longstanding challenge in healthcare. Genetics determine innate disease risk, but Polygenic Risk Scores (PRS) that capture this risk fail to account for acquired risk due to environmental factors. Medical imaging provides another avenue to assess an individual's disease susceptibility, as it captures components of both innate and acquired risk. However, systematic efforts to combine multiple data modalities for comprehensive disease prediction are limited. Here we use multimodal data from 30,224 individuals from the UK Biobank, including imaging, genetics, metabolomics, and electronic health records to develop robust deep learning models for forecasting 7-year incidence for Osteoarthritis (OA) and Osteoporosis (OP), two highly prevalent complex musculoskeletal diseases. We first examined the predictive power of individual data modalities and found that PRS alone is a weak predictor for OA and OP, achieving an accuracy (α) of 0.56 and 0.54 respectively, reflecting its limitation by heritability. Demographic data (age, sex, BMI) also showed weak predictive power for OA ($\alpha = 0.46$) but was a stronger predictor for OP ($\alpha = 0.70$). Lastly, NMR metabolic biomarkers were poor predictors for OA ($\alpha = 0.52$) but performed better for OP ($\alpha = 0.75$), partly due to their information on blood calcium content. We then utilized DXA imaging data from 6 different modalities (spine, lumbar, full body opaque, full body transparent, knee, hip) to generate low-level embeddings that were most associated with incidence status. These image embeddings were strong predictors for both OA ($\alpha = 0.78$) and OP ($\alpha = 0.79$). However, adding additional clinically relevant image derived phenotypes, such as mean joint space or body fat distribution did not improve prediction, suggesting that our embeddings largely captured these variables implicitly. In addition, unlike PRS prediction which

decreased in accuracy by ~10% when employed on non white British individuals, our image based predictors demonstrated no bias in prediction accuracy. Finally, we combined multimodal data including genetics, demographics, blood assays, and imaging and observed that using this integrated approach for predicting 7-year incidence of OA and OP provided the best performance, with accuracies of 0.83 for OA and 0.86 for OP. By utilizing all forms of ascertainable information available for a patient in a single model, our study demonstrates the power of multimodal data in enhancing disease forecasting and lays the foundation for personalized interventional medicine.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5129W: Multi-Trait Analysis of the FinnGen Population

Authors:

A. Stiemke[1], F. Farias[1], **N. Zhao**[1], .. FinnGen Research Project[2], I. SWANSON[1], S. Longerich[1]; [1]Merck & Co., Cambridge, MA, [2]FinnGen Res. Project, Helsinki, Finland

Abstract:

Genome-Wide Association Studies (GWAS) are a common approach used to identify genomic variants that are statistically associated with a risk for a disease or a particular trait. However, single phenotype analyses are often limited by sample size as GWAS require large sample sizes to have enough statistical power to detect significant single nucleotide polymorphisms (SNPs). To increase the power of low or under-sampled disease populations, we combine multiple related phenotypes, borrowing the genetic correlation between them. By combining related phenotypes, multi-trait analyses can increase power to identify phenotype-specific genetic association. This method can also reveal novel pleitropic loci between the related phenotypes. These insights can establish an underlying link in the formation of related disease and further broaden our understanding of disease heritability.

We developed a multi-trait pipeline initially used to process the FinnGen research project. The FinnGen research project has produced, curated, and analyzed an enormous amount of data, and GWAS for over 2,000 individual phenotypes, many of which are closely related. The Finnish population's standing as a founder population allows for easier discovery of rare SNPs less prevalent in European and admixed populations. The power of the low-case phenotypes can be elevated using a multi-trait method to further expand previous analyses.

The pipeline was developed with a "submit and forget" mindset with all processing and pre-processing contained in a single master pipeline allowing for seamless running of all data. After ensuring minimal required genetic correlation using LDSCore, two multi-trait models, Multi-Trait

Analysis of GWAS (MTAG) and ASSociation analysis based on subsETs (ASSET), with complementary population assumptions were run to ensure validity of results. SNPs significant in both analyses were further annotated and fine-mapped for causality using Genome-wide Complex Trait Analysis (GCTA).
Preliminary results from our multi-trait analysis of hypertension, ischemic heart disease (IHD), and transient ischemic attack (TIA) show a significant increase of significant SNPs detected in the low-powered TIA indication. Additionally, a significant pleiotropic genetic locus emerged between all indications. These results will improve clinical risk assessments of at-risk individuals and increase therapeutic options for these populations.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5130W: Multi-trait GWAS leverages power to identify childhood adiposity signals and reveals life-stage effects of leptin-melanocortin pathway genes

Authors:

K. Kentistou[1], J. Sundfjord[2], L. Kaisinger[1], A. Lupu[2], N. Bargas[2], R. Karimi[2], J. Tadross[1], L. Steuernagal[3], G. Dowsett[1], Y. Zhao[1], B. Lam[1], J. Bruening[3], G. Yeo[1], P. Njølstad[2], F. Day[1], M. Vaudel[2], J. Perry[1], K. Ong[1], S. Johansson[2]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Univ. of Bergen, Bergen, Norway, [3]Max Planck Inst. for Metabolism Res., Cologne, Germany

Abstract:

Child obesity poses a significant burden on childhood and later life health. The genetic correlations between early or later child BMI and adult BMI are only moderate (rg_{1year}=0.22, rg_{8year}=0.55), suggesting differing regulatory mechanisms from adults and potentially different therapeutic approaches. However, genetic signal identification has been hindered by the comparatively smaller sample sizes of published studies (n_{max}=61,111 children) and the heterogeneous age of childhood BMI measurements.
To improve discovery power, we applied genomic structural equation modelling (gSEM) to combine measured child BMI (n_{max}~72,000) in the Norwegian Mother, Father and Child Cohort Study (MoBa), with recalled childhood adiposity in UK Biobank (n_{max}~444,345) and a large meta-analysis (n_{max}~632,955) of pubertal onset in women (combined $n_{eff.}$=599,924). Using only the MoBa longitudinal growth data from 6 weeks to 8 years, we identified 152 signals associated with childhood BMI, while our combined gSEM GWAS identified a further 472 independent signals. One quarter of our childhood adiposity signals show not even nominal association with

adult BMI (P>0.05). We then applied a clustering approach to characterise signals with differing effects across childhood and identified 4 distinct clusters, one affecting mainly infant BMI and three affecting later childhood BMI.
Using our variant to gene mapping method, we identified 434 genes as confidently linked to these signals, which were enriched for hormone pathways and feeding behaviour. This included key regulators of the leptin-melanocortin pathway and GDF15 and its receptor GFRAL, which showed stronger effects on child than adult BMI. These results were supported by whole-genome sequence analyses of rare loss-of-function (LOF) variants in ~490,000 diverse-ancestry participants. Again, we found LOF variants in leptin-melanocortin pathway components that had stronger effects on child compared to adult BMI. Finally, we integrated our genetic associations with human hypothalamic single-cell data and identified several neuronal cell populations that were specifically enriched for genes acting on childhood adiposity but not adult adiposity, highlighting a distinct neuronal circuitry.
In summary, we leveraged relevant traits to assemble the largest GWAS of child adiposity, identifying genes and mechanisms that exhibit specific or predominant effects in childhood. We hypothesize that homeostatic regulators of body weight may have stronger impacts during childhood. Our data also allow future studies to discern the distinct drivers of adult BMI.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5131W: Nanopore long-read sequencing of Lewy body dementia cases from the DementiaSeq1 cohort for structural variation and integrative genetics study

Authors:

E. Rice[1,2], C. Alba[1,2], A. Ray[3], K. Kaivola[3], M. Rasheed[4], J. Ding[4], J. R. Gibbs[4], R. Chia[4], B. J. Traynor[4], S. W. Scholz[3], C. L. Dalgard[1]; [1]The American Genome Ctr., Collaborative Hlth.Initiative Res. Program, Uniformed Services Univ., Bethesda, MD, [2]Henry M. Jackson Fndn. for the Advancement of Military Med., Bethesda, MD, [3]Neurodegenerative Diseases Res. Section, Natl. Inst. of Neurological Disorders and Stroke, NIH, Bethesda, MD, [4]Lab. of Neurogenetics, Natl. Inst. of Aging, NIH, Bethesda, MD

Abstract:

Large-scale genomics studies, such as DementiaSeq1, have advanced the understanding of the genetic etiologies contributing to Lewy body dementia (LBD). However, the limited case size and multifactorial aspects of this neurodegenerative disease contribute to missing determinants of LBD. As an orthogonal strategy to investigate the DementiaSeq1 cohort, a subset of LBD cases

has been selected for deep molecular profiling that includes bulk comprehensive transcriptomics, array-based genome-wide methylation, single-cell RNA-seq, spatial transcriptome profiling, and, described here, nanopore-based long-read whole genome sequencing. This subset of approximately 200 units comprises brain tissue from pathologically confirmed individuals with LBD (n = 131) and frontotemporal dementia (n = 67), alongside controls (n = 10). High-molecular weight genomic DNA was isolated following protocols optimized for Oxford Nanopore workflows, and library preparation by ligation-based methods was conducted before sequencing on a PromethION platform. Sequencing runs were completed on R10.4.1 flow cells over the course of four days, with wash and reload cycles on days two and three. Basecalling was performed using Dorado high accuracy model with modified bases, and a Q9 quality filtering threshold. Long-read WGS data was compared to existing short-read whole genome sequencing data from DementiaSeq1 to confirm single nucleotide variant concordance. More importantly, long-read WGS data demonstrated high confidence structural variation (SV) calls that overlap candidate SV calls from the short read WGS data. For example, a ~300 bp deletion in *TCPN1* was identified, as well as novel candidate SV, potentially within loci associated with neurodegenerative diseases. Furthermore, integrated analysis of SV with transcriptome data from matching tissue aliquots were conducted to evaluate discordance of transcript expression levels with these genetic alterations. As an additional extension to the publicly available data sharing initiative of DementiaSeq1 and future DementiaSeq2, these long-read WGS data will be provided in several data repositories for the research community utilization and study.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5132W: Novel Genetic Variants Linked to Liver Cirrhosis and Cancer: Evidence of Gene-Environment Mediation and Interaction in the Million Veteran Program ⭐

Authors:

M. Vujkovic[1,2], D. Kaplan[1,2], J. Ghouse[3], M. Levin, G[1,2], J. Brancale[4], A. Lewis[5], O. Veatch[6], C. Schneider[7], B-L. Chang[1], A. Verma[1,2], K. Wangensteen[8], D. Gill[9], C. Konkwo[4], T-H. Schwantes-An[10], A. Shaked[1], K. Olthoff[1], M. Hoteit[1], D. Dochtermann[11], S. Pyarajan[11], K. Maxwell[1,2], D. Rader[1], Y. Tanaka[12], B. Voight[1,2], S. Vilarinho[4], L. Bastarache[5], S. Stender[3], P. Tsao[13,14], T. Morgan[15], J. Lynch[16], K-M. Chang[1,2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [3]RigsHosp.et, Copenhagen Univ. Hosp., Copenhagen, Denmark, [4]Yale Univ., New Haven, CT, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Univ. of Kansas Med. Ctr., Kansas City, KS, [7]RWTH Aachen Univ., Aachen, Germany, [8]Mayo Clinic, Rochester,

MN, [9]Imperial Coll. London, London, United Kingdom, [10]Indiana Univ. Sch. of Med., Indianapolis, IN, [11]Boston VA Med. Ctr., Boston, MA, [12]Kumamoto Univ., Kumamoto, Japan, [13]Stanford Univ., Palo Alto, CA, [14]Palo Alto VA Med. Ctr., Palo Alto, CA, [15]VA Long Beach Hlth.care System, Long Beach, CA, [16]Salt Lake City VA Med. Ctr., Salt Lake City, UT

Abstract:

Over the past two decades, metabolic-dysfunction associated steatosis (formerly known as non-alcoholic fatty liver disease or NAFLD) has become the leading cause of chronic liver disease (CLD), cirrhosis, and hepatocellular carcinoma (HCC). Here, we conducted the largest multi-ancestry genome-wide meta-analysis of cirrhosis and HCC, including 35,481 cirrhosis cases and 1,914,739 controls, as well as 7,126 HCC cases and 1,757,505 controls among several large biobanks, including the Million Veteran Program (MVP). We identified 27 risk loci for cirrhosis (9 novel) and 12 for HCC (4 novel), with 8 loci overlapping
(*PNPLA3*, *HSD17B13*, *TM6SF2*, *APOE*, *HLA*, *IFNL3*, *CD81*, *HFE*).
Given that many loci are known risk factors for CLD, we hypothesized that the relationship between these variants and disease progression might be mediated by underlying disease causes. We applied counterfactual mediation analysis to estimate the direct and indirect effects of variants on cirrhosis through mediators including metabolic dysfunction, alcohol use, viral hepatitis, and other causes of liver disease in 11,738 cirrhosis cases and 444,882 controls in MVP. Most loci showed evidence of partial mediation, indicating that genetic variants contribute to cirrhosis directly and indirectly by increasing susceptibility to CLD. Complete mediation was observed in two loci only (e.g., ADH1B via alcohol use, and HFE via hemochromatosis).
Finally we examined whether individuals with liver disease and a high polygenic risk score (PRS, top 5th percentile) for cirrhosis were at increased risk of disease progression using multi-state Markov survival models using the transitional states of CLD, cirrhosis, HCC, and liver mortality over a 20 year follow-up period in 134,841 Europeans (EUR), 20,068 Hispanics (HIS), and 39,397 African Americans (AA) in MVP. A high PRS was associated with progression from CLD to cirrhosis (EUR hazard ratio (HR)=1.94, $P=2x10^{-68}$; HIS HR=1.55, $P=4x10^{-06}$; AA HR=1.3, $P=5x10^{-04}$) and from cirrhosis to HCC (EUR HR=1.65, $P=7x10^{-08}$; HIS HR=1.46, P=0.05; AA HR=1.26, P=0.19). In a separate survival model including interaction terms in Europeans, metabolic dysfunction showed interactions with PRS on cirrhosis (P=0.005), with alcohol use on cirrhosis ($P=5x10^{-28}$), and with viral hepatitis on cirrhosis ($P=9x10^{-11}$), HCC (P=0.001) and liver mortality ($P=2x10^{-07}$). Collectively, our study implicates novel genes with disease progression in CLD. Metabolic dysfunction is both an independent risk factor as well as accelerator of progression when combined with a high genetic burden, underscoring the need for aggressive management of cardiometabolic risk factors in people with CLD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5133W: Optimization and Validation of Multi-Ancestry Polygenic Risk Scores

Authors:

P. Yazdi, J. Lerga-Jaso, B. Novkovic; KureAI, Los Angeles, CA

Abstract:

Polygenic risk scores (PRS) have ushered in a new era in genetic epidemiology, offering insights into individual predispositions to a wide range of diseases. This study aimed to develop and benchmark multi-ancestry PRS models capable of predicting disease risk across diverse populations. Leveraging trans-ethnic GWAS meta-analysis, we generated novel summary statistics for 30 medically related traits and assessed the predictive performance of four PRS algorithms. Algorithm efficacy across traits varied, with Stacked C+T (SCT) performing better when trait prevalence was higher, while LDpred and Lassosum performed better when trait prevalence was lower. We subsequently benchmarked and validated our PRS models in the page-mec dataset, demonstrating that they are well calibrated with only a minor drop in performance in non-European samples. We then directly tested and validated our models against the PRS from the eMERGE network demonstrating their superior performance in 7 out of 8 phenotypes tested. Subsequent integration of PRS algorithm outputs through logistic regression and incorporation of additional medical information enhanced model accuracy. Notably, including ancestry information further improved predictive performance, underscoring its importance in PRS model development. Our findings highlight the potential of multi-ancestry PRS models in clinical settings, demonstrating superior predictive accuracy and broad applicability across different ancestral backgrounds.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5134W: PBMC profiling in single ventricle/hypoplastic left heart patients by single-cell analysis

Authors:

H-Q. Qu[1], K. Ostberg[1], D. J. Slater[1], F. Wang[1], J. Snyder[1], C. Hou[2], J. Connolly[1], M. March[2], J. Glessner[3], C. Kao[1], H. Hakonarson[2]; [1]Children s Hosp. of Philadelphia, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Childrens Hosp. of Philadelphia/PENN, Philadelphia, PA

Abstract:

Background: Single ventricle and hypoplastic left heart syndrome (SV/HLHS) are congenital heart defects, necessitating restructuring of circulation to optimize blood flow and oxygenation. Affected patients require lifelong medical monitoring and management to address potential complications and optimize their health. Peripheral blood mononuclear cells (PBMCs) mediate the response to viral infections and affect cardiovascular complications in SV/HLHS, including heart failure, arrhythmias, and vascular dysfunction. Assessing transcriptome profiles in PBMCs may offer valuable insights into the pathogenesis and provide new biomarkers for improving disease management and assessing prognosis. **Methods**: Using single-cell RNA sequencing (scRNA-seq), we studied PBMCs from 36 cases with single ventricle/hypoplastic left heart and 40 controls without heart diseases. The experiments were done in 4 independent batches (8 cases, 8 controls; 20 cases, 15 controls; 4 cases, 11 controls; 4 cases, 6 controls), using 10X Chromium Single Cell Gene Expression assay (10x Genomics, Single Cell 3' v3). Sequencing was performed using the Illumina NovaSeq6000. Data analysis was conducted using the Seurat R package, employing SCTransform for data normalizing and scaling (Satija, Farrell et al. 2015, Butler, Hoffman et al. 2018). Uniform manifold approximation and projection (UMAP) (Becht, McInnes et al. 2019) was used to group cells into clusters. Cell typing was done using the celldex: DatabaseImmuneCellExpressionData function(Aran, Looney et al. 2019). To mitigate batch effects, the 4 batches were analyzed separately. **Results**: A total of 112 genes were consistently identified of differential expression (DE) ($|\log2FC|\geq0.25$, $Padj<0.05$) in each of the 4 batches, including 66 upregulated and 46 downregulated genes. Notably, among the 66 upregulated genes, two gene sets, HALLMARK_TNFA_SIGNALING_VIA_NFKB (FDR=1.51E-05) and HALLMARK_APOPTOSIS (FDR=8.80E-04), were significantly overrepresented, including 14 genes (*BNIP3L, CCNL1, CD69, CTNNB1, DNAJA1, EIF1, IRF1, JUN, KLF6, NFKBIA, PPP1R15A, SAT1, SQSTM1, TNFAIP3*). Further analysis identified DE of these genes in different cell types, with the most significant changes observed in naïve CD8+ T cells, naïve CD4+ T cells, and NK cells. **Conclusion**: We identified DE genes and cell types in PBMCs affected by SV/HLHS. These changes may be related to increased susceptibility to infections, altered inflammatory and immune responses, and contribute to myocardial remodeling and cardiac dysfunction in SV/HLHS. Targeting these genes and cell types could potentially improve the management of these diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5135W: Phenomics Network Analysis for Depression Reveals Novel Comorbidity Architecture, Genomic Relationships, and Pleiotropic Variants

Authors:

G. Chen[1], Q. Zhong[1], Z. Yang[2], P. Jain[3], P. Drineas[4], P. Paschou[1]; [1]Dept. of Biological Sci., Purdue Univ., West Lafayette, IN, [2]Univ. of Helsinki, Helsinki, Finland, [3]Lee Kong Chian Sch. of Med., Nanyang Technological Univ., Singapore, Singapore, [4]Dept. of Computer Sci., Purdue Univ., West Lafayette, IN

Abstract:

Depression is one of the most prevalent psychiatric disorders and is a leading cause of health ailments worldwide. It is highly heritable and is frequently comorbid with other mental and physical traits. This observation motivated us to look deeper into the genetic and phenotypic connections between depression and other traits to identify correlations as well as potentially causal connections between them. In this project, we utilize electronic health records (EHRs) from the UK biobank and All of Us to systematically evaluate relationships between depression and other heritable traits from both a phenotypic and a genetic aspect. We compressed available ICD diagnose codes into more than 1,500 heritable phecodes and constructed a comorbidity network connecting depression and other disorders on over 300,000 participants of European ancestry. We investigated the genetic correlation for each connection in the network. We also investigated potentially causal relationships using mendelian randomization for all pairs of significantly correlated disorders and uncovered horizontal pleiotropic genetic variants and genes contributing to disease etiologies. We found gastroesophageal reflux disease (GERD), constipation, headache, anxiety and bipolar disorder contribute to be direct causes for depression, with GERD and headache lying at the center of the causal network. Genes broadly expressed in various tissues, such as NEGR1, TCF4, and BTN2A1 underlie the pathways leading not only to depression but also to other related disorders. We highlight the broad connections between depression and diverse traits, indicating a complex etiology and possible existence of subtypes for depression. Our findings highlight the value of cross-trait analysis in better understanding the neurobiology of complex psychiatric disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5136W: Pima/O'odham Informed Consent in Genomic Studies: a Scoping Review

Authors:

E. Manriquez, K. Tsosie, C. Banta, R. Cu:k Ba'ak; Arizona State Univ., Tempe, AZ

Abstract:

Objective: We conducted a Scoping Review of who is using O'odham/Pima genomic datasets that correlate with Diabetes Mellitus to assess who is using said data, the proportionality of Community Engagement within these studies (if any at all), and to answer the question, could these samples still be used in subsequent studies.Methods: We conducted a PubMed search to locate articles that extracted biological samples from contemporary Pima/O'odham individuals (i.e. blood, saliva, plasma, adipose tissue, urine, etc.), as well as articles that utilized those datasets. To keep track of such data, we used Covidence and Excel to keep record of methodical values through a numerical system that assigned values to data/methods of research articles (i.e. blood = 1, saliva = 2, adipose tissue = 7, etc).Results: We identified 332 articles that matched our criteria, 223 of said articles are Primary (directly sourced genetic material) while 40 are Secondary (used data made publicly available or retrieved data from the 223 mentioned). Conclusions: Although our Scoping Review might not fully encompass the entirety of O'odham/Pima studies in current academia, it does rather point to an ethical concern within the field. That is, how often are the individuals whom samples are being taken of are fully aware of what will happen to said samples, and aware of the key possibility of them being shared outside these pre-ordained groups (i.e. 40 secondary articles), or even kept for future studies that don't fit the pre-recognised purpose. Thus, more accountability should be pointed to researchers who fail to give full informed consent to Indigenous participants.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5137W: Polygenic prioritization of monogenic IBD genes

Authors:

C. Barnhill[1], X. Zeng[2], S-G. Ji[3]; [1]BridgeBio, Houston, TX, [2]BridgeBio Pharma, San Francisco, CA, [3]BridgeBio Pharma Inc., Arlington, MA

Abstract:

Inflammatory bowel disease (IBD) is a common disease with high unmet need. Over 50,000 children in the United States have IBD. A small subset of children who develop IBD-like intestinal

inflammation before six years of age have a single gene defect that affects immune function or disturbs epithelial barrier function, collectively known as monogenic IBD (mIBD). While anti-tumor necrosis factor (TNF) agents have been effective in treating IBD, up to one-third of patients treated with anti-TNF agents do not respond. Of those who do respond, 13 percent develop loss of response each year. Our goal is to combine common variant evidence from large IBD genome-wide association studies (GWAS) with evidence from monogenic, Mendelian forms of IBD to identify genes and pathways that address the gap in the current treatment landscape.We incorporated generalized gene-set analysis of GWAS data and computed a polygenic priority score (PoPS) on a recent meta-analysis of IBD, Crohn's disease (CD), and ulcerative colitis (UC) in individuals of European and East Asian ancestry from the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). We analyzed the PoPS scores for 102 mIBD genes with high and moderate penetrance and found that *NFKBIA*, *TNFAIP3*, and *IL10RA* were enriched across all three of the GWAS traits.In this work, we utilized common variant associations to prioritize targets that were previously identified in monogenic IBD. This approach can be used to identify genetic subsets of common complex diseases to find monogenic drivers of the disease, increasing our motivation to develop therapies for patients with both Mendelian and complex forms diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5138W: Polygenic risk score model based on polymorphisms related to metabolic syndrome offers valuable genetic insights

Authors:

Y. Park[1], S. Moon[1], J. Choi[1], J. Kim[1], H-Y. Son[2], S-W. Im[3], J-i. Kim[2]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of, [3]Kangwon Natl. Univ., Gangwon, Korea, Republic of

Abstract:

Background and Purpose: Genome-wide association studies(GWAS) regarding metabolic syndrome(MetS) are relatively scarce in the current literature. This study aimed to extract SNPs from a large population GWAS of MetS, create a polygenic risk score(PRS) model based on these SNPs, validate the PRS in an independent cohort, and then perform additional fine-mapping within Koreans and European ethnicities. **Methods:** Based on the Korean Genome and Epidemiology Study(KoGES) city cohort, 58,600 Korean subjects with available biochemical information and demographic variables for MetS were included. These subjects were genotyped with the Korean Biobank Array and imputed with the Northeast Asian Reference Database 2

panel. We underwent GWAS on MetS and extracted a total of 11 statistically independent SNPs from GCTA-COJO (Genome-wide complex trait analysis tool, COnditional and JOint multiple-SNP analysis), performed fine-mapping (based on the Sum of Single Effects from PolyFUN), and created a PRS model. The PRS model was established from the odds ratios derived from age, sex-adjusted GWAS, and the dosage for each SNP of each genotyped participant. The PRS was validated within another independent KoGES rural cohort(N=8,105) and the frequency of MetS was calculated for each PRS quintile based on various MetS criteria. **Results:** We conducted GWAS on MetS and its diagnostic components with covariates of age and sex. 11 SNPs from 5 different loci including *APOA5, CETP, LPL, APOE,* and *SNHG32* genes were extracted from GCTA-COJO. The 11 SNPs extracted included rs651821 (fine-mapping posterior inclusion probability 0.99 in Koreans, but not in the UK Biobank) from *APOA5*. The PRS model was validated in the independent KoGES rural cohort and revealed a consistent increase of MetS case frequency for each increasing PRS quintile from 1-bin to 5-bin(38, 41, 44, 47, 51(%), respectively, average 44%). Because there was a large proportion of MetS(3 out of 5 criteria) patients, more severe subsets of MetS - 4 out of 5 diagnostic criteria applied, 15, 17, 20, 24, and 25(%) respectively, and for 5 out of 5 diagnostic criteria - 2.9, 3.8, 4.4, 6.7, 7.9(%), for PRS quintiles respectively, was also analyzed and revealed that the quotient of 5-bin to 1-bin was significantly higher in the 5 out of 5 diagnostic criteria compared to other diagnostic criteria applications. **Conclusions:** A PRS model of 11 SNPs derived from MetS GWAS and GCTA-COJO of the KoGES city cohort was validated in an independent rural cohort of identical ancestry. This genomic model could be integrated into the hospital clinical setting, and contribute to the advanced precision medicine of metabolic syndrome and obesity in the future.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5139W: Polygenic risk score predicts incident cardiovascular disease in people with diabetes at borderline or intermediate 10-year ASCVD risk ⭐

Authors:

J. Choi[1], H. Lee[1], T. Rhee[2], J. Park[1], J. Lee[1], Y. Jo[1], S. Lee[1], S. Kwak[2]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract:

People with type 2 diabetes (T2D) and 10-year atherosclerotic cardiovascular disease (ASCVD) of 20% or higher are recommended for intense lipid management. However, the role of genetic predisposition to ASCVD on determining the intensity of statin therapy in patients with T2D at

borderline (5-7.5%) or intermediate (7.5-20%) risk for incident ASCVD is unclear. We aimed to investigate the utility of polygenic risk score (PRS) in predicting incident ASCVD in people with T2D at borderline and intermediate risk from UK Biobank. We calculated 10-year ASCVD risk using pooled cohort risk equations and PRS for coronary artery disease using the Polygenic Risk Score-Continuous Shrinkage Priors method (PRS-CS) in participants of UK Biobank. We performed time-to-event analysis for incident ASCVD and assessed improvement of C-statistic and continuous net reclassification improvement (NRI) when PRS was added to 10-year ASCVD risk. In addition, risk of ASCVD was estimated for those who were in the top 10 percentile of PRS compared to the remainders. Statin medication status and first 10 genetic principal components those showing a statistically significant association with incident ASCVD were adjusted in statistical models. During a median follow-up of 14 years, the incidence of ASCVD in people with T2D was 5.9% and 8.6% in borderline (N=2,619) and intermediate (N=13,173) group, respectively. The normalized mean PRS for those who developed ASCVD was higher in each group: 0.50 (95% CI -1.36-2.36) vs 0.12 (95% CI -1.86-2.10) in borderline and 0.26 (95% CI -1.71-2.23) vs 0.11 (95% CI -1.86-2.07) in intermediate group (all P<0.001). Standardized PRS was associated with incident ASCVD in each group when conducted time-to-event analysis: HR 1.43 (95% CI 1.18-1.72, P<0.001) and 1.09 (95% CI 1.03-1.16, P<0.01). Moreover, those who were in the top 10 percentile had increased risk of ASCVD: HR 1.97 (1.23-3.26, P<0.01) and 1.25 (1.04-1.84, P<0.05), respectively. Adding PRS to 10-year ASCVD risk improved the C statistic from 0.59 (95% CI 0.54-0.65) to 0.64 (CI 0.59-0.69) in borderline and from 0.54 (CI 0.52-0.56) to 0.56 (CI 0.54-0.58) in intermediate group (all P < 0.01). Adding PRS to 10-year ASCVD risk resulted in an intermediate level of NRI in borderline group (0.34, 95% CI 0.16-0.53, P<0.001) and weak but significant level in intermediate group (0.11, CI 0.04-0.17, P<0.01). In conclusion, genetic information in people with T2D at borderline and intermediate ASCVD risk could be used to identify those who are at higher risk for incident ASCVD and require high intensity statin treatment. This study was supported by a grant from the National Human Genome Research Institute (U01HG011723)

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5140W: Polygenic Risk Score Predicts Lung Function Response to PM2.5 Exposure in Children

Authors:

J. Kelchtermans, H. Qu, F. Mentch, M. March, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Rationale: The relationship between Particulate Matter 2.5 (PM2.5) exposure and asthma exacerbations is well-established. However, genetic variability in the response to PM2.5 is poorly understood, complicating efforts to mitigate the impact of PM2.5 exposure. We hypothesized that a polygenic risk score (PRS) could predict short-term changes in lung function after exposure to PM2.5.
Methods: We drew spirometry tests completed by pediatric patients between January 1st, 2008, and December 31st, 2022, from a de-identified database. For these patients, a PRS was calculated based on a GWAS comparing children with asthma exacerbations following PM2.5 spikes to those with asthma but no temporal relationship between exacerbations and PM2.5 exposure. Forced expiratory volume in one second (FEV1) over forced vital capacity (FVC) ratio was compared between patients with a high PRS z-score (>1) and others. Differences were assessed using t-tests both on days with PM2.5 < 12 µg/m3 ('baseline') and following days with PM2.5 > 12 µg/m3 ('PM2.5 exposed'). A linear regression model was created to adjust for age, sex, ethnicity, clustering at the patient level, and lifetime incidence of PM2.5 > 12 µg/m3.
Results:There were 18,233 tests from 5,268 patients that met inclusion criteria The cohort was primarily male (54%), with most frequent recorded race 'White' (62%) or 'Black or African American' (24%). Mean age at time of testing was 12.3 ± 4.1 years. No significant differences in FEV1/FVC were found at baseline. In the 4,982 (27%) spirometry tests classified as PM2.5 exposed, patients with a high PRS had significantly lower FEV1/FVC ratios than others (p = 0.00045). This remained significant after correcting for age, sex, ethnicity, and lifetime incidence of PM2.5 > 12 µg/m3 (P=0.0059).
Discussion: This study presents the first PRS to predict changes in spirometry measurements after modest increases in PM2.5 exposure in children. Linking genetic sensitivity with a widely used clinical test offers a framework to assess how therapeutic interventions interact with genetic sensitivity to determine the impact of PM2.5 exposure. Follow-up studies should explore how gene-gene and gene-environment interactions modulate the clinical consequences of sensitivity to air pollution.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5141W: Polygenic score-wide association study of comorbidities to incident dementia

Authors:

S-H. Jung[1,2], D-g. Lee[1], J. Woerner[1], Y. Nam[1], L. Shen[1], H-H. Won[2], D. Kim[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Sungkyunkwan Univ, Seoul, Korea, Republic of

Abstract:

Background: Previous studies have reported various prodromal symptoms and comorbidities that precede the development of dementia. In this study, we comprehensively investigate the impact of genetic predispositions to multiple comorbidities on the risk of incident dementia. **Methods:** Our study included 377,653 participants of European descent from the UK Biobank, comprising 370,183 controls and 7,470 cases of incident dementia. We utilized large-scale genome-wide association study (GWAS) data from the FinnGen Consortium to generate the polygenic score (PGS) for quantifying genetic predispositions. From 2,405 GWASs, we selected those with a heritability greater than 1% using LD Score regression. We then analyzed pairwise genetic correlations, retaining the more heritable factor in cases where the collinearity (genetic correlation) exceeded 0.9. Consequently, we included 331 eligible PGSs in the Cox proportional hazard model-based polygenic score-wide association study (PolyWAS) for incident dementia. **Results:** We identified 53 significant comorbidity-PGSs (Bonferroni corrected p-value=0.05/331) through PolyWAS analysis, adjusted for the covariates age, sex, *ApoE* ε4 status, and dementia PGS. Notably, cardiometabolic-related diseases categorized under the ICD-10 code blocks 'I' (diseases of the circulatory system) and 'E' (endocrine, nutritional and metabolic diseases) showed the most significant associations with the risk of incident dementia. Depression and mood disorders under category 'F' (mental, behavioral and neurodevelopmental disorders) were risk factors in Alzheimer's disease (AD) but not in Vascular dementia (VD). In addition, the genetic predisposition to 'Medical abortion' under the category 'O' emerged as a risk factor for all-cause dementia and its subtypes, including AD and VD. All significant PGSs were positively associated with incident dementia. **Conclusion:** Our study implemented a novel association test approach using comorbidity-PGSs to investigate the genetic predispositions associated with numerous comorbidities and their impact on incident dementia and its subtypes. The proposed PolyWAS approach can better advance our knowledge of disease mechanisms and enhance our understanding of the shared genetics of comorbidities.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5142W: Polygenic trait-wide analysis identifies complex traits associated with immune-mediated inflammatory diseases and their clinical phenotypes

Authors:

A. Aterido[1,2], S. Martínez-Hilario[1], A. Gómez[2], E. Domènech[3], J. Cañete[4], J. Tornero[5], J. Gisbert[6], A. Fernández-Nebro[7], J. Carrascosa-Carrillo[8], J. Calvo[9], E. Fonseca[10], J. Rodríguez[11], P. Carreira[12], E. Ricart[13], V. Garcia[14], F. Blanco[15], N. Palau[2], R. Tortosa[2], M. Lopez-Lasanta[2], S. Marsal[1,2], A. Julià[1,2]; [1]IMIDomics, Inc, Barcelona, Spain, [2]Rheumatology Res. Group, Vall Hebron Res. Inst.,

Barcelona, Spain, [3]Gastroenterology Dept., Hosp. Univ.ri Germans Trias i Pujol, Badalona, Spain, [4]Hosp. Clinic de Barcelona and IDIBAPS, Barcelona, Spain, [5]Rheumatology Dept., Hosp. Univ.rio Guadalajara, Guadalajara, Spain, [6]Gastroenterology Dept., Hosp. Univ.rio de la Princesa, Madrid, Spain, [7]Rheumatology Dept., Hosp. Regional Univ.rio Carlos Haya, Málaga, Spain, [8]Dermatology Dept., Hosp. Univ.ri Germans Trias i Pujol, Badalona, Spain, [9]Rheumatology Dept., Hosp. Univ.rio Araba, Álava, Spain, [10]Dermatology Dept., Complejo Hosp.ario Univ.rio de A Coruña, A Coruña, Spain, [11]Rheumatology Dept., Hosp. Univ.ri de Bellvitge, Barcelona, Spain, [12]Rheumatology Dept., Hosp. Univ.rio 12 de Octubre, Madrid, Spain, [13]Gastroenterology Dept., Fundació Clínic per a la Recerca Biomèdica, Barcelona, Spain, [14]Gastroenterology Dept., Hosp. Univ.rio Reina Sofía, Córdoba, Spain, [15]Rheumatology Dept., Complejo Hosp.ario Univ.rio de A Coruña, A Coruña, Spain

Abstract:

Background Immune-mediated inflammatory diseases (IMID) are characterized by self-directed inflammatory responses. GWAS have identified many SNPs associated to complex traits like IMID, which can be integrated into Polygenic Scores (PGS) to estimate the individual's susceptibility to manifest a trait. PGS-wide analysis is starting to be conducted in large cohorts. In IMID, however, how polygenic predisposition to complex traits shapes autoimmunity is unknown **Objective** To determine which polygenic factors underlying complex traits also contribute to IMID clinical heterogeneity **Methods** Genotyped data from 8,019 subjects (1510 controls; 1285 Rheumatoid arthritis (RA); 1058 Psoriasis (PS); 1071 Psoriatic arthritis (PsA); 874 Lupus (SLE); 1298 Crohn's disease (CD) and 912 Ulcerative colitis (UC) patients) and clinical data for 889 IMID phenotypes were used. All patients were collected by the IMID Consortium. For each patient, >39M SNPs were imputed to compute the PGS of 2,110 complex traits. Using case-control and case-case study designs, each PGS was tested for association to each disease and phenotype. Statistical analysis was adjusted for confounding factors and, subsequently, False Discovery Rate (FDR) method was applied **Results** Polygenic trait-wide case-control analysis detected 487 significant associations, confirming a higher IMID risk in patients than in controls (FDR≤6.6e-40). In RA, patients showed a strong genetic predisposition to high serum levels of IL6 (FDR=2.1e-5), the target of an approved drug. SLE patients exhibited a marked predisposition to high levels of cytokine fractalkine or adrenomedullin (FDR<9.6e-4), supporting them as new targets for SLE. Consistent with prior findings on genetic risk sharing across IMID, PGS of Multiple Sclerosis (FDR=8.7e-8) and Systemic Sclerosis (FDR=5.8e-07) were found to be higher in SLE than in controls. CD was associated with genetic predisposition for high levels of inflammatory proteins like PTX3 (FDR=1.1e-2). At the phenotype level, SLE showed the largest number of PGS-associated phenotypes (n=31, FDR<0.05). In CD, the PGS of BMI was positively associated with disease activity (FDR<1.8e-3). We also detected confirmatory phenotype associations like a higher hypertension-PGS in patients with high blood pressure (FDR<4.7e-3).

Notably, we found nominal associations with clinical response to TNF blocking. Responder patients are genetically more predisposed to high plasma levels of MMP7 in CD (P=5.2e-4) and CD40 in PsA (P=2.4e-3) than refractory patients **Conclusion** By leveraging the power of PGS, we provide new phenotypic and molecular associations in IMID suggesting novel drug targets

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5143W: Potential Causal Link between Hyperuricemia and Incident Chronic Kidney Disease in Eastern Asians: Insights from Large-scale Mendelian Randomization Analyses

Authors:

H-L. Chen[1,2], C-T. Chao[3,4], C-C. Lin[1], H-Y. Chiang[1,2], Y-T. Lin[1,2], Y-C. Lin[1,2], H-C. Yeh[5], I-W. Ting[5], Y-C. Hong[5], D. Chang[5], W-C. Syu[6], Y-T. Lin[6], C-C. Kuo[1,2,5]; [1]Big Data Ctr., China Med. Univ. Hosp., China Med. Univ., Taichung, Taiwan, [2]Dept. of BioMed. Informatics, China Med. Univ., Taichung, Taiwan, [3]Div. of Nephrology, Dept. of Internal Med., Natl. Taiwan Univ. Hosp. and Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [4]Div. of Nephrology, Dept. of Internal Med., Min-Sheng Gen. Hosp., Taoyuan City, Taiwan, [5]Div. of Nephrology, Dept. of Internal Med., China Med. Univ. Hosp., Taichung, Taiwan, [6]Atgenomix, Taipei, Taiwan

Abstract:

Background: Elevated serum uric acid (SUA) has been associated with incident chronic kidney disease (CKD). However, the causality between SUA and CKD remains controversial, as prior clinical trials evaluating whether uric acid-lowering agents reduce the risk of CKD have shown inconclusive results. Therefore, we investigated whether SUA is an independent and causal risk factor for CKD through a large Mendelian randomization (MR) analysis using data from three key biobanks in Taiwan and Japan.**Methods:** China Medical University Hospital established the largest phenome-genome data platform in 2020, known as the iHi Platform. This platform includes data from 3 million patients and 344,202 genetic chip information entries. Patients with hyperuricemia who were free of CKD were identified using SUA levels, with hyperuricemia defined as SUA > 7.0 mg/dL for males and > 6.0 mg/dL for females. Cox proportional hazards modeling was used to evaluate the association between hyperuricemia and the development of CKD, which was defined as a follow-up estimated glomerular filtration rate measurement of less than 60 mL/min/1.73 m^2. A primary and sensitivity two-sample MR analyses were performed using a genetic instrument containing 73 and 90 independent SNPs associated with SUA from the Taiwan Biobank and Biobank Japan, respectively. Summary statistics for SUA levels were obtained from a genome-wide association study (GWAS) conducted by Taiwan Biobank

(n=88,879), and CKD data were sourced from iHi Platform (n=223,134). The MR analysis adhered to the classic MR assumptions. The intercept test from the MR-Egger regression indicated no significant evidence of horizontal pleiotropy. **Results:** A total of 298,560 eligible patients were analyzed, with a median age of 44.6 years (IQR: 33.5-56.3) and an average follow-up of 3.57 years. The fully adjusted hazard ratio for incident CKD in patients with hyperuricemia compared to those without was 1.22 (95% confidence interval (CI), 1.16-1.28). The primary MR analysis, MR-Inverse-Variance Weighted (MR-IVW), indicated that each 1 standard deviation increase in genetically predicated SUA levels increased the odds of CKD (odds ratio [OR] 1.30, 95% CI 1.19-1.41). This finding was robust across multiple sensitivity analyses, including six other different MR methods and sex stratification. Sex modified the causal impact of SUA on CKD, showing a weaker OR of MR-IVW (1.16, 95% CI 1.08-1.25) among women. **Conclusions:** Our findings, in contrast to the previous neutral results from a large Caucasian-based MR study, confirm the potential causal role of hyperuricemia in the development of CKD in East Asian populations.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5144W: Potential genetic regulators of a novel plasma-based biomarker for Alzheimer's disease

Authors:

D. Goldstein[1], N. Aytan[2], J. Palmisano[3], Y. Tripodis[3], K. W. Turk[2,4], G. Heo[5,6], M. K. O'Connor[2,7], L. E. Goldstein[2], W. Qiu[2], N. W. Kowall[2], A. C. McKee[2,8], C. Cruchaga[5,6], J. Mez[2], M. L. Alosco[2], L. A. Farrer[1,2,9,3,10], T. Stein[8,4], G. R. Jun[1,9,3]; [1]Dept. of Med. (BioMed. Genetics), Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [2]Dept. of Neurology, Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [3]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [4]VA Boston Hlth.care System, Boston, MA, [5]Dept. of Psychiatry, Washington Univ. Sch. of Med. in St. Louis, St. Louis, MO, [6]NeuroGenomics and Informatics Ctr., Washington Univ. Sch. of Med. in St. Louis, St. Louis, MO, [7]VA Bedford Hlth.care System, Bedford, MA, [8]Dept. of Pathology & Lab. Med., Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [9]Dept. of Ophthalmology, Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [10]Dept. of Epidemiology, Boston Univ. Sch. of Publ. Hlth., Boston, MA

Abstract:

Introduction: Identification of blood-based biomarkers that are indicative of Alzheimer’s disease (AD) pathology may help with early detection of AD. We previously identified a novel plasma biomarker (MMP-8) associated with both antemortem clinical and postmortem pathological

diagnosis of AD in brain donors from the Boston University Alzheimer's Disease Research Center (BUADRC). The purpose of this study was to gain insight on the molecular basis for MMP-8 through the investigation of genetic regulators of plasma MMP-8 expression. **Methods:** We generated proteome data for ~7,000 proteins in both antemortem plasma and postmortem brain tissue from 103 participants of the BUADRC. All participants completed comprehensive neuropsychological evaluations and dementia severity was measured by global cognitive dementia rating (CDR) scores. A pathological AD diagnosis (PathAD) was established using NIA Reagan criteria. We conducted association tests of protein level in plasma and brain with dementia (i.e., CDR ≥1 for dementia, CDR <1 for control) and autopsy-confirmed AD status as binary outcomes, adjusted by sex and either age at last antemortem exam or death. We combined association results in plasma and brain by meta-analysis. Significantly associated proteins ($p<0.05$) from analyses in plasma, brain, and combined specimens for both PathAD and dementia diagnosis in BUADRC were validated with an independent dataset including ~5000 individuals. A genome-wide association study (GWAS) was conducted using the plasma level of the top validated protein as a quantitative outcome in the BUADRC, adjusting for sex, age at death, and principal components 1 through 3. **Results:** We identified 86 proteins for further validation, among which MMP-8 ($P_{meta\text{-}analysis} = 5.6\times10^{-6}$) and GPDA ($P_{meta\text{-}analysis} = 5.2\times10^{-6}$) were the top-ranked proteins. Findings with 24 out of these proteins were validated ($P<0.05$) with clinical diagnosis of AD in the replication dataset, including plasma levels of GPDA ($P = 5.6\times10^{-9}$ and MMP-8 ($P = 5.0\times10^{-3}$). Associations between GWAS and plasma MMP-8 level identified genome-wide significant associations with SNPs in *FOXN3* (rs17125790, $P=4.9\times10^{-9}$), *BNC2* (rs76974599, $P=8.9\times10^{-9}$), *ERC1* (rs150275887, $P=3.2\times10^{-8}$), and *IGF2* (rs4930041, $P=4.6\times10^{-8}$).

Conclusion: We confirmed MMP-8, a neutrophil collagenase and member of the matrix metalloproteinase family, as a potential plasma-based biomarker for pathological and clinical diagnosis of AD. Plasma MMP-8 level may be modulated by variants in *FOXN3*, *BNC2*, *ERC1*, and *IGF2*. Our findings suggest pathways that may modulate a novel plasma biomarker for AD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5145W: Prediction of coronary heart disease in Hispanics/Latinos using state-of-the-art polygenic risk scores: Findings from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL)

Authors:

C. G. Hutten[1,2], F. J. Boehm[1], J. A. Smith[3], B. W. Spitzer[4], S. S. Wassertheil-Smoller[5], C. Isasi[5], J. Cai[6], J. Unkart[7], J. Sun[2], V. W. Persky[2], M. Daviglus[8], M. Argos[9,2]; [1]Div. of Cardiology, Dept. of

Internal Med., Univ. of Michigan Med., Ann Arbor, MI, [2]Dept. of Biostatistics and Epidemiology, Sch. of Publ. Hlth., Univ. of Illinois at Chicago, Chicago, IL, [3]Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [4]CardioVascular Inst., Beth Israel Deaconess Med. Ctr., Boston, MA, [5]Dept. of Epidemiology and Population Hlth., Albert Einstein Coll. of Med., Bronx, NY, [6]Dept. of Biostatistics, UNC at Chapel Hill, Chapel Hill, NC, [7]Dept. of Surgery, SUNY Downstate Med. Ctr., Brooklyn, NY, [8]Inst. for Minority Hlth.Res., Univ. of Illinois at Chicago, Chicago, IL, [9]9. Dept. of Environmental Hlth., Sch. of Publ. Hlth., Boston Univ., Boston, MA

Abstract:

Background: Coronary heart disease (CHD) is a leading cause of death for the Hispanic/Latino population in the United States. We evaluated the association and prediction of 4 polygenic risk scores (PRS) with incident myocardial infarction (MI) in a Hispanic/Latino study sample. **Methods:** We used participant data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). We assessed PRS from the PGS catalog, derived using a few methods (LDpred, AnnoPred, stacked clumping and thresholding, and LDPred2). Cox proportional hazards models were used to evaluate the associations between each PRS and time to adjudicated MI, adjusted for age, sex, first 5 principal components, and accounted for complex survey design. Concordance statistics (c-index) were used to assess predictive accuracy of each PRS compared with, and in addition to traditional risk factors for CHD (obesity, hypercholesterolemia, hypertension, diabetes, and smoking). Analyses were stratified by self-reported Caribbean- (Puerto Rican, Dominican or Cuban heritage) and Mainland- (those of Mexican, Central American, or South American heritage) Hispanic/Latino subgroups.**Results:** Among 9055 participants in the analytic sample (mean age (SD) 47.6(13.1), 62.2% female) the incidence of adjudicated MI was 1.0% through 2019 (n = 95). Each PRS was more strongly associated with MI among Mainland participants. Among the Mainland subgroup, one SD-increase in LDPred2 PRS was associated with a 2.69 higher rate (95% CI 1.71, 4.20) of MI with a c-index of 0.897 (95% CI 0.848, 0.946). This was a modest increase in c-index of traditional risk factors alone (0.880, 95% CI 0.827, 0.933). The AnnoPred PRS had the highest c-index among the Caribbean sample (0.721, 95% CI 0.647, 0.795), however, one SD-increase in PRS was modestly associated with the rate of MI (HR=1.14, 95% CI 0.82, 1.60).**Conclusion:** Performance of PRS for CHD is lacking for Hispanics/Latinos who originate from the Caribbean and have a substantial proportion of African genetic ancestry. This threatens continued widening of health disparities in these populations. A PRS developed using AnnoPred, leveraging functional annotations, performed better than other PRS in this group, suggesting a potential strategy to prioritize PRS construction in diverse populations. These results further highlight the need for additional studies to evaluate and optimize risk prediction of CHD based on cumulative genetic risk in diverse Hispanic/Latino populations.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5146W: Prenatal autoimmune endocrine disorders and risk of autism

Authors:

A. Kepinska[1], B. Mahjani[1,2]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Karolinska Inst.t, Solna, Sweden

Abstract:

Objective: In perinatal autoimmune endocrine disorders, a dysregulated immune system erroneously targets the endocrine system, leading to hormonal imbalances in pregnant individuals. These processes potentially impact offspring perinatal neurodevelopment and may predispose the offspring to autism spectrum disorders (ASD). However, limited studies have accounted for genetic confounding, the impact of genetic factors on both exposures and outcomes. We used family history of autoimmune endocrine disorders (ADs) and ASD, a proxy for genetic factors.
Methods: We used data from the Swedish nationwide registers, which included 481,855 mothers (birthing parents assigned female at birth). Among these, 4100 mothers were diagnosed with an endocrine AD during pregnancy. First, we conducted a nested case-control study that included 15,291 children with an ASD diagnosis and 611,640 controls matched based on birthyear, maternal age at birth, and sex. We used the Cox model to estimate hazard ratios (HRs) with 95% CIs of offspring ASD diagnosis. Then, we conducted a family-based study that included 608,676 full siblings to estimate familial risk of ASD diagnosis if a sibling also has an ASD diagnosis, adjusted for maternal AD diagnosis, maternal age at birth, and sex.
Results: The case-control model revealed statistically significant associations between prenatal insulin-dependent diabetes mellitus (HR = 1.62, 95% CI: 1.30–2.01, $p < 0.001$), Addison's disease (HR = 3.11, 95% CI: 1.00–9.63, $p = 0.0496$), and offspring ASD. Sibling-based analysis confirmed that an ASD diagnosis in one child is a strong predictor of ASD in their sibling (OR range: 8.22-8.23) in models adjusted for maternal ADs. Significant maternal predictors included prenatal hypothyroidism (OR = 2.17, 95% CI: 1.40–3.19, $p < 0.001$) and Addison's disease (OR = 4.69, 95% CI: 1.00-15.10, $p = 0.02$).
Conclusions: The association between maternal endocrine ADs and their offspring's ASD, including among full siblings, suggests that the observed sibling correlation in ASD diagnoses can be partly explained by the influence of maternal ADs, indicating a genetic or environmental pathway mediated through maternal health. Future research could explore genetic components

shared by ASD and multiple endocrine ADs, as well as unique to ASD and each AID, to clarify how different hormonal dysregulations and their timing impact ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5147W: Rare Deleterious ATXN2 Variants in Extreme Adult Obesity

Authors:

S. Almansoori; Imperial Coll. London, London, United Kingdom

Abstract:

Obesity is a growing problem worldwide, highlighting the importance of understanding its genetic underpinning. Ataxin-2 (*ATXN2)* has repeatedly been identified as an important player in metabolic regulation, with deficiency in animal models causing adult-onset obesity, diabetes and altered lipid profiles, with sex-specific differences, including increased female embryonic lethality. To examine the implications of deleterious *ATXN2* variants for human adult-onset obesity, a diabesity-focused custom genotyping array, including 471 rare, predicted-deleterious *ATXN2* variants, was applied to 1714 unrelated UK adults with severe obesity. The UK Biobank (UKB) exome sequence dataset was also analysed for further insights. Eleven rare, predicted-deleterious variants (CADD-PHRED>20) were identified in 20 individuals with severe obesity, 90% of whom were female (vs 75.5% amongst non-carriers). The variants were over-represented compared to gnomAD frequencies. The UKB dataset revealed 83 carriers. Female carriers were proportionally rarer than in the rest of the UKB (43.4% vs. 54.5%, $p = 0.047$) indicating ascertainment or survival bias. Our study provides evidence for the presence of rare, predicted deleterious *ATXN2* variants in severe adult obesity, and suggests potential sex-specific effects, as observed in mouse models. Further studies are needed to investigate the sex bias observed in UKB.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5148W: Rare variants identify novel autism genes and different genetic architecture in simplex and multiplex autism families ⭐

Authors:

L. Qiao[1,2], P. Feliciano[3], J. Hagen[3], S. D. Barns[3], C. Bradley[4], G. Caprara[5], P. Carbone[6], M. Hale[7], E. Neuhaus[8], H. Wiese[9], E. Wodka[10], Y. Shen[2,11,12], W. K. Chung[1,3]; [1]Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, [2]Dept. of Systems Biology, Columbia Univ. Irving Med. Ctr., New York, NY, [3]Dept. of Pediatrics, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [4]Dept. of Pediatrics, Med. Univ. of South Carolina, Charleston, SC, [5]Ctr. for Autism Res., Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Dept. of Pediatrics, Univ. of Utah, Salt Lake City, UT, [7]Dept. of Psychology, Univ. of Miami's Ctr. for Autism and Related Disabilities (UM-CARD), Coral Gables, FL, [8]Dept. of Psychiatry and Behavioral Sci., Univ. of Washington, Seattle, WA, [9]Dept. of Psychiatry and Behavioral Sci., Rush Univ. Med. Ctr., Chicago, IL, [10]Ctr. for Autism Services, Sci. and Innovation, Kennedy Krieger Inst., Baltimore, MD, [11]Dept. of BioMed. Informatics, Columbia Univ. Irving Med. Ctr., New York, NY, [12]JP Sulzberger Columbia Genome Ctr., Columbia Univ. Irving Med. Ctr., New York, NY

Abstract:

Autism is a highly heritable complex condition often associated with other neurodevelopmental phenotypes. Genome and exome sequencing studies have identified ~200 genes associated with an increased probability of autism, most of which contribute through *de novo* coding variants and the implicated genes are strongly associated with cognitive impairment. There is evidence supporting the contribution of rare inherited variants to autism, particularly when transmitted from unaffected mothers. However, the contribution degree of rare inherited variants has been inconsistent due to uncertainty of the effects of most missense variants and insufficient sample size. In this study, we analyzed the largest exome sequencing study of autism with more than 120,000 individuals including 44,000 cases from SPARK and 14,000 autism trios from other published autism cohorts. We used MisFit and GeneBayes for estimating the selection coefficient of missense and loss of function (LoF) variants, calculated weights of variants grouped by selection coefficients, and used these weights in DeNovoWEST for autism gene discovery. We identified 18 putatively novel genes in autism with a false discovery rate (FDR) ≤0.01, eight of which are established neurodevelopmental disorder genes. Most autistic individuals carrying *de novo* damaging variants in *POLR2A* (n = 9/10) reported attention deficit hyperactivity disorder and typical language abilities (n= 8/10). IQ scores were available for five of the 10 *POLR2A* individuals and all were in the average range (IQ > 90). We observed that unaffected individuals have significantly lower *de novo* missense burden in multiplex than simplex families (ratio rate=0.66, $p=7.4\times10^{-3}$) under strong selection, indicating a depletion of *de novo* missense variants under strong selection in unaffected siblings of multiplex families. In our analysis of rare inherited variants, we observed a significant over-transmission of ultra-rare LoF and missense variants under strong and moderate selection to autistic individuals, mainly in simplex families ($p < 7.1\times10^{-8}$). The over-transmission rate is highly correlated with the estimated selection coefficient for both rare missense and LoF variants. The estimated heritability of rare

inherited variants in simplex cases is ~9%, which is much larger than previous estimates. The estimated value for multiplex cases is ~4%, which is likely an under-estimate due to the fact that nearly all multiplex families in the study have unaffected parents. These results provide new insights into the differences in genetic architecture between simplex and multiplex autism families and the role of rare inherited variants in autism.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5149W: Revolutionizing precision health in Type 2 Diabetes: AI integration of multimodal medical imaging and genome-wide genetic data in a large-scale biobank

Authors:

H-C. Yang, Y-J. Huang, C-h. Chen; Academia Sinica, Taipei, Taiwan

Abstract:

Effective risk assessment and prevention improve patients' quality of life and reduce national insurance costs. The convergence of medical imaging and genetic data within expansive biobanks, propelled by artificial intelligence, is reshaping precision health. This research delves into the realm of precision health for Type 2 Diabetes (T2D) by scrutinizing medical images (abdominal ultrasonography and bone density scan images) in conjunction with genetic information (whole-genome single nucleotide variation) to assess T2D risk in a cohort of 17,785 Han Chinese participants from the Taiwan Biobank. Medical imaging data and genetic data are cleaned through a series of quality control and data preprocessing procedures. Densely connected convolutional network deep-learning models were used for imaging analysis. In addition, a graphic neural network was applied to account for the intra-individual dependency of multiple images available for each individual. Bayesian statistical learning models were used for a polygenic risk score (PRS) in genetic analysis. We fused these modalities utilizing the eXtreme Gradient Boosting machine-learning method. We compare the performance based on pixel data of the studied medical images with the image feature variables annotated by medical technologists; our findings from imaging analysis underscore the superiority of pixel-based analysis over feature-centric methods in terms of accuracy, ushering in automated and cost-effective processes. Furthermore, the incorporation of multi-modality analysis enhances accuracy compared to single-modality approaches. These strides, when combined with medical imaging, genetic, and demographic information, open up promising avenues for fusion modeling, encompassing artificial intelligence deep learning, machine learning, and statistical learning, to refine disease risk assessment. The ultimate model yields a compelling Area under

the Receiver Operation Curve of 0.954 (ACC = 0.875, SEN = 0.882, and SPE = 0.875) in accurately assessing the risk of T2D. Moreover, our results show that the odds ratio of T2D and its corresponding confidence interval revealed a positive correlation between the multi-image risk score (MRS) and T2D, as well as between the PRS and T2D. Based on MRS and PRS, we identified specific high-risk subgroups within the study population. In summary, this research represents the first integration of multimodal imaging pixels and genome-wide genetic variation data for T2D risk assessment. This methodology and its results significantly advance our comprehension of precision health in the context of T2D.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5150W: Ribosomal DNA copy number modifies the genetic risk for Type 2 diabetes via dysregulated ribosome biogenesis in pancreatic islets

Authors:

A. Raj, N. Thayer, M. Hotz, J. Brown, K. Wright, I. Lam, N. Fong, W. Kong, D. Hendrickson; Calico Life Sci. LLC, South San Francisco, CA

Abstract:

Ribosomal RNA (rRNA) are the key building blocks for ribosomes, macromolecules responsible for protein synthesis in all living organisms. In humans, the 18S, 28S, and 5.8S subunits of rRNA are encoded as tandem repeats of a single 45S subunit on the short arms of the five acrocentric chromosomes (13, 14, 15, 21, and 22), while the 5S subunit is encoded on chromosome 1. The first complete human reference released by the Telomere-to-Telomere consortium has enabled us to quantify genetic variation in these highly repetitive regions of the genome using short-read whole genome sequencing (WGS) data. We quantified the copy number (CN) of the 5S and 45S subunits using WGS measurements in 490,380 participants in the UK Biobank (UKB). We observed both subunits to be highly heritable, similar to our observations in the 1000 Genomes samples with matched ancestry. Additionally, we observed the 5S subunit to be significantly more heritable than the 45S subunit, consistent with recent observations of high recombination between the short arms of the acrocentric chromosome. We identified 45S CN to be associated with increased incidence of several metabolic, cardiovascular, respiratory, and neurovascular diseases including type 2 diabetes, myocardial infarction, obstructive chronic bronchitis, and cerebrovascular occlusion. In addition, 45S CN was also associated with a higher risk for mortality due to cerebrovascular diseases. In contrast, 5S CN was associated with increased risk for essential hypertension alone. We stratified participants by the T2D risk variant (rs1635852)

at JAZF1, a gene implicated in T2D via dysregulation of ribosome biogenesis, and observed that high 45S CN increases T2D risk specifically in individuals with one copy of the disease-linked allele (T) of the risk variant. Furthermore, using deep proteome measurements in a subset of UKB participants, we observed that 45S CN was positively associated with the plasma levels of proteins that are specifically expressed in the pancreas and goblet cells in the small intestine and trachea. We quantified 5S and 45S CN in 50 donor-derived pancreatic islet microtissues and observed that 45S CN was associated with higher baseline insulin secretion, specifically in females. Additionally, in JAZF1 T/T donors, high CN of both 5S and 45S subunits were linked to high baseline insulin secretion, consistent with our observations in incidence of T2D. Our results highlight the contribution of variation in structural components of ribosomes to variation in complex traits and suggest a novel mechanism by which ribosomal variation modifies genetic risk for complex diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5151W: Sex-specific causal inference assessment of immune cells and protein levels in neurodegenerative diseases

Authors:

F. Lona Durazo[1], R. Byrne[2], R. McLaughlin[2], S. Gagliano Taliun[3]; [1]Montréal Heart Inst., Montreal, QC, Canada, [2]Trinity Coll. Dublin, Dublin, Ireland, [3]Université de Montréal, Montreal, QC, Canada

Abstract:

Sex differences, in terms of prevalence, symptoms and disease progression, are established in the aetiology of complex neurodegenerative diseases, such as in amyotrophic lateral sclerosis (ALS) and Parkinson's disease (PD), but the underlying biology driving these differences remains poorly understood. There is emerging evidence, through genetic and functional analyses, affirming the role of the immune system in such diseases, but assessment of sex differences remains overlooked.

We aimed to evaluate the causal effect of immune-related cell and protein levels on two neurodegenerative diseases with large-scale sex-stratified genome-wide association data available: ALS (N females = 67,957; N males = 65,692) and PD (N females = 104,082; N males = 110,616). To do so, we applied a robust causal inference approach, two-sample Mendelian Randomisation (MR).

First, we carried out sex-stratified MR using summary statistics of immune cell levels (for five types of white blood cells) from the UK Biobank (N females ~ 245,494; N males ~ 206,770). We identified a significant effect of neutrophil count on female-only ALS (Inverse-variance weighted

random effects (IVW) odds ratio (OR) = 0.910; p = 0.044), supporting previous studies. Sensitivity analyses assessing horizontal pleiotropy warrant caution in interpretation.
Next, we used sex-combined plasma protein *cis* quantitative trait loci (*cis*-pQTLs) from a large sample (N = 7,213). We focused on 527 proteins with significant *cis*-pQTLs (FDR < 0.05) from the Gene Ontology biological process 'immune system process' (GO 0002376) to test for a causal relationship between each of these proteins and either ALS or PD in sex-stratified and sex-combined data. After performing sensitivity analyses, we identified a significant effect of FCGR2A and FYN protein levels on sex-combined PD (IVW OR = 1.08; p = 2e-4 and IVW OR = 1.4; p = 0.002, respectively). These findings were supported by colocalization analysis (posterior probability > 0.70). *FCGR2A* is expressed in neutrophils, macrophages and microglia in the brain, and it is involved in initiating cellular responses against pathogens, whereas *FYN* is expressed in T cells and in astrocytes and it is involved in regulating cell growth.
Our results point to a causal effect of increased levels of two immune-related proteins in PD, in a non-sex-specific manner. There were no significant findings for ALS. A possible explanation for the lack of sex-specific signals may be insufficient power due to smaller sample sizes. We are currently testing immune-related proteins derived from a different source, cerebrospinal fluid, to explore the influence of immune-related proteins in the brain.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5152W: Sex-Stratified Single-Cell Analysis Uncovers Gene Expression Differences in Schizophrenia, Bipolar Disorder, and Autism

Authors:

Y. Xia[1], X. Li[1], R. Malik[1], R. Meng[1], D. Clarke[2], P. Emani[1], M. Gerstein[1]; [1]Yale Univ., New Haven, CT, [2]yale Univ., new haven, CT

Abstract:

Background: Understanding the genetic underpinnings of psychiatric disorders, such as schizophrenia (SCZ), bipolar disorder (BD), and autism spectrum disorder (ASD), is critical for developing targeted treatments. Recent advances in single-cell RNA sequencing (snRNA-seq) provide unprecedented resolution to explore cell-type-specific and sex-specific gene expression differences in these disorders.
Methods: In this study, we utilized single-nucleus RNA sequencing (snRNA-seq) data from the PsychENCODE consortium (n=388) to perform a sex-stratified analysis at single-cell resolution. We investigated differentially expressed genes (DEGs) in schizophrenia (SCZ), bipolar disorder (BD), and autism spectrum disorder (ASD) across 17 brain cell types, including 9 excitatory

neuron subtypes, 5 inhibitory neuron subtypes, and 4 non-neuronal cell types (astrocytes, microglia, oligodendrocytes, and oligodendrocyte precursor cells). Post differential analysis, we evaluated the druggability of these genes, their developmental trajectories, and their evolutionary patterns to gain insights into their potential as therapeutic targets. Our approach included the assessment of sex-specific DEGs and their functional roles in disease pathology.
Results: We identified a higher number of significant DEGs in SCZ among males compared to females, with effect sizes of DEGs being larger in females. In sex-specific differential expression tests, cell types L6.CT and L6.IT.Cars exhibited the highest number of significant DEGs. Notably, we discovered several sex-specific DEGs across SCZ, BD, and ASD. For example, *LIN28B*, a gene involved in the regulation of sex hormones, was found to be female-specific in L6.CT excitatory neurons. Additionally, *ROBO2*, a known SCZ risk gene that influences dopamine-producing neurons, displayed significant sex-specific expression patterns.
Conclusion: Our study highlights the importance of sex-stratified analyses in understanding the genetic architecture of psychiatric disorders at single-cell resolution. These findings suggest that targeted therapies could benefit from considering sex-specific gene expression differences, potentially leading to more personalized and effective treatments for SCZ, BD, and ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5153W: Sexually Dimorphic Association of CPS1 with Atherosclerosis in Humans and Mice

Authors:

J. Hartiala[1], J. Hilser[1,2], N. Woodward[1,2], P. Huang[1,2], J. Gukasyan[1,2], Z. Wang[3,4,5], Z. Fouladian[6], S. Charugundla[6], W. Mack[1,7], H. Hodis[1,7,8], K. Reue[9,10], A. Lusis[6,9], L. Ma[11], J. Björkegren[11,12], FinnGen, S. Ripatti[13], A. Palotie[13], W. Tang[3,4,5], S. Hazen[3,5], H. Allayee[1,2]; [1]Dept. of Population & Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Dept. of Biochemistry & Molecular Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [3]Dept. of Cardiovascular & Metabolic Sci., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [4]Dept. of Ctr. for Microbiome and Human Hlth., Cleveland Clinic, Cleveland, OH, [5]Dept. of Cardiovascular Med., Heart, Vascular and Thoracic Inst., Cleveland Clinic, Cleveland, OH, [6]Dept. of Med., Immunology, & Molecular Genetics, David Geffen Sch. of Med. of UCLA, Los Angeles, CA, [7]Atherosclerosis Res. Unit, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [8]Dept. of Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [9]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [10]Molecular Biology Inst., Univ. of California, Los Angeles, CA, [11]Dept. of Genetics & Genomic Sci., Inst. of Genomics and Multiscale Biology, Icahn Sch. of Med. at Mount Sinai, New York, NY, [12]Integrated Cardio

Metabolic Ctr., Dept. of Med., Karolinska Inst.t, Karolinska Univ.ssjukhuset, Huddinge, Sweden, [13]Inst. for Molecular Med. Finland, FIMM, Univ. of Helsinki, Helsinki, Finland

Abstract:

Background: Over 300 susceptibility loci have been identified for coronary artery disease (CAD) but the contribution of genetic factors to CAD risk through sexually dimorphic patterns is not well known. Based on prior metabolomics studies, we evaluated sex-stratified association of genetically decreased activity of *CPS1*, the rate-limiting enzyme of the urea cycle, with atherosclerosis in humans and mice. **Methods:** Meta-analyses were performed to test a *CPS1* variant (rs715) for association with risk of CAD in 652,486 women and 561,291 men separately using data from the UK Biobank (UKBB), GeneBank, and FinnGen cohorts, and summary statistics from BioBank Japan and the CARDIoGRAM+C4D Consortium. Rs715 was independently tested for association with carotid atherosclerosis in 1,430 post-menopausal women from 5 clinical trials. Sex-stratified *CPS1* expression was evaluated in the STARNET cohort (n=600) and 89 inbred mouse strains. *Cps1* deficient (*Cps1*$^{+/}$) and wildtype (WT) mice were characterized for differences in plasma metabolomics and for atherosclerosis by aortic lesion and *en face* analyses. **Results:** A strong sexually dimorphic association (P-heterogeneity=2.3x10^{-5}) was observed between rs715 and decreased risk of CAD in women (OR=0.95, 95% CI 0.93-0.96; P=2.6x10^{-12}) compared to men (OR=0.99, 95% CI 0.98-0.998; P=0.02). Rs715 was also associated with decreased carotid atherosclerosis (beta=-0.046, SE=0.011; P=1.6x10^{-5}) in post-menopausal women, which was not modulated by hormone replacement therapy. Compared to WT, male *Cps1*$^{+}$/$^{-}$ mice had decreased atherosclerotic lesions at the aortic root (452,855±23,550 vs. 343,481±29,752µm^2 /section; p=0.009) and along the entire aorta (6.4±1.2 vs 2.3±0.6%; p=0.01). By comparison, heterozygous *Cps1* deficiency had no effect on aortic lesions in female mice, and there were no sex differences in amino acid or urea cycle metabolite levels between *Cps1*$^{+}$/$^{-}$ and WT mice. *CPS1* mRNA levels in liver were ~20% lower in women than men (828±385 vs 1024±582 TPM; p=4.6x10^{-6}) whereas hepatic *Cps1* expression was ~28% lower in male mice than in females (12.2 ±0.13 vs 12.5±0.09 $\log_2$ units; P=2.7x10^{-51}). However, gonadectomy in either male or female mice did not alter *Cps1* expression, suggesting that the sex-specific phenotypic differences in humans and mice were not regulated by sex hormones. **Conclusions:** Our results provide compelling evidence for the sexually dimorphic contribution of genetically decreased *CPS1* activity to atheroprotective traits in humans and mice. Further studies will be needed to elucidate the underlying biological mechanisms for the opposite and differential effects observed in males and females.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5154W: SGLT2 Inhibitors Attenuate Risk of *APOL1*-mediated Chronic Kidney Disease

Authors:

A. Williams[1], C. Zeng[1], S. Goleva[1], D. Schlueter[1,2], B. Waxse[3], J. Keaton[1], T. Tran[1], H. Mo[1,4], T. Ferrara[1], J. Denny[1,5]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Univ. of Toronto Scarborough, Toronto, ON, Canada, [3]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [4]Cohort Analytics Core, Bethesda, MD, [5]All of Us Res. Program, Bethesda, MD

Abstract:

Apolipoprotein L1 (*APOL1*) risk variants (RV) are a major cause of chronic kidney disease (CKD) and end stage kidney disease (ESKD) in individuals of West African ancestry. Sodium Glucose Cotransporter 2 inhibitors (SGLT2i), a first-line treatment for individuals with diabetes mellitus (DM), have been shown to decrease CKD risk. This study evaluated whether SGLT2i attenuates risk of CKD and ESKD in individuals with *APOL1* RV. We identified 8,702 participants with whole genome sequencing data from the *All of Us* Research Program, who had DM, no prior CKD or ESKD, and who started DM medications from March 2013 to July 2022. Among these participants, 5,897 (41.8%) received metformin, 2,928 (20.7%) insulin, 1,921 (13.6%) sulfonylureas, 1,218 (8.6%) SGLT2i, 1,204 (8.5%) glucagon-like peptide-1 agonists, and 948 (6.7%) dipeptidyl peptidase-4 inhibitors. The cohort prescribed SGLT2i totaled 1,218 (two *APOL1* RV= 79; one *APOL1* RV= 357; zero *APOL1* RV= 782) and not prescribed SGLT2i totaled 7,484 (two *APOL1* RV= 603; one *APOL1* RV= 2,346; zero *APOL1* RV= 4,535). We used electronic health record phenotyping algorithms to define CKD and ESKD. We used Kaplan-Meier estimators and Cox regression models to assess the association between *APOL1* RV and the risk of CKD and ESKD stratified by SGLT2i exposure, adjusting for median HbA_{1c}, sex and age. For survival analyses, time zero was first medication class date and event date was the first CKD or EKSD code. We performed binomial tests to evaluate the probability of developing CKD and ESKD at 3 time points. SGLT2i was associated with a lower risk of CKD in individuals with two *APOL1* RV (HR: 0.34; 95%CI 0.12-0.91; p=0.03) and individuals with zero *APOL1* RV (HR: 0.62; 95%CI 0.46-0.83; p=0.001). Participants with two *APOL1* RV and prescribed SGLT2i were less likely to develop CKD (p=0.04) and trending towards less ESKD (p=0.09) compared to participants prescribed other medications. Notably, no participants with two *APOL1* RV and prescribed SGLT2i developed ESKD. The probability of developing CKD occurred less than expected when participants were prescribed SGLT2i at all time points (p &lt 2.2E-16). The probability of developing ESKD occurred less than expected when participants were prescribed SGLT2i at the 5- (p &lt 2.2E-16) and 7.5-year (p&lt6E-6) time points. SGLT2i attenuated *APOL1*-mediated CKD and ESKD risk. Moreover, the protective effect in *APOL1*-mediated CKD may be

higher than in DM kidney disease. SGLT2i should be considered as a therapy to protect against *APOL1*-mediated kidney disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5155W: Shaping the Genetic Landscape of ATXN2: Insights from a Novel Autosomal Recessive CAG Repeat Expansion at the Spinocerebellar Ataxia Type 2 Locus in a Homozygous Acadian Family

Authors:

J. Saucier[1,2], M. Al-Qadi[2,3], E. P. Allain[4,5,6], P-P. Robichaud[4,5,6], L. Chamard Witkowski[2,4], M. Alvarez[4], A. Marrero[4], N. Crapoulet[4,6], A-M. Dion-Coté[5], **M. Ben Amor**[2,4]; [1]Université de Montréal, Montréal, QC, Canada, [2]Ctr. de formation médicale du Nouveau-Brunswick, Moncton, NB, Canada, [3]Dalhousie Univ., Halifax, NS, Canada, [4]Dr. Georges-L.-Dumont Univ. Hosp. Ctr., Moncton, NB, Canada, [5]Université de Moncton, Moncton, NB, Canada, [6]Atlantic Cancer Res. Inst., Moncton, NB, Canada

Abstract:

Spinocerebellar ataxias (SCA) are a heterogeneous group of cerebellar ataxias transmitted in an autosomal dominant manner. SCA2 is the second most prevalent SCA worldwide, a complex ataxia with a high prevalence of extracerebellar signs, including parkinsonism, dystonia, hyporeflexia and cognitive impairment. The pathogenesis of this autosomal dominant disorder involves a CAG repeat expansion on chromosome 12q23-24, with the fully pathogenic range at 37-39 repetitions. ATXN2 is a powerful modulator of neurological disease; CAG repeat mutations in the non-pathogenic range of SCA2 have been identified as risk factors for Parkinson's disease, Amyotrophic Lateral Sclerosis and Frontotemporal Dementia, while fully penetrant alleles with CAA interruptions may lead to SCA2 with predominant parkinsonism.
We report a novel presentation to *ATXN2* mutations in an Acadian family from New Brunswick, Canada: a homozygous CAG expansion repeat, within the fully penetrant range of SCA2, with autosomal recessive inheritance. Three related patients, including two siblings, were referred to our genetics clinic for an adult-onset neurodegenerative disorder with a variable phenotype. Symptoms included ataxia, dysarthria, dysphagia, aphasia, spasticity, hyperreflexia, atypical parkinsonism, incontinence and signs of frontal lobe dysfunction. All patients had some degree of intellectual disability or learning difficulties during childhood. Whole genome sequencing and subsequent repeat-primed PCR identified a homozygous 39/39 CAG trinucleotide repeat expansion in the *ATXN2* gene, with 4 CAA interruptions, in all three probands. Genetic testing of family members revealed four unaffected carrier parents over the age of 65 and one adult

sibling with heterozygous 39 CAG mutations. Asymptomatic carriers who consented to evaluation (3/5) had normal neurological exams, neuropsychological assessments and cerebral MRI. Considering the age of the carrier parents with mutations in the pathogenic range of SCA2, clinical and radiological signs of the disease were expected. Thus, we describe a novel *ATXN2* mutation at the SCA2 locus, of autosomal recessive inheritance, resulting in a neurodegenerative disorder in the homozygous state.
To further characterize this mutation, we will conduct genetic testing to identify unaffected carriers in the probands' family and region. We aim to determine the prevalence of *ATXN2* mutations in this population and elucidate molecular mechanisms, namely disease-modifier genes and epigenic mutations, that could explain the inheritance mode and clinical features of this Acadian family, which diverges from existing literature.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5156W: Shared genetics of plasma protein level network guides composite risk scores for coronary artery disease and type 1 diabetes

Authors:

T. Westbrook, J. Woerner, Penn Medicine Biobank, Y. Nam, D. Kim; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: Polygenic risk scores (PRS) derived from genome-wide association study (GWAS) summary statistics have been shown to be effective in stratifying individuals' risks for a phenotype. Although single PRS models can successfully categorize individuals with high genetic risk, incorporating additional genetic information is necessary as the full extent of complex disease heritability has not yet been captured. Recent studies have developed composite PRS (cPRS) models that combine polygenic scores for phenotypes or endophenotypes associated with the target trait, demonstrating increased performance in risk stratification. Circulating plasma protein levels are used as disease biomarkers, but research on risk models utilizing their genetic relationships with complex diseases as endophenotypes remains challenging. **Methods:** We developed cPRS models for two polygenic disorders, coronary artery disease (CAD) and type 1 diabetes (T1D), using summary statistics of protein quantitative trait locus (pQTL) mapping for 2,922 plasma proteins obtained from 34,557 European UK Biobank (UKB) participants. To identify which plasma proteins have significant associations with each disease on a genetic basis, we constructed genetic association networks between CAD or T1D and the proteins, where nodes are connected by their shared significant SNPs. Then we used a

personalized PageRank algorithm to assign a proximity score to each node and selected the top ranked proteins (25 for CAD, 18 for T1D) via an elbow plot. Applying PRScs to the pQTL summary statistics for the selected proteins, we developed cPRSs for CAD and T1D using data from European ancestry participants in the UKB without plasma proteomic data (n > 200,000) and in the Penn Medicine BioBank (n = 28,076). **Results:** Among the proteins selected by analyzing pQTLs with the personalized PageRank method, known heart-associated genes *HEG1* and *LDLR* were most strongly associated with CAD occurrence and immune-related gene *MICB* was most strongly associated with T1D in the UKB sample. For both CAD and T1D, the cPRS model that used only pQTL summary statistics had similar prediction performance to the disease PRS, and combining the cPRS and disease PRS models slightly improved their AUCs. **Conclusion:** These results demonstrate that pQTL summary statistics can enhance risk prediction for complex diseases. Future work will involve improving the selection of proteins used in the cPRS. Expanding this approach to include diverse populations and other omics data could further enhance the accuracy and applicability of cPRS models for many complex diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5157W: Single-cell multiomics study in hiPSC-derived neurons from large cohorts reveals novel gene regulatory mechanisms of schizophrenia ⭐

Authors:

S. Zhang[1,2], C. Li[1,3], H. Zhang[1], L. Liang[4], Z. Wang[4], A. Kozlova[1], Z. P. Pang[5], A. Sanders[1], X. He[4], J. Duan[1,2]; [1]North Shore Univ Hlth.System, Evanston, IL, [2]Pritzker Sch. of Med., The Univ. of Chicago, Chicago, IL, [3]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA, [4]Dept. of Human Genetics, The Univ. of Chicago, Chicago, IL, [5]Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ

Abstract:

Genome-wide association studies (GWAS) have identified many risk genes for neuropsychiatric disorders (NPD), such as schizophrenia (SCZ). However, functional interpretation of these GWAS findings remains challenging; a major hurdle is that polygenic risk manifests its effect conditionally upon neural activation, i.e., context-dependent. Human induced pluripotent stem cell (hiPSC)-derived neurons are a tractable cellular model for ascertaining context-dependent polygenic effects, e.g., neural activation that may mimic the physiological effects of environmental stimuli. Here, we modelled neural activation by depolarisation using high

potassium chloride (KCl) in co-cultured excitatory/inhibitory neurons of 100 hiPSC lines (28 from SCZ donors), followed by assaying single-cell multiomes (scRNA/ATAC-seq) of over 700,000 neurons at 0hr (unstimulated), 1hr (early response), and 6hr (later response) of high KCl exposure. We linked genes with open chromatin peaks for each main neural subtype (GABAergic inhibitory, *NEFM*+ or *NEFM*-excitatory neurons) and confirmed a significant correlation between peak accessibility and target gene expression. Compared to static genes/peaks, dynamic ones, specifically those activity-upregulated, were more enriched for NPD GWAS risk, with the most robust enrichment for SCZ. Based on the dynamic gene expression pattern across the three time points, we found NPD risk genes tend to be continuously upregulated (i.e., with up-up dynamics). We further analysed context-specific SCZ-associated differentially expressed genes (DEGs) using MAST for single neurons of 28 SCZ cases and 72 controls. We found that 3-5%, 4-6%, and 15-23% of genes were SCZ-associated DEGs in *NEFM*+ excitatory, *NEFM*- excitatory, and GABAergic neurons, respectively, most of which were cell activity-specific. Notably, we found that SCZ-associated DEGs in high-activity neurons were more enriched for synapse-related GO terms and NPD risk genes. Interestingly, upstream regulatory sequences of the stimulation-specific SCZ-associated DEGs were enriched for binding sites of TCF4, which is a strong GWAS risk gene and was considered the “master regulator” of other SCZ-related ones. Moreover, the upregulated genes in SCZ cases in the stimulated (only at 6hr) *NEFM*+ excitatory neurons showed strong enrichment for GO terms related to cholesterol synthesis. Single-neuron DEG analysis for 28 SCZ cases and 28 matched controls gave similar results. Our study suggests that many NPD genes may only elicit disease-relevant effects upon neuronal activation, providing novel insights into how polygenic risk factors interact with environmental stimuli for NPD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5158W: Single-nucleus multimodal dissection of Alzheimer's Disease identifies brain region-specific dysregulation of the regulome.

Authors:

P. Dong, J. Bendl, C. Casey, A. Hong, M. Alvia, Z. Shao, J. K. Massaguer, V. Haroutunian, D. Lee, G. Hoffman, J. Fullard, P. Roussos; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Alzheimer's disease (AD) is a highly heritable and pathologically heterogeneous neurodegenerative disorder, posing one of the most significant societal and biomedical challenges of our time. Neurological and imaging studies have identified divergent pathological

progressions, including the deposition of amyloid-β (Aβ) plaques in the neocortex and tau tangles in the limbic area. Understanding the cellular and molecular dynamics of AD progression across these regions is critical for uncovering the underlying mechanisms and developing effective treatments. To address this, we conducted multimodal single-nucleus (sn) profiling to capture gene expression and chromatin accessibility within the same cells across three brain regions crucial for AD progression: the entorhinal cortex (EC), superior temporal gyrus (STG), and dorsolateral prefrontal cortex (PFC). Our study utilized deeply phenotyped specimens from 150 AD cases and 50 age-matched controls, covering the full spectrum of AD pathology stages. After rigorous quality control, our analysis yielded over 2 million nuclei representing 27 cell types and 82 cell subtypes. This comprehensive resource enabled us to model cellular and molecular variations across brain regions during AD progression. We identified brain region-specific vulnerabilities in somatostatin inhibitory neurons, layer 2/3 and layer 4 intratelencephalic excitatory neurons, and increased abundance of astrocytes and multiple vascular cells. The EC exhibited the largest scale of alterations in both cellular abundance and molecular activity. We also identified AD-risk candidate genes, such as APOE in astrocytes, that were selectively dysregulated across brain regions. Furthermore, we constructed gene regulatory networks, revealing differential gene regulatory programs in response to AD across the brain regions. In conclusion, we present an integrative cross-brain region multimodal map of AD progression in the human brain at single-cell resolution, highlighting brain region-specific cellular and gene regulatory programs in AD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5159W: SLC30A8 rare variant modify contribution of common genetic and lifestyle factors toward type 2 diabetes

Authors:

B-J. Kim, H-M. Jang, Y. Park, Y. Kim; Korea NIH, Chungwon-gun, Korea, Republic of

Abstract:

To explore the modifying effect of rare genetic variants on the risk of type 2 diabetes (T2D) in the context of common genetic and lifestyle factors, we conducted a comprehensive analysis of genetic and lifestyle factors associated with T2D in a cohort of 146,284 Korean individuals. Among them, 4,603 individuals developed T2D during the follow-up period of up to 18 years. We calculated the polygenic risk score (PRS) for T2D based on an independent genome-wide association study on T2D. A rare protective variant, I349F at *SLC30A8,* was validated in 146K Koreans and individuals carrying them were identified. Additionally, we assessed the Healthy

Lifestyle Score (HLS) by combining various indicators of a healthy lifestyle, including physical activity, obesity, smoking status, healthy diet, and sodium-intake status. Cox proportional hazards model was used to analyze the association between genetic components (PRS and I349F) and HLS in incident cases of T2D. Furthermore, we examined the combinations of PRS-I349F, PRS-HLS, and HLS-I349F in relation to incident T2D cases. A high PRS (top 20%) and an unfavorable lifestyle were associated with an increased risk of incident T2D. Individuals with an unfavorable lifestyle and the top 20% T2D-PRS had the highest risk of incident T2D. However, these established risk factors were significantly influenced by the presence of a rare allele, I349F. Rare-allele carriers had a T2D prevalence of 5.4%, whereas non-carriers had a prevalence of 11.7%. A rare allele modified the effect of T2D -PRS by reducing the prevalence of T2D from 23.18% to 12.70% in individuals carrying rare protective allele in the top quintile of T2D-PRS. Furthermore, rare-allele carriers showed a reduced risk of prevalent T2D among the intermediate and unfavorable HLS groups compared to the favorable group. Rare-allele carriers exhibited a similar trend of reduced risk of incident T2D cases. Considering the combined influence of common and rare genetic variants along with lifestyle factors improved the predictability of T2D within the Korean population. Our findings highlight the significance of rare genetic variations in both PRS and HLS in the context of the development of T2D. The conventional use of common genetics-based PRS and HLS is not effective in predicting T2D among carriers of a rare variant. Considering the accumulation of rare variant information from recent sequencing efforts, these findings have important implications for personalized prevention and intervention strategies as well as subtyping of individuals at risk of developing T2D.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5160W: *SOD1* Related Amyotrophic lateral sclerosis Progression A Case Report

Authors:

A. Rivera-Cameras[1,2], M. Gallegos-Arreola[3,4], O. Padilla - de la Torre[5], A. Alcantar-Aranda[6], I. Dávalos-Rodríguez[3,4]; [1]Departamento Ciclo de Vida, Genética y Med. Genómica, Unidad Académica de Cs. de la Salud, Univ. Autónoma de Guadalajara, Guadalajara, Mexico, [2]División de Genética, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, México., Guadalajara, Mexico, [3]División de Genética, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, Mexico, [4]Doctorado en Genética Humana, CUCS, Univ. de Guadalajara, Guadalajara, Mexico, [5]Departamento de Neurología, UMAE Hosp. de Especialidades, Centro

Médico Natl. de Occidente, IMSS, Guadalajara, Mexico, [6]Alta Especialidad Neurogenética, Univ. Autónoma de México,, Mexico, Mexico

Abstract:

INTRODUCTION: Amyotrophic lateral sclerosis (ALS), OMIM #147450, is a progressive neurodegenerative disorder characterized by motor neuron degeneration leading to weakness, fasciculations, and bulbar symptoms. Variants in *SOD1* have been described since 1993 and are the second most common cause of ALS. More than 200 pathogenic variants in this gene have been reported, with approximately 20% of familial cases and 2% of sporadic cases associated with *SOD1* mutations. **OBJECTIVE:** To describe the molecular correlation between *SOD1* and clinical characteristics in ALS. **CASE REPORT**: A 48-year-old man presented with initial symptoms in 2019, reporting calf cramps in his right leg. Over the year, he experienced progressive weakness in the right pelvic limb, accompanied by muscle atrophy and difficulty walking and climbing stairs. Subsequently, similar symptoms appeared in the left leg, indicating bilateral progression of weakness. As his condition worsened, the patient became increasingly dependent on walking aids. The diagnostic complexity of ALS, the atypical clinical presentation initially raised of muscular dystrophy due to the absence of bulbar symptoms or fasciculations. Genetic studies confirmed the diagnosis, revealing a variant (rs121912443) [c.140A>G p.(His47Arg)] in *SOD1*. This molecular knowledge illuminates ALS's underlying pathophysiology and its correlation with clinical symptoms. **CONCLUSIONS:** For accurate diagnosis and effective management, it's essential to include ALS in the list of potential conditions when patients present with unusual symptoms and muscle weakness. Integration of genetic studies, alongside collaboration with a multidisciplinary team, is key.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5161W: Spatial Multi-omic architecture of Alzheimer’s disease : Lessons from Human and Mouse model systems

Authors:

V. Swarup[1], E. Miyoshi[1], S. Morabito[2], E. Head[1], K. Green[1]; [1]Univ. of California Irvine, Irvine, CA, [2]Univ. of California, Irvine, Irvine, CA

Abstract:

The molecular origins of Alzheimer's disease (AD), shaped by a complex interplay of environmental and genetic determinants, remain poorly understood. Our investigation employs spatial transcriptomics (ST) and spatial proteomics alongside single-nucleus RNA sequencing

(snRNA-seq) to scrutinize the transcriptomic environments of both late-onset sporadic AD and AD related to Down Syndrome (AD in DS). Analyzing AD in DS provides a novel angle that could help reconcile discrepancies between conventional genetic mouse models and sporadic forms of AD. Our findings delineate marked transcriptomic discrepancies that could underlie the distinct pathologies observed predominantly in specific cortical layers. We identified disease processes that are both spatially localized and temporally transient through the application of spatial co-expression network analysis, revealing a dysregulated glial inflammatory gene module in the upper cortical strata associated with genetic susceptibility factors and amyloid-linked pathways in AD. The study further examines disruptions in cell-cell interactions, highlighting the role of particular genes within this inflammatory module in disturbed signaling networks. Additionally, we have generated ST data from an amyloid-centric AD mouse model, enabling a cross-species analytical approach to identify amyloid-related transcriptomic alterations within a defined conformational framework. This research highlights the critical role of incorporating both cellular and spatial resolutions in AD studies, uncovering complex transcriptomic variations across different stages and manifestations of the disease, which are essential for decoding AD pathogenesis and advancing therapeutic strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5162W: Spectrum of rare genetic variation in etiology of non-syndromic autism spectrum disorder

Authors:

H. Sheth[1], J. Shah[1], D. Jain[2], S. Shah[3], H. Patel[4], K. Patel[5], D. Solanki[6], A. Iyer[7], P. Mahtre[8], S. Mehta[3], S. Bajaj[9], D. Patel[10], D. Mondal[11], J. Sheth[1], S. Ramdas[11], F. Sheth[1]; [1]FRIGE's Inst. of Human Genetics, Ahmedabad, India, [2]Shishu Child Dev. And Early Intervention Ctr., Ahmedabad, India, [3]Royal Inst. of Child NeuroSci.s, Ahmedabad, India, [4]Zydus Hosp., Ahmedabad, India, [5]Speciality Homeopathy Clinic, Ahmedabad, India, [6]Mantra Child Neurology and Epilepsy Hosp., Bhavnagar, India, [7]NeuroKids, Ahmedabad, India, [8]Tender Kids Ctr. for Child Dev., Navi Mumbai, India, [9]The Purple Gene Clinic, Mumbai, India, [10]Charotar Inst. of ParaMed. Sci., Changa, India, [11]Indian Inst. of Sci., Bengaluru, India

Abstract:

Autism spectrum disorder (ASD) is a multifactorial neurodevelopmental disorder characterized by impairments in social communication and repetitive and restricted behaviors. Although ASD is defined behaviorally rather than genetically, contribution of genetic variants in its etiology can be identified. Genetic variants ranging from extremely rare to common, including *de novo* and

inherited variants, impact genes in a manner that together or on their own can result in an ASD phenotype. Implementation of high-throughput short and long read sequencing technologies provides opportunities to assess genome wide patterns of rare genetic variants in the etiology of ASD. We present here rare putative causative genetic variants in etiology of ASD that are discovered by Illumina whole exome sequencing (WES) and Oxford Nanopore long-read whole genome sequencing (WGS) in a cohort of 101 patient-parent trios. Of 101 trios, 30 trios (29.7%) received genetic diagnosis of ASD through detection of pathogenic or likely pathogenic (P/LP) single nucleotide variants (SNVs; n=27/30, 90%) and copy number variants (CNVs; n=3/30, 10%) using WES. 66.6% (n=3 for CNVs and n=17 for SNVs) of these cases had *de novo* variants. *De novo* SNVs were found primarily in genes previously associated with ASD (n=16/17; 94%) with the *MECP2* gene being the most recurrently mutated (n=6/17; 35%). Additionally, WES identified 22 variants of uncertain significance (n=21/101 trios; 20.8%). Of this, a homozygous frameshift variant in the *NEUROG1* gene led to a novel genotype-phenotype correlation with cranial dysinnervation disorder along with ASD phenotype (OMIM#620469). Of 101 trios, 23 trios without a confirmed genetic diagnosis through WES gave consent for WGS. WGS was conducted at a mean coverage of 10x and structural variants (SVs) were detected by aligning reads against the hg38 reference genome using minimap2 and using consensus SV calls from 5 SV callers- cuteSV, sniffles2, SVIM, nanovar and npInv. *A priori* benchmarking using the GIAB H002 dataset showed ~79%, 79% and 80% values for precision, recall and F1 score statistics for SV calls at 10x coverage, respectively, with lowest values for SVIM. On average, ~200,000 SVs were called per sample by 5 callers, which gave ~200 consensus filtered SV calls (SVs overlapping a protein coding gene with minor allele frequency of ≤1% in the global population datasets). No putative P/LP SVs associated with ASD phenotype were detected in 23 trios suggesting limited role of rare SVs in the etiology of ASD. Results from our study suggests *de novo* SNVs and CNVs are the primary drivers of genetic etiology of ASD, thereby supporting WES as a first tier clinical genetic test.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5163W: Targeting Dementia Subtypes using Comparative Genomics and Drug Repositioning

Authors:

Z. Shameer[1], A. Titus[2], S. Moidhunny[3], K. K. Yadav[4]; [1]Urbana High Sch., Urbana, MD, [2]Dept. of Preventive Cardiology, Houston Methodist, Houston, TX, [3]Univ. Of Miami, Miller Sch. of Med., Miami, FL, [4]Texas A&M Univ., Houston, TX

Abstract:

Over 134 million Americans aged 65 or older are affected by dementia each year, with four clinical subtypes: vascular, frontotemporal, pre-senile, and lewy body dementia. Identifying shared and unique genes, pathways, and biological network modules across disease subtypes can improve screening, diagnosis, and treatment options. In this research, we developed a computational pipeline analysis to understand the molecular landscape of the different subtypes of dementia to identify unique and shared genetic pathways and driver genes. Disease-gene associations were compiled using the DiSGeNet database. Subtype lists were used to perform biological functional enrichment analysis using the KEGG pathway database. Data was compiled and statistical analyses were performed using Excel, and visualized using Venny. Based on extensive comparison across all the four subtypes, we have found a core set of 25 genes driving involved in all subtypes includes granulin precursor (GRN) involved in neuronal development, Apolipoprotein E (APOE) - a key gene involved in neurodegeneration and Superoxide dismutase [Cu-Zn] (SOD1), an antioxidant enzyme involved in reducing cellular toxicity driven by reactive oxygen species - a key component of neurodegenerative diseases. Furthermore, 90 shared pathways that include insulin resistance, glioma, and hepatitis B. Drug repositioning revealed associations with known drugs targeting neurodegenerative diseases like donepezil and econazole as well as novel associations with antifungals (econazole and butenafine) and immuno-modulatory agent (mepacrine). Further exploration of the data could help develop genomics-oriented preventive treatments for dementia, while also providing an avenue for drug repositioning once major pathways are found.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5164W: The DNA of distress: Uncovering genomic associations of allostatic load

Authors:

C. Okeh, Z. Rodriguez, L. Guare, Penn Medicine BioBank, Regeneron Genetics Center, S. Setia-Verma; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Allostatic load (AL) quantifies the physiological burden or "wear and tear" on the body, reflecting cumulative stress across various biological systems. Traditionally, there has been no consensus on the biomarkers that best represent AL, resulting in variable measures that depend heavily on the specific population studied and the data available. This study introduces a standardized and robust method for calculating AL using electronic health record (EHR) data, which facilitates a more uniform application across different populations. We selected 16

biomarkers indicative of stress-related physiological activity within the cardiovascular, immune, metabolic, and other systems. Each biomarker was measured through outpatient lab tests, with all recorded instances considered. We computed an average for each biomarker per individual, which was then standardized against a distribution derived from our study population. This allowed us to categorize the biomarkers into high-risk or low-risk based on their deviation from the norm, subsequently synthesizing these into a composite AL score for each patient. Further, we conducted a genome-wide association study (GWAS) stratified by ancestry (European and African) to identify genetic variants associated with AL. Through our GWAS, we identified 11 significant loci in the African ancestry analysis and 9 significant loci in the European ancestry analysis. Meta-analysis of ancestry stratified GWAS revealed one locus at genome-wide significance mapping to the gene *UBE2Q1*, which has known implications in protein ubiquitination pathways—critical for cellular stress responses and protein homeostasis. *UBE2Q1* has been previously associated with immunological traits such as white blood cell count, coordinating lysosomal ubiquitination and cell repair. The immune system of the body holds biomarkers that are reliable signifiers of wear and tear within the body. The identification of *UBE2Q1* as a locus influencing AL underscores the genetic basis of physiological stress responses and highlights the molecular pathways involved in stress adaptation and maladaptation. This insight extends our understanding of how genetic factors contribute to the variability in allostatic load among individuals. It opens new avenues for research into preventative and therapeutic strategies aimed at mitigating the adverse health effects associated with high allostatic load. This study not only establishes a replicable method for AL calculation using EHR data but also advances our genetic understanding of physiological stress, which could lead to personalized medicine approaches in managing stress-related disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5165W: The genetic determinants and genomic consequences of non-leukemogenic somatic point mutations ⭐

Authors:

J. Weinstock[1], S. Chaudhry[2], M. Ioannou[3], M. Viskadourou[2], P. Reventun[2], Y. Jakubek[4], P. Natarajan[5], S. Jaiswal[6], A. Bick[7], W. Post[2], P. Scheet[8], P. Auer[9], T. Karantanos[3], A. Battle[1], M. Arvanitis[2]; [1]Dept. of BioMed. Engineering, Johns Hopkins Univ., Baltimore, MD, [2]Dept. of Med., Cardiology Div., Johns Hopkins Univ., Baltimore, MD, [3]Div. of Hematological Malignancies, Dept. of Oncology, Sidney Kimmel Comprehensive Cancer Ctr., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [4]Univ. of Kentucky, Lexington, KY, [5]Ctr. for Genomic Med. and Cardiovascular Res. Ctr., Massachusetts Gen. Hosp., Boston, MA, [6]Dept. of Pathology, Stanford Univ., Stanford,

CA, [7]Vanderbilt Univ. Med. Ctr., Nashville, TN, [8]Dept. of Epidemiology, Univ. of Texas M.D. Anderson Cancer Ctr., Houston, TX, [9]Div. of Biostatistics, Data Sci. Inst., and Cancer Ctr., Med. Coll. of Wisconsin, Milwaukee, WI

Abstract:

Clonal hematopoiesis (CH) is defined by the expansion of a lineage of genetically identical cells in blood. Genetic lesions that confer a fitness advantage, such as point mutations or mosaic chromosomal alterations (mCAs) in genes associated with hematologic malignancy, are frequent mediators of CH. However, recent analyses of both single cell-derived colonies of hematopoietic cells and population sequencing cohorts have revealed that CH frequently occurs in the absence of previously implicated driver genetic lesions. To characterize CH without known driver genetic lesions, we used 51,399 whole genomes from the NHLBI TOPMed consortium to perform simultaneous germline and somatic mutation analyses among individuals without leukemogenic point mutations (LPM), which we term CH-LPMneg. We quantified CH by estimating the total mutation burden. Because estimating somatic mutation burden without a paired-tissue sample is challenging, we developed a novel statistical method, the Genomic and Epigenomic informed Mutation (GEM) rate, that uses external genomic and epigenomic annotations to help distinguish artifactual signals from true somatic mutations. We observed that chromatin state annotations from CD34+ cells were especially useful in distinguishing true mutations from artifacts. We performed a genome-wide association study (GWAS) of GEM to discover the germline determinants of mutation burden in blood. After fine-mapping and variant-to-gene analyses, we identified seven genes associated with CH-LPMneg (TCL1A, TERT, SMC4, NRIP1, PRDM16, MSRA, SCARB1), and one locus associated with a sex-associated mutation pathway (SRGAP2C). We performed a secondary analysis excluding individuals with mCAs, finding that the genetic architecture was largely unaffected by mCA inclusion. Functional analyses of SMC4 and NRIP1 implicate altered HSC self-renewal and proliferation as the primary mediator of mutation burden in blood. We then performed multi-tissue transcriptomic analyses, finding that expression levels of 404 genes are associated with GEM. Finally, we performed phenotypic association meta-analyses across four cohorts, finding that GEM is associated with increased white blood cell count and increased risk for incident peripheral artery disease (hazard ratio = 1.15, pvalue = 3 x 10-3), but is not significantly associated with incident stroke (hazard ratio = 1.03) or coronary disease events (hazard ratio = 1.00). Overall, GEM enables an improved quantification of mutation burden from WGS, and discovery of genetic, genomic, and phenotypic correlates of CH-LPMneg.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5166W: The impact of lifestyle intervention on cardiovascular and kidney outcomes in genetically high-risk obese individuals ⭐

Authors:

M. Shivakumar[1], Y. Kim[1], J. Woerner[1], S-H. JUNG[1], S. Jeong[2], D. Kim[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univeristy of Pennsylvania, Philadelphia, PA

Abstract:

Individuals with a high genetic high-risk of obesity are known to develop cardiometabolic diseases at higher rates and experience worse clinical outcomes. Modifiable lifestyle factors, such as physical activity, may offer potential in controlling the development and progression of these diseases. However, the effectiveness of managing these factors to reduce the occurrence of diseases and improve cardiovascular outcomes remains unclear. In this study, we investigated the effect of lifestyle modulation on the development of kidney diseases and cardiometabolic outcomes.

From the UK Biobank (UKB), we identified patients at a >75% polygenic risk percentile using body mass index (BMI) polygenic risk scores (PRS) (N = 101,954, European). For these patients we further extracted cohorts of coronary heart disease (CHD), chronic kidney disease (CKD), renal failure, diabetes, and acute kidney damage. We generated respective datasets by omitting samples with pre-existing conditions at baseline. We subsequently utilized follow-up data to define lifestyle factors, namely moderate activity, vigorous activity, and smoking. We compared health outcomes in individuals who improved their lifestyle factors (e.g. though more physical activity) to those with unchanged lifestyle factors. We then ran association with diseases as outcomes and lifestyle factor as predictors, adjusting for age and sex to evaluate whether positive changes in lifestyle factors are associated with disease outcomes.

Alterations in physical activity levels after baseline demonstrated a significant influence on disease incidence. Patients who were initially inactive but began engaging in vigorous physical activity were observed to have decreased risk of diabetes (P = 0.004, OR = 0.71), CKD (P = 0.038, OR = 0.8), and acute kidney injury (P = 0.032, OR = 0.6). Similarly, maintaining a moderate level of activity post-baseline was shown to reduce the risk of CKD (P = 0.027, OR = 0.81), and acute kidney injury (P = 0.0054, OR = 0.61). Additionally, diabetic and hypertensive patients incorporating vigorous activity into their lifestyles had significantly lower CHD risks (P = 0.014, 0.04, OR = 0.48,0.83), and hypertensive patients that included moderate activity experienced significantly lower CKD risk (P = 0.008, OR = 0.8).

In our study, we found that changes in lifestyle factors can indeed mitigate the risk of both kidney and cardiometabolic diseases among individuals with high BMI PRS. This emphasizes the

transformative role of lifestyle modifications in preventative health strategies for those genetically predisposed to obesity.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5168W: Genome-Wide Association Study Identifies Variants Associated with Neural Pathway Morphology

Authors:

A. Barrows[1], L. Shao[2], B. Xu[1], R. Loughnan[3], J. Ahern[2], W. Thompson[1], C. Fan[1], F-C. Yeh[4], H. Zheng[1]; [1]Laureate Inst. for Brain Res., Tulsa, OK, [2]Univ. of California, San Diego, San Diego, CA, [3]Univ. of California, San Diego, Ventura, CA, [4]Univ. of Pittsbrugh, Pittsburgh, PA

Abstract:

Introduction: Individual variations in white matter (WM) tracts are linked to a range of cognitive and neuropsychiatric traits. While previous genome-wide association studies (GWAS) have predominantly focused on WM microstructure, the genetic determinants of WM tract morphology remain largely unexplored. This study investigated the genetic architecture underlying the tractography shape metric, specifically the surface area of WM fiber tracts, a critical factor in understanding their role in behavioral and psychiatric phenotypes.
Methods: We computed the total surface area for 80 white matter tracts using deterministic tractography on diffusion-weighted imaging data from the UK Biobank, encompassing a sample of 48,244 individuals. We analyzed the genetic architecture of principal components of the estimated area of major fiber tract bundles that explain 50% of the phenotypic variance. This was followed by genome-wide association studies (GWAS) and post-GWAS analyses, including SNP heritability estimation and genetic correlation analyses with diffusivity metrics and 11 psychiatric disorders.
Results: Our GWAS identified 17 genome-wide significant loci ($P < 2.3 * 10^{-8}$), including a novel locus on chromosome 17 that is linked to WNT3 (lead SNP rs199526; $p = 3*10^{-20}$), which is critical for neural pathway development and maintenance and the previously identified KANSL1 (lead SNP rs3890609; $p = 7 *10^{-9}$), which is implicated in white matter microstructure diffusivities. The SNP heritability of WM tract surface area ranged from 6% to 38%, indicating a significant contribution of common genetic variants. Genetic correlations with classical diffusion metrics (e.g., fractional anisotropy and mean diffusivity) were modest, ranging from -0.197 to 0.292, except for the first principal component (beta= 0.30728), suggesting distinct genetic contributions to WM surface area not captured by traditional diffusion measures. However, correlations with 11 psychiatric disorders were limited, indicating in the polygenic level, the

shape of the fiber bundles are more relevant to the typical development process.
Conclusion: Our study reveals a distinct genetic architecture for the surface area of white matter tracts, separate from those affecting WM microstructure. These findings provide new insights into the genetic underpinnings of brain morphology. Understanding these genetic influences is crucial for unraveling the biological pathways linking brain structure to behavior and mental health.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5169W: Tractor Analysis of African American Inflammatory Bowel Disease Identifies Ancestry Specific Associations

Authors:

S. Buyske[1], C. Simpson[2], A. Choi[3], S. R. Brant[3]; [1]Rutgers Univ., Piscataway, NJ, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [3]Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ

Abstract:

Introduction: Inflammatory Bowel Disease (IBD) is an immune mediated inflammatory disease of the gastrointestinal tract. In European and East Asian populations, GWAS meta-analyses have established over 300 genetic loci. However, in African Americans, our GWAS (Brant et al 2017) and subsequent WGS study (Somineni et al 2021) altogether identified 3 novel loci and replicated 41 loci. Due to African (AFR) and European (EUR) admixture in African Americans, a haplotype signal from one ancestry tagging a risk allele may be confounded by a similar haplotype from the other ancestry tagging a protective allele. We therefore re-used our GWAS data using Tractor (Atkinson et al 2021) to analyze the AFR and EUR ancestral genomes separately and to identify potential ancestry-specific novel loci.
Methods: The sample consisted of 923 IBD cases and 3063 controls genotyped on the Axiom platform and 1095 cases and 1552 controls on the Omni platform (totaling 2018 cases and 3615 controls), all African Americans. After phasing, we inferred local ancestry using FLARE (Browning et al 2023) and used Tractor separately by platform to obtain ancestry-specific effect sizes and p-values, which were imputed to non-genotyped loci using RAISS (Julienne et al 2019) and then meta-analyzed. We considered p-values below 10-5 to be suggestive. For known loci, p < 10-4 was evidence of replication within the established locus interval. Note that the effective sample size for the AFR background is about four times that for the EUR background.
Results: No results were genome-wide significant. Novel suggestive loci were detected in AFR ancestry near *SORBS2* (Chr 4), *FRMD1* (Chr 6), *IL6* (Chr 7), and *OXGR1* (Chr 13), and for EUR ancestry near *RASGEF1B* (Chr 4). For AFR ancestry, 18 established loci replicated, 7 meeting

suggestive evidence (including *PTGER4*, *IL12B* but with peak at *EBF1*, *HLA-DQB1*, and *ADO*), including some at distal locus margins (e.g. *CCR9*, 508Kb from *CCR3*). In EUR ancestry 6 loci replicated (including *ATG16L1*, *ETS1*, *IL26*, and *DOCK9*), 1 with suggestive evidence (at *IL12B*). *IL12B*, *ETS1* and *ATG16L1* replicated in both ancestries.
Discussion: This first examination of IBD loci separately for the African and European ancestral genomes further confirmed evidence for previously established African American loci at *HLA* and *PTGER4 (*the latter showing evidence only in the AFR genome), replicated for the first time key loci in IBD (e.g *ATG16L1* and *SP140*), potentially focused associations (e.g. *CCR9*), and detected provocative novel loci.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5170W: Transcriptomic and proteomic analysis of isogenic *ARC* gene deficiency SH-SY5Y cells

Authors:

M-C. Cheng; Taipei Veterans Gen. Hosp. Yuli Branch, Hualient Country, Taiwan

Abstract:

Objective: ARC (activity regulated cytoskeletal associated protein) is a critical player in multiple neuronal signaling pathways, interacting with the N-methyl-D-aspartate receptor (NMDAR) complex. Its dysregulation can significantly impact learning, memory, and normal brain functions. However, its role in psychiatric disorders, especially schizophrenia, remains largely unexplored, making it a promising area for investigation. This study aims to delve into the mechanism underlying the effects of the *ARC* gene in SH-SY5Y cells, offering a novel perspective on the pathogenesis of schizophrenia.**Methods:** To investigate the function of the *ARC* gene in vitro, we generated *ARC*-deficiency SH-SY5Y cell lines via CRISPR/Cas9-mediated gene editing and conducted RNA sequencing and label-free LC-MS/MS analysis to identify the differentially expressed genes and proteins in isogenic *ARC*-deficiency SH-SY5Y cells. **Results:** Isogenic SH-SY5Y cell lines with *ARC* mutation were generated using CRISPR/Cas9 genome editing. After single-cell isolation, three clones ($ARC^{c.-30\text{-}c.6delTTTGCCGCCTGCCGGAGCACCTGCGCACAGATGGAG/WT}$, $ARC^{c.-24\text{-}c.-3delGCCTGCCGGAGCACCTGCGCAC}$, and $ARC^{c.-15\text{-}c.1delGCACCTGCGCACAGA/WT}$) were identified. The two clones had heterozygous mutations ($ARC^{c.-30\text{-}c.6delTTTGCCGCCTGCCGGAGCACCTGCGCACAGATGGAG/WT}$ and $ARC^{c.-15\text{-}c.1delGCACCTGCGCACAGA/WT}$), while the clone had a homozygous mutation ($ARC^{c.-24\text{-}c.-3delGCCTGCCGGAGCACCTGCGCAC}$) in SH-SY5Y cells. The homozygous mutation clone was confirmed to cause downregulation of the ARC protein in SH-SY5Y cells; thus, it was used for subsequent RNA sequencing and label-free LC-MS/MS analysis. A combination of RNA sequencing and label-free

LC-MS/MS analysis revealed that the *ARC* gene deficiency disturbed multiple genes, including *ANXA2*, *AHNAK*, *FLNA*, *FLNB*, *CNN2*, *MYH9*, *S100A11*, which were involved in remodeling the cytoskeleton, maintaining cell shape, neuroplasticity, inflammation, and apoptosis. **Conclusion:** Our preliminary data has unveiled a significant finding that the *ARC* gene deficiency can lead to differential expression of genes involved in neuroplasticity, inflammation, and apoptosis. This discovery deepens our understanding of the role of the *ARC* gene in psychiatric disorders and provides a potential pathway for developing new therapeutic strategies. Besides, the transcriptomic and proteomic profiles for the *ARC*-deficiency SH-SY5Y cells further contribute to the growing body of evidence on the novel mechanisms underlying the effect of the *ARC* gene and the molecular pathways of the pathophysiology of schizophrenia.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5171W: Translating genomic risk models to clinical trials in a learning health system: the Prostate CAncer integrated Risk Evaluation (P-CARE) model ⭐

Authors:

J. Vassy[1], A. M. Dornisch[2,3], R. A. Karunamuni[2,3], C. A. Brunette[1], M. E. Danowski[4], I. P. Garraway[5,6], M. Gatzen[7,8], R. L. Hauger[3,2], C. Kachulis[7,8], A. S. Kibel[9], N. J. Lennon[7,8], J. A. Lynch[10], K. N. Maxwell[11,12], C. C. Teerlink[13,14], G. Xu[2,3], Million Veteran Program, T. M. Seibert[15,3]; [1]Harvard Med. Sch. at VA Boston Hlth.care System, Boston, MA, [2]Univ. of California San Diego, La Jolla, CA, [3]San Diego VAMC, San Diego, CA, [4]VA Boston Hlth.care System, Boston, MA, [5]VA Greater Los Angeles, Los Angeles, CA, [6]UCLA David Geffen Sch. of Med., Los Angeles, CA, [7]Broad Inst. of MIT and Harvard, Cambridge, MA, [8]Broad Clinical Labs, Cambridge, MA, [9]Brigham & Women's Hosp, Boston, MA, [10]Salt Lake City VA Med. Ctr., Salt Lake City, UT, [11]Univ. of Pennsylvania, Philadelphia, PA, [12]Philadelphia VAMC, Philadelphia, PA, [13]Univ Utah Sch Med, Salt Lake City, UT, [14]VA Salt Lake City Hlth.care System, Salt Lake City, UT, [15]Univ. of California San Diego, San Diego, CA

Abstract:

Background: As healthcare moves away from a one-size-fits-all approach, individual risk prediction is an important first step in disease prevention and early detection. Biobank-linked healthcare systems can generate knowledge about genomic risk and test the impact of implementing that knowledge in care. Here, we describe the development of a clinical translation pipeline for genomics-informed prostate cancer screening in a national healthcare system. **Methods:** We used data from 646,991 participants of the Veterans Affairs (VA) Million

Veteran Program (MVP) to develop and validate a clinical prostate cancer risk prediction model (P-CARE) based on a polygenic score, family history, and continuous genetic ancestry. The model was externally validated in 18,457 PRACTICAL Consortium participants. Additional validation was performed in 90,069 samples from unrelated males in the All of Us Research Program to develop a clinical laboratory assay on a blended genome-exome (BGE) platform for both the P-CARE model and rare prostate cancer-associated variants in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *EPCAM*, *HOXB13*, *MLH1*, *MSH2*, *MSH6*, *PALB2*, *PMS2*, and *TP53*. **Results:** In overall and ancestry-stratified analyses (57,111 AMR, 122,557 AFR, 10,671 ASN, 456,652 EUR), the final polygenic score of 601 single-nucleotide polymorphisms was associated with any, metastatic, and fatal prostate cancer in MVP and PRACTICAL. For example, P-CARE values at ≥80th percentile in the multiancestry cohort overall were associated with hazard ratios (HR) of 2.75 (95% CI 2.66-2.84), 2.78 (2.54-2.99), and 2.59 (2.22-2.97) for any, metastatic, and fatal prostate cancer in MVP, respectively, compared to the median. When high- and low-risk P-CARE groups were defined as HR>1.5 and HR<0.75 for metastatic prostate cancer, the 271,152 (41.9%) high-risk vs.158,770 (24.5%) low-risk participants in MVP had a 45.3% vs. 13.7%, 8.9% vs. 1.9%, and 3.4% vs. 0.8% cumulative cause-specific incidence of any, metastatic, and fatal prostate cancer by age 90, respectively. Among 118,058 participants self-identifying as Black/African-American, 2.4% and 80.0% were in the low- and high-risk groups. Clinical BGE reports and recommendations have been developed to communicate the P-CARE and monogenic results to participants and primary care providers. **Conclusions:** A model consisting of a polygenic score, family history, and genetic ancestry describes a clinically important gradient of prostate cancer risk, identifying individuals at higher and lower risk. The clinical assay and reports are now being implemented in a randomized trial of precision prostate cancer screening in the VA (clinicaltrials.gov NCT05926102).

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5172W: Type 2 Diabetes Is Causally and Independently Associated with Pulmonary Tuberculosis: A Mendelian Randomization Study

Authors:

H. Lee[1], S. Park[2], J. Choi[3], J. PARK[4], J. Lee[2], J-i. Kim[5], K. Park[2], S. Kwak[2]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [3]Seoul Natl. Univ., Jongno-gu, Korea, Republic of, [4]SEOUL Natl. Univ., seoul, Korea, Republic of, [5]Seoul Natl. Univ Coll. of Med, Seoul, Korea, Republic of

Abstract:

Tuberculosis (TB) is one of the leading causes of mortality from infectious disease, resulting in the loss of 1.3 million lives worldwide at 2022. An estimated 370,000 new cases of TB were suggested to be attributable to type 2 diabetes (T2D), which is one of the five main risk factors of TB. However, whether T2D is causally associated with TB is largely unknown.
We utilized two-sample Mendelian randomization (MR) investigate causal relationship between T2D and pulmonary TB. MR analysis was conducted in the East Asian population, where pulmonary TB is still endemic and the infection pressure is similar. Eighteen independent genome-wide significant variants of T2D were selected as instrumental variables from the meta-analysis of the Korean Genome and Epidemiology Study and the Taiwan Biobank (N=135,387, F-statistics=70.2). Genome-wide association study of pulmonary TB from the Biobank Japan (N=212,453) was used as outcome.
Summary-level MR analysis showed that having a higher genetic risk of T2D was significantly associated with higher pulmonary TB risk (inverse-variance weighted, beta=0.324, $P=3.4x10^{-5}$). This association was also significant in sensitivity analyses (MR-Egger P=0.046; weighted median $P=8.9x10^{-4}$; weighted mode P=0.023; MR-PRESSO $P=6.7x10^{-4}$; MR-RAPS; $P=9.9x10^{-4}$). MR-Egger intercept test showed no evidence of horizontal pleiotropy (P=0.36). T2D's causal effect on pulmonary TB was not mediated by other major risk factors of TB (smoking, alcohol intake and body mass index) and ten other cardiometabolic traits by multivariable MR analysis. Increased genetic risk of HbA1c was also significantly associated with increased pulmonary TB risk (43 non-erythrocytic variants, beta=0.604, P=0.029). Higher genetic risk for fasting glucose showed directional concordance with increased risk for pulmonary TB but did not reach statistical significance (beta=0.266, P=0.114).
Genetically predicted T2D is significantly and independently associated with higher risk of pulmonary TB. This MR analysis supports the causal relationship between T2D and pulmonary TB.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5173W: Uncoupling the genetics driving obesity and cardiometabolic disease

Authors:

A. Thornburg[1], R. Mitchell[1], N. Sakabe[2], I. Salamone[1], L. Zhang[1], M. Nobrega[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ Chicago, Chicago, IL

Abstract:

Cardiometabolic complications due to obesity, such as metabolic syndrome, type 2 diabetes, and cardiovascular disease create a major health burden and warrant a better understanding of the biology underlying their progression from obesity. Interestingly, there are individuals who are obese who do not have overt cardiometabolic complications, referred to as metabolically healthy obese (MHO). Previous work showed that differences in adipose biology distinguish MHO from metabolically unhealthy obese (MUO), placing adipocyte biology at the genesis of these opposing health outcomes. Research studies have also identified different genetic variants distinctly associated with MHO or MUO, suggesting a genetic component underlying the uncoupling of obesity and cardiometabolic disease. Previous work has identified noncoding genetic variants associated with MHO and MUO. We plan to identify the target genes of these noncoding variants and interrogate their role in adipocyte biology. To this end, we are currently generating Hi-C, ATAC-seq, and RNA-seq data from primary human mesenchymal stem cells that can be differentiated into adipocytes. We will then use these omics data to identify the likely target genes of the MHO and MUO associated regulatory variants. Once we have identified a list of candidate genes, we will then perturb these genes in adipocytes to assess how modulation of expression of these genes impact adipocyte biology. We are currently developing a CRISPRa and CRISPRi protocol in primary human mesenchymal stem cells to perturb these genes. We will then link perturbation of these genes to adipocyte biology by assessing molecular and cellular phenotypes such as differentiation efficiency, insulin sensitivity, and triglyceride levels. We expect to observe that genes associated with MHO will have distinct effects on adipocyte biology compared to genes associated with MUO. The results from this work will provide a better understanding of the genes and underlying biology that uncouples obesity and cardiometabolic disease. These findings can then be used to inform drug target design and improve personalized treatments for obesity-associated cardiometabolic disease.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5174W: Understanding the common genetic architecture of spontaneous coronary artery dissection using genetically related phenotypes

Authors:

I. Tarr[1,2], S. Hesselson[1,3], M. Troup[1,3], P. Young[1,3], L. Ma[4], Y. Gao[1,3], M. Bax[1,3], K. Junday[1,2], R. Johnson[1,3], L. McGrath-Cadell[1,2], D. Muller[1,5,3], S. Iismaa[1,3], D. Fatkin[1,3,5], J. Kovacic[1,3,5,4], R. Graham[1,3,5], E. Giannoulatou[6,3]; [1]Victor Chang Cardiac Res. Inst., Sydney, Australia, [2]Sch. of Clinical Med., Univ. of New South Wales, Sydney, Australia, [3]Univ. of New South Wales, Sydney,

Australia, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]St Vincent's Hosp., Sydney, Australia, [6]Victor Chang Cardiac Res. Inst., Darlinghurst, Australia

Abstract:

Spontaneous coronary artery dissection (SCAD) is a poorly understood yet significant cause of myocardial infarctions and predominantly affects women. The genetic etiology of SCAD is complex: a recent SCAD genome-wide association study (GWAS), identified 17 common variants, genetic overlap with other traits, and a high degree of heritability - most of which is unidentified. Understanding this largely female disease is important, yet the traditional approach of relying on the recruitment of very large cohorts is difficult in the context of SCAD. Here we aim to leverage the known genetic architecture for traits associated with SCAD to gain further insight into common variants contributing to SCAD risk, using two complementary approaches: analysing polygenic risk scores (PRS) on patient-level data, and performing a multi-trait GWAS.

PRS for comorbid traits (N_{PRS} = 21, N_{traits} = 11) - traits with genetic and/or epidemiological association with SCAD - and control PRS ($N_{control\ PRS}$ = 3), were calculated in case (N = 289) and control (N = 1127) cohorts, and associations between SCAD status and PRS were analysed via logistic regression. Within PRS analyses we also first 1) stratified cases by clinical comorbidity phenotype, and 2) removed from PRSs all variants in linkage disequilibrium (LD) with known SCAD loci. A multi-trait GWAS using MTAG software was applied to GWAS of SCAD, genetically correlated traits, and traits indicated by PRS analysis.

Control PRS performed as expected. PRS for three other vascular conditions, fibromuscular dysplasia (odds ratio [OR] = 1.46, 95% confidence interval [CI] = 1.07 - 1.99, p = 0.0177), cervical artery dissection (OR = 1.59, 95% CI = 1.32 - 1.92, p = 1.40E-6), and abdominal aortic aneurysm (OR = 1.24, 95% CI = 1.01-1.53, p = 0.04) were associated with SCAD. PRS for diastolic blood pressure and cholesterol were significantly associated with SCAD, even following LD pruning of PRS variants and in the subset of SCAD cases without known hypertension (OR = 1.71, 95% CI = 1.31-2.23, p = 7.29E-5) or cholesterol issues (OR = 0.70, 95% CI = 0.54 - 0.91, p = 0.0056). Preliminary MTAG analysis confirmed the suitability of this approach for the GWAS data selected. Complete results for SCAD and other traits will be detailed, as will validation of findings in separate disease-specific cohorts.

Using a PRS-based analysis of SCAD cases and controls, we have demonstrated genetic overlap between SCAD and cardiovascular diseases or related traits, which are not explained by known shared loci or comorbid cases. With a multi-trait GWAS targeting these associated traits, we expect to identify additional common variants contributing to SCAD.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5175W: Understanding the relationship between giant cell arteritis and polymyalgia rheumatica using population biobank data

Authors:

M. Liu[1], A. Blumenfeld[1], N. Smaoui[1], X. Zheng[2], F. Agliano[3], M. Kleyman[4], B. Riley-Gillis[5]; [1]Abbvie, North Chicago, IL, [2]AbbVie Inc / Genomics Res. Ctr., Lake Bluff, IL, [3]AbbVie, Cambridge, MA, [4]Abbvie, Worcester, MA, [5]AbbVie, Glenview, IL

Abstract:

Giant Cell Arteritis (GCA) and Polymyalgia Rheumatica (PMR) are immune-modulated rheumatological conditions associated with large vessel vasculopathy that are highly associated with each other. Disease occurs in populations that are over 50 years old and while onset is often acute, the presentation of symptoms is distinct between GCA and PMR. Patients with GCA tend to report pain and stiffness around the head and neck (headaches) while those with PMR report stiffness and pain in the joints. Both diseases require a 2 to 3-year corticosteroid taper treatment period, however long-term use is highly detrimental to overall health due to the degenerative side effects of this drug such as osteoporosis. To understand the etiology of GCA and PMR, we aim to identify key genes and variants that predispose an individual to disease and explore factors that are associated with incidence of disease.

To start, we identified cases of GCA (N=728) and PMR (N=3970) as having a prescription for corticosteroid, over the age of 50 and of European ancestry in the UK Biobank from their primary care read codes and hospital records. Next, we chose age-matched and ancestry-matched controls that do not have any musculoskeletal disorders (N = 47206). Using these case and control criteria, we ran 3 case-control GWAS for GCA, PMR and a combination of both GCA and PMR. In all 3 GWAS, we found HLA to be the most significant region in line with other immune-related diseases. We have also found the region around IL6ST to be significant in PMR which is consistent with current treatment options.

Next, we performed comorbidity analyses on our cohorts to identify additional diseases or medications that may be associated with or predictive of GCA or PMR. Shared comorbidities include osteoporosis, which may be explained by corticosteroid use, as well as distinct comorbidities such as headache for GCA and myalgia for PMR were identified. Longitudinal analyses of medication and disease codes before onset of GCA and PMR found that history of vaccinations might be protective for these two diseases.

Our study presents a comprehensive genetic and phenomic evaluation of GCA and PMR, two diseases that are closely related yet with many unique symptoms and comorbidities. We have replicated what is known in literature in our GWAS and found interesting results from our analysis of medical records in the UK Biobank.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5176W: Understanding the role of APOE genotype in the genetic regulation of gene expression and splicing in human microglia

Authors:

T. Naito[1,2], J. Humphrey[1], D. Tatarakis[3], C. Ha[3], K. Raju[3], K. Monroe[3], T. Sandmann[3], M. Belloy[4], D. A. Knowles[2,5], S. V. Andrews[3], T. Raj[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]New York Genome Ctr., New York, NY, [3]Denali Therapeutics, South San Francisco, CA, [4]Washington Univ., Saint Louis, MO, [5]Columbia Univ., New York, NY

Abstract:

Microglia, the resident innate immune cells of the central nervous system, are genetically implicated in multiple neurodegenerative diseases, including Alzheimer's disease (AD). Previously, we mapped expression quantitative trait loci (eQTLs) and mRNA splicing quantitative trait loci (sQTLs) in human microglia (microglia genome atlas, "MiGA"), identifying several loci where common genetic variants in microglia-specific regulatory elements explain AD risk loci identified by genome-wide association studies (GWAS). However, there is a limited understanding of how microglia eQTLs and sQTLs may differ by APOE genotype, a critical knowledge gap given that APOE is the dominant genetic risk factor for AD. To address this, we will conduct eQTL and sQTL mapping analyses within stratified groups defined by APOE genotype. We have assembled the largest collection to date of paired RNA-seq and genotype data from human microglia (N = 653 samples). This dataset was derived from pooling together both internal and publicly available samples from healthy controls and various neurological and psychiatric conditions (including AD), across several brain regions. Next, we will conduct eQTL and sQTL mapping separately in carriers of the APOE4 allele (N = 199) and those who do not carry the APOE4 allele (N = 454). These QTL analyses will utilize our recently generated large-scale RNA isoform atlas in human microglia assembled through long-read sequencing in 30 samples ("isoMiGA"), enabling comprehensive cataloging of microglia eQTLs and sQTLs. Through subsequent colocalization and fine-mapping analyses of the generated e/sQTLs with similarly APOE-stratified AD GWAS data, we aim to nominate loci that may exert influence on AD through changes in gene expression, and/or via isoform and splice site usage, either across or within distinct APOE genotypes. This work will refine the mechanism of genetic risk factors for AD within APOE genotype-defined populations, and therefore inform the identification of prospective therapeutic targets for intervention that are specific to these groups.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5177W: Unraveling the Genetic Architecture of Inflammatory Bowel Disease Through Large-Scale Exome Sequencing ⭐

Authors:

R. Zhu[1,2], M. Yu[1,2]; [1]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of Harvard and MIT, Cambridge, MA

Abstract:

Inflammatory bowel disease (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), has a complex genetic basis. Genome-wide association studies have identified over 300 loci associated with IBD. Initial sequencing efforts have revealed rare coding variants associated with CD, offering novel insights into disease mechanisms. However, the impact of low-frequency and rare coding variants in UC and IBD broadly remains underexplored. We conducted whole exome sequencing on 37,994 IBD cases (20,898 CD, 14,086 UC, and 2,465 IBD-uncertain) and 65,995 controls at the Broad Institute using the Nextera (Illumina) and Twist (Twist Biosciences) capture kits. After stringent quality control, we analyzed 10,821,747 variants, including 8,007,347 non-synonymous ones, using REGENIE v.2.2.4 with a logistic mixed model. We applied high-confidence variants (> 5% MAF, 99% call rate, LD-pruned) and Firth correction to address case-control imbalance. Sex and the top five principal components were included as covariates. We performed the analyses on CD and UC separately and for each exome capture kit, then meta-analyzed samples across capture kits using an inverse-variance weighted fixed-effect model. A significant threshold (P < 1.9E-7) was set for single-variant analysis, considering 257,178 non-synonymous variants with MAF (gnomAD NFE) between 0.0001 and 0.1. We identified 246, 170, and 399 variants (annotated to 40, 39, and 40 unique genes) significantly associated with CD, UC, and IBD respectively. Most top hits (P < 1E-20) are previously reported IBD-associated variants or those in LD with them, confirming the robustness of our QC and analysis. Our analysis revealed potentially novel loci associated with IBD. For example, a splice variant in ADAM17 (MAF = 0.0097), a frameshift variant (S709Lfs*22) in CSF2RB (MAF = 0.0011), and a missense variant (R113W) in PTCD1 (MAF = 0.0046), all significantly associated with CD and not previously reported. Subsequent exome sequencing was conducted at the Sanger Institute on an independent cohort of 27,883 IBD cases (10,749 CD, 13,295 UC, 3,839 IBD-U) and 168,355 non-IBD controls from the UK Biobank. We will perform a meta-analysis to combine this with the Broad Institute's callset for replication and power. In conclusion, here we report novel genes associated with IBD from the initial analysis of the exome-sequencing data from 37,994 IBD cases and controls. These findings will be validated and expanded with additional samples,

totaling 80,000 IBD cases and 460,000 controls. This study will provide novel insights into IBD's genetic architecture and prompt investigation into novel therapeutics.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5178W: Whole genome sequencing analysis identifies rare non-coding variants with large effects and regulatory regions associated with circulating protein levels ⭐

Authors:

R. Beaumont[1], K. Chundru[1], I. jackson[2], K. Patel[2], A. Murray[2], A. Wood[3], C. Wright[2], M. Weedon[2], T. Frayling[2,4], G. Hawkes[2]; [1]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [2]Univ. of Exeter, Exeter, United Kingdom, [3]Coll. of Med. and Hlth., Exeter Univ., Exeter, United Kingdom, [4]Université de Genève, Geneva, United Kingdom

Abstract:

Background: The role of rare non-coding genetic variation in common phenotypes is largely unknown, due to a lack of whole-genome sequence data, and the difficulty of categorising non-coding variants into biologically meaningful units.
Methods: To begin addressing these challenges, we performed a cis association analysis of 2,912 circulating protein levels in up to 53,021 UK Biobank participants using whole-genome sequence data. We tested 1.1 billion single variants, and preformed 123 million rare non-coding aggregate-based tests, including both annotation-based units and sliding windows agnostic to functional region annotations.
Results: We identified 2,006 independent rare non-coding single variant associations with circulating protein levels (P < 2.95x10-10) after conditioning on protein-coding and common associated variants. Rare non-coding aggregate testing identified 310 independent regulatory regions (P < 8.71x10-9) of which 179 (58%) were not detected through single variant testing, and 29 of these were detected only as sliding windows, indicating the possible presence of new regulatory features in these regions. Of regions not detected through single variant testing, 23.5% were from gene-based annotations, 56.4% were intergenic/intronic regulatory annotations, 6.1% were RNA units, and 14.0% were sliding-window based. Further, unlike coding aggregate associations, 98.4% of independent rare-variant non-coding aggregates were identified through tests allowing for sparse causality, highlighting the difficulty of non-coding annotation. Unlike protein-coding variation, rare non-coding genetic variation was equally likely to increase or decrease protein levels, and in some cases had effect sizes equal in magnitude to predicted loss-of-function coding variation. In addition, we observed an enrichment of rare non-

coding pQTLs in the 5' and 3' untranslated regions of protein-coding genes, which despite containing non-coding exons are not usually captured by exome-sequencing. The effect sizes of rare non-coding genetic variants did not show substantial evidence of variation due to linkage disequilibrium with coding genetic variants, suggesting they are less affected confounding by epitope effects than coding associations.
Conclusions: Our results have important implications for the identification and role of rare non-coding genetic variation associated with common human phenotypes, including, the importance of testing aggregates of non-coding variants.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5179W: Whole genome sequencing analysis of severe, early-onset chronic obstructive pulmonary disease

Authors:

K. Kim[1], X. Hu[2], S. Chun[3,4], S. Gharib[5,6], E. Oelsner[7], L. Rasmussen-Torvik[8], T. Sofer[9,10], H. Xu[11], B. Yu[12], R. Barr[7], A. Carson[13], R. Kalhan[14], R. Kaplan[15], D. Lomas[16], S. London[17], D. Meyers[18], A. Norwood[13], G. O'Connor[19], V. Ortega[18], B. Psaty[5,20], S. Rich[2], P. Sliwinski[21], Trans-Omics in Precision Medicine, E. Silverman[1], A. Manichaikul[2], M. Cho[1]; [1]Brigham and Women s Hosp., Boston, MA, [2]Univ. of Virginia, Charlottesville, VA, [3]Boston Children s Hosp., Boston, MA, [4]Div. of Pulmonary Med., Boston Children's Hosp., Boston, MA, [5]Cardiovascular Hlth.Res. Unit, Dept. of Med., Univ. of Washington, Seattle, WA, [6]Div. of Pulmonary, Critical Care and Sleep Med., Dept. of Med., Univ. of Washington, Seattle, WA, [7]Dept.s of Med. and Epidemiology, Columbia Univ. Med. Ctr., New York, NY, [8]Northwestern Univ., Chicago, IL, [9]Beth Israel Deaconess Med. Ctr. / Harvard Med. Sch., Boston, MA, [10]Div. of Sleep and Circadian Disorder, Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [11]Boston Univ., Boston, MA, [12]UTHlth.Houston, Houston, TX, [13]Dept. of Med., Univ. of Mississippi, Jackson, MS, [14]Dept. of Preventive Med. and Div. of Pulmonary and Critical Care Med., Dept. of Med., Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [15]Dept. of Epidemiology & Population Hlth., Albert Einstein Coll. of Med., Bronx, NY, [16]Univ. Coll. London, London, United Kingdom, [17]NIEHS, Research Triangle Park, NC, [18]Ctr. for Individualized Med., Mayo Clinic, Phoenix, AZ, [19]Pulmonary Ctr., Boston Univ. Sch. of Med., Boston, MA, [20]Dept.s of Epidemiology and Hlth.Systems and Population Hlth., Univ. of Washington, Seattle, WA, [21]Natl. Tuberculosis and Lung Disease Res. Inst., Warsaw, Poland

Abstract:

Introduction:Chronic obstructive pulmonary disease (COPD) is a leading cause of respiratory morbidity and mortality worldwide, imposing a significant burden on public health. Variants

in *SERPINA1* which result in alpha-1 antitrypsin deficiency, identify a subset of COPD variants that can be characterized by low allele frequency, early-onset of disease, and large effect size. While extensive research has uncovered common genetic factors associated with COPD, our understanding of rare genetic variants and severe COPD remains limited due to challenges such as the abundance and unclear effect of rare variants. Recent advancements including large-scale whole genome sequencing, as well as focusing on severe cases (especially those with early-onset of disease) and healthy controls offer promising approaches to overcome these challenges. We leveraged these strategies to identify and characterize rare variants that play a crucial role in COPD and lung function impairment. Methods:We performed a whole genome sequencing association study of severe early-onset COPD (age of ascertainment <= 55), severe COPD, and spirometric phenotypes (FEV1 and FEV1/FVC) in eight population-based studies (ARIC, CARDIA, CFS, CHS, FHS, HCHS/SOL, JHS and MESA) and seven COPD-enriched studies (COPDGene, ECLIPSE, Boston Early-Onset COPD Study, LTRC, Transcontinental COPD Genetics Study (Poland), ICGN, SPIROMICS). We investigated genome-wide single variants using the genome-wide regression test implemented in regenie. We considered a P-value of $5x10^{-8}$ for genome-wide significance. Results: After sample quality control, we retained 53,028 participants including 34,751 European, 12,508 African, 1,088 Latin American, and 674 East Asian ancestry. There were 5,063 severe COPD cases in this sample, among whom 1,109 were early-onset. In preliminary results, excluding variants within 1 Mbp windows of previously reported variants, we identified 8 associations in 6 genomic regions with severe early-onset COPD, or lung function impairment. Our top association for severe early-onset COPD was rs74297390 (intergenic to *POM121L12*, MAF = 9.24%, Beta = 0.41, P = 4.3×10^{-9}). For association with FEV1 and FEV1/FVC ratio, we identified variants near *DKK2* and *AKR1E2*, respectively at P < 5×10^{-8}, which were both associated with FEV1/FVC ratio in Shrine et al. (2019) with P < 8×10^{-6} and with consistent direction of effect. Conclusions: In a whole genome sequencing study focused on severe early-onset and severe COPD, we identified new loci for COPD and lung function. Ongoing work includes rare variant and variant set-based analysis and adjustment for polygenic risk scores.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5180W: Whole genome sequencing identifies deleterious novel variants in craniofacial enhancers of Africans with orofacial clefts

Authors:

E. Aladenika[1], M. Olujitan[2], L. Cook[3], A. Alade[4], W. Awotoye[5], L. Gowans[6], M. Eshete[7], T. Naicker[8], W. Adeyemo[9], A. Oladayo[1], T. Busch[1], A. Visel[10], J. Murray[11], P. Mossey[12], A.

Adeyemo[13], A. Butali[1]; [1]Univ. of Iowa, Iowa City, IA, [2]Univ. of Iowa, Coralville, IA, [3]Environmental Genomics and Systems Biology Div., Lawrence Berkeley Natl. Lab., Berkeley, CA, [4]Univ. of Iowa, Rockville, IA, [5]Univ. of Iowa, Iowa, IA, [6]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [7]Addis Ababa Univ., Addis Ababa, Ethiopia, [8]Univ. of KwaZulu-Natal and Inkosi Albert Luthuli Central Hosp., Durban, South Africa, [9]Dept. of Oral and Maxillofacial Surgery, Univ. of Lagos, Lagos, Nigeria, [10]Lawrence Berkeley Natl Lab., Berkeley, CA, [11]U of Iowa, Iowa City, IA, [12]Dept. of Orthodontics, Univ. of Dundee, Dundee, United Kingdom, [13]NIH, Bethesda, MD

Abstract:

Background: Orofacial clefts (OFC) are the most common congenital craniofacial defect, affecting about 1 in 700 newborns worldwide. Both genetic and environmental factors play a role in OFC development. While whole genome sequencing and genome-wide association studies have found common genetic variants linked to OFC, they explain only a small part of its heritability. Furthermore, about 90-95% of OFC-associated variants are in non-coding regions. Enhancers that regulate gene expression by binding transcription factors (TF) are in the non-coding region of the genome. Based on these, we hypothesize that rare and novel variants in these enhancers may disrupt TF binding and affect the expression of craniofacial genes. Methods: Whole genome sequencing was conducted on 130 African case-parent trios. Variants were identified and annotated using GATK. Putative craniofacial (CF) enhancers with very rare (MAF <0.1%) and novel variants were filtered using Varseq software and intersected with in-vivo validated enhancers. Variants with a CADD score of ≥15 were considered deleterious. Priority was given to variants predicted by RegulomeDB to affect transcription factor binding. Genes near these enhancers were identified using the UCSC genome browser and their relevance to craniofacial development was evaluated using MGI, OMIM, and CleftGeneDB. Genes implicated in any of these databases were prioritized.In addition, we are also conducting a case-control rare variant burden analysis to identify CF enhancers with a burden of rare variants. Results:We analyzed 4,579 putative enhancers and found 7,441 very rare and novel variants in 2,925 of them. Among these, 1,795 enhancers had at least two variants, and 175 overlapped with in vivo validated enhancers. Sixty-nine of these were active in facial structures. Twenty-five of these 69 enhancers contained 46 pathogenic variants. RegulomeDB predicted 5 of these variants to affect transcription factors involved in craniofacial development. Two novel variants (chr14:53572496 and chr20:7516913) were found near the BMP4 and *BMP2* genes in enhancers active in developing mouse faces at embryonic day 11.5. Additionally, a very rare variant, rs887296504 (MAF=0.0002416), was found in an enhancer near *HS2ST1*. *BMP4* and *BMP2* are known cleft candidate genes, and *HS2ST1* has mouse models with craniofacial phenotypes. Conclusions: Our study suggests that novel and very rare variants in validated craniofacial enhancers near *BMP4*, *BMP2*, and *HS2ST1* may affect gene regulation and contribute to OFC

development. Experiments to validate the functional and mechanistic implications of these variants are ongoing.

Session Title: Complex Traits and Polygenic Disorders Poster Wednesday Session

Board 5181W: Whole-genome sequencing identifies *SIKE1, CD36,* and *IFFO1* as candidate susceptible genes for systemic lupus erythematosus

Authors:

Y. Lei[1], L. Shao[2], F. Wang[1], H. Su[1], X. Dang[1], Q. Song[2], W. Yang[3]; [1]LKS Faculty of Med., HKU, Hong Kong, China, [2]Dept. of Rheumatology, Lupus Res. Inst., Affiliated Hosp., Jining Med. Univ., Jining, Shandong province, China, [3]Univ. of Hong Kong, Pokfulam, China

Abstract:

Systemic lupus erythematosus (SLE) is an inflammatory autoimmune disease marked by autoantibodies and multi-organ damage. Genetic factors play a significant role, with heritability estimates ranging from 43% to 66% across populations. Genome-wide association studies (GWAS) have identified over 130 loci linked to SLE, but these account for only 17% to 28% of heritability, suggesting undiscovered loci. Most GWAS use DNA microarrays to detect common variants, often omitting rare variants. This study used whole-genome sequencing (WGS) to compare rare loss-of-function (LOF) variants between SLE and control samples, finding significant differences in the genes *SIKE1* and *CD36*. Additionally, GWAS analysis identified a locus in *IFFO1* associated with SLE.

A total of 799 Northern Chinese individuals were included, comprising 200 SLE cases and 599 controls. All samples underwent 30x and 150bp paired-end WGS. Sequencing quality was checked with FASTQC, alignment was performed using the Burrows-Wheeler Aligner, SNVs/Indels were called using the GATK best practice pipeline, and joint calling was conducted with GLnexus. Variants with a total depth below 5 were set to missing, and those with an alternative allele depth below 5 were set to homozygous reference. After quality control, gene-based Fisher's exact test for LOF variants and GWAS analysis for common variants were conducted.

The gene-based Fisher's exact test showed significant differences in *SIKE1* and *CD36* between cases and controls: 18 in SLE samples vs. 5 in controls for *SIKE1* (P = 2.00E-6) and 6 in SLE samples vs. 0 in controls for *CD36* (P = 5.84E-4). *SIKE1* suppresses *IKBKE*, which is vital in neuropsychiatric SLE pathogenesis. *CD36* is involved in clearing apoptotic cells, contributing to SLE when deficient. GWAS analysis found rs7306500 in *IFFO1* associated with SLE, with an odds

ratio of 0.486 and P of 1.12E-7. *IFFO1* is linked to neuronal injury and inflammation, common in SLE. However, due to the small sample size, the P values for *SIKE1*, *CD36*, and *IFFO1* do not surpass the multiple testing threshold.
Most identified SLE-associated loci are common variants, with a gap between estimated and explained heritability. While rare variants have been linked to SLE, these studies often have small sample sizes, and no large-scale WGS study has been done for SLE. This study conducted WGS for 200 SLE patients and 599 controls, finds significant differences in rare LOF variants in *SIKE1* and *CD36*, and identifying *IFFO1* as a new candidate SLE-associated locus. Larger sample sizes and replication datasets are needed to confirm these findings and exclude population stratification or sampling bias.

Session Title: Epigenetics Poster Wednesday Session

Board 1002W: 3D-genome impacts of TAD-boundary-disrupting ultra-rare structural variants identified in schizophrenia ⭐

Authors:

L. Porras[1], M. Farrell[2], I. Rodriguez-Lausell[1], G. Iglesias-Maldonado[1], C. Leveque[1], N. Ancalade[2], C. Stoddard[3], G. Martínez[1], E. Tuliao[1], P. Giusti-Rodríguez[1], J. Szatkiewicz[4]; [1]Univ. of Florida, Gainesville, FL, [2]Univ. of North Carolina, Chapel Hill, NC, [3]Univ. of Connecticut, Farmington, CT, [4]Univ North Carolina, Chapel Hill, Chapel Hill, NC

Abstract:

Schizophrenia (SCZ) affects 1% of the U.S. population and involves a complex genetic structure, including 10 large rare structural variants, about 10 exonic variants, and over 250 common variants. Our recent study combined whole genome sequencing (WGS) with brain-specific epigenomic data to evaluate contributions to SCZ etiology from a variety of genetic variants accessible to WGS but not by previous technologies. We found an excess of ultra-rare SVs in SCZ localizing to boundaries of topologically associated domains (TADs). Disrupted TAD boundaries can cause misaligned enhancer-promoter contacts and dysregulated gene expression, linked to diseases like cancer and brain abnormalities.
We aimed to characterize two of those ultra-rare SVs (8 kb deletion from chr10; 6.5 kb deletion from chr13) to understand their effects on gene expression and 3D-genome organization. Using CRISPR/Cas9, we engineered these SVs in isogenic iPSCs. These iPSCs were differentiated into neuronal precursor cells (NPCs), excitatory glutamatergic neurons (CNs), and medium spiny neurons (MSNs). Hi-C and RNA-sequencing libraries were prepared and sequenced for both wild-type and SV-modified cells.

RNA-seq analysis confirmed successful differentiation, with each cell type clustering together, regardless of deletion. Hi-C data, analyzed using HiCRep to calculate the stratum-adjusted correlation coefficient (SCC), showed consistent clustering with RNA-seq data. CN and MSN cells were notably closer to each other than to NPC cells.
We analyzed the deletions' effects independently for each cell type. The deletions caused more differentially expressed genes (DEGs) in CN and MSN cells compared to NPCs. However, pathway analysis revealed similar pathways across all three cell types, often associated with synaptic activity. Hi-C analysis revealed differences in A/B compartment composition at 100 kb resolution. We also observed changes in insulation scores for TADs and TAD boundaries (50 kb), frequently interacting regions (FIREs; 50 kb), and chromatin loops (10 kb resolution). Ongoing analysis focuses on integrating and further parsing gene expression and Hi-C data to identify genes and biological pathways impacted by these deletions and determine their significance in the context of SCZ.

Session Title: Epigenetics Poster Wednesday Session

Board 1004W: An atlas of genetic effects on the monocyte methylome across European and African populations

Authors:

W. Zhang[1,2], X. Zhang[3], C. Qiu[3], Z. Zhang[1], K-J. Su[3], Z. Luo[3], M. Liu[1], B. Zhao[4], L. Wu[5], Q. Tian[3], H. Shen[3], C. Wu[1], H-W. Deng[3]; [1]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [2]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Tulane Univ., New Orleans, LA, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Univ. of Hawaii at Manoa, Honolulu, HI

Abstract:

Deciphering the genetic regulation of DNA methylation (DNAm) is essential for understanding the mechanisms underlying complex diseases. Current datasets often suffer from incomplete coverage of methylation sites and the use of tissues containing heterogeneous cell populations. Here, we present a comprehensive human methylome atlas based on deep whole-genome bisulfite sequencing (WGBS) and whole-genome sequencing (WGS) of purified monocytes from 298 European Americans (EA) and 160 African Americans (AA) in the Louisiana Osteoporosis Study. Monocytes play crucial roles in various diseases, including osteoporosis, Alzheimer's disease, inflammatory disorders, and cardiovascular conditions. Our atlas enables the analysis of over 25 million methylation sites. We identified 1,383,250 and 1,721,167 methylation sites in EA and AA, respectively, with 880,108 sites overlapping between the two populations, showing associations with common variants in cis-regions. While the identified cis-methylation

quantitative trait loci (cis-meQTLs) exhibited population specificity mainly due to differences in minor allele frequencies, the shared cis-meQTLs demonstrated high concordance between the two ancestries. Trans-meQTL analyses revealed an additional layer of genetic regulation. Notably, cis-heritability estimates revealed that the mean value for the AA population (0.09) was significantly higher than that for the EA population (0.04). We developed population-specific DNAm imputation models using Elastic Net, enabling methylome-wide association studies (MWAS) for 1,976,046 and 2,657,581 methylation sites in EA and AA, respectively. The performance of our MWAS models was validated through a systematic multi-ancestry analysis of 41 complex traits from the Million Veteran Program. Our findings bridge the gap between genomics and the monocyte methylome, uncovering novel methylation-phenotype associations and their transferability across diverse ancestries. The identified meQTLs, MWAS models, and data resources are freely available at www.gcbhub.org.

Session Title: Epigenetics Poster Wednesday Session

Board 1005W: An improved DNA methylation signature for Kabuki Syndrome for classification of KMT2D variants

Authors:

M. Wang, S. Choufani, S. Helal, R. Weksberg; The Hosp. for Sick Children, Toronto, ON, Canada

Abstract:

Germline pathogenic variants (PV) in *KMT2D* cause Kabuki syndrome type 1 (KS1), a neurodevelopmental disorder characterized by distinctive facial features, growth retardation and intellectual disability. While truncating variants are common, missense variants are also observed in KS1, posing challenges in their clinical assessment. Our prior research elucidated that PVs in epigenetic genes alter DNA methylation (DNAm) patterns, and identified DNAm signatures for over 60 neurodevelopmental disorders, underscoring their clinical relevance in the interpretation of variants of uncertain significance (VUS). The power and diagnostic utility of the signature is dependent on how it is derived, therefore we significantly increased the number of KS1 cases and controls to generate an improved DNAm based classifier. A total of 113 individuals with suspected KS1 and 327 typically developing individuals were recruited for DNAm profiling using Infinium HumanMethylationEPIC BeadChip. Comparing DNAm profiles between 25 *KMT2D* truncating variants and 43 sex- and age-matched controls unveiled 589 differentially methylated CpG probes (DMPs, $|\Delta\beta| > 10\%$ and FDR < 0.05) that were designated as the DNAm signature for KS1. We then trained a support vector machine (SVM) model on the discovery cohort to establish a classifier that was further tested on an independent KS1 and

control cohorts. All 31 *KMT2D* truncating variants were classified as “positive” (SVM > 0.5), while all 284 controls were classified as “negative” (SVM < 0.5), demonstrating a 100% specificity and sensitivity of this DNAm signature. Applying the DNAm signature to classify VUS identified 15 of 42 missense variants and all 5 intronic variants as positive. Remarkably, all 15 positive missense variants were predicted as deleterious by several *in silico* tools, including CADD, REVEL and AlphaMissense. Notably, DNAm classification showed a better correlation with clinical severity. Among 14 missense variants with available phenotypes, 6 exhibited minimal KS1 phenotypic features, all of which were classified as benign via DNAm signature. However, about half were predicted as deleterious via *in silico* tools, underscoring the superior performance of the DNAm signature in classifying the pathogenicity of *KMT2D* missense variants. Additionally, one out of three individuals without a previously reported *KMT2D* variant and two out of four individuals lacking an available genetic test were also classified as positive for KS1. In conclusion, the DNAm signature constitutes a valuable tool for delineating clinical syndromes and interpreting variants in individuals with uncertain clinical diagnoses of KS1.

Session Title: Epigenetics Poster Wednesday Session

Board 1006W: Atlas of single-cell ATAC plus gene expression reveals differential epigenetic regulation across many tissue types and species

Authors:

M. Garbati, S. Wang, A. Bartels, D. Lee, E. Wang; Active Motif, Inc, Carlsbad, CA

Abstract:

Single-Cell Multiome is a novel technique that combines single-cell gene expression analysis with single-cell open chromatin mapping to provide genome-wide sequencing of both the transcriptomal and epigenetic landscapes. Open chromatin is mapped using ATAC-seq (assay for transposase-accessible chromatin), taking advantage of the enzyme Tn5 transposase, which, in a process called tagmentation, inserts sequencing adapters into open regions of the genome *in situ* and epigenetic modifications control these open/closed chromatin states. Using nuclei as a starting material, Tn5 tagmentation is performed in bulk before these nuclei are loaded into a microfluidic channel where they individually enter barcoded gel beads so that only one nucleus interacts with only one bead. Within each gel bead, single-nucleus cDNA and ATAC libraries are constructed, and the unique barcodes within the gel beads allow these two libraries be traced back to the same cell with certainty.
Using this technology, our lab has profiled over 16 different cell or tissue types, including PBMC, heart, liver, various brain regions, and more, across five different species: human, mouse, pig,

snake, and hamster. These combined datasets have been integrated with complex bioinformatics analyses, revealing: 1) distinct ways that epigenetic states influence the expression of the same genes differently across cell types, 2) similar or identical cell types functioning differently across various tissue types, and 3) different functions of the same cell type within the same tissues across different species.Our growing body of work represents perhaps the most comprehensive atlas of combined single-cell epigenetic and gene expression data ever constructed. Collectively, these data contribute to a better understanding of how chromatin states influence gene expression in different cells and tissues.

Session Title: Epigenetics Poster Wednesday Session

Board 1007W: Characterizing Epigenomic Differences in Pulmonary Hypertension Subtypes

Authors:

T. Asir[1], A. Titus[2], S. Khader[3], K. Yadav[4]; [1]Univ. of Wisconsin-Madison, Madison, WI, [2]Dept. of Preventative Cardiology, Houston Methodist, Houston, TX, [3]Imperial Coll. London, South Kensington, United Kingdom, [4]Texas A&M Univ., Houston, TX

Abstract:

Pulmonary hypertension (PH) is a critical condition characterized by elevated blood pressure in the lungs, leading to increased cardiac workload and subsequent cardiac damage. Affecting 50-70 million individuals globally, and with an average prognosis of 2-3 years, understanding the underlying mechanisms is essential for developing advanced treatments. Epigenetic modifications, including DNA methylation and histone modification, play a pivotal role in PH development by altering gene expression without changing the DNA sequence. These modifications can lead to abnormal cellular functions, vascular remodeling, and inflammation, contributing to disease progression. In this study, we conducted an in-depth epigenomic analysis of four PH subtypes using an integrative bioinformatics approach. The subtypes investigated were Familial Primary Pulmonary Hypertension (FPPH), Familial Pulmonary Arterial Hypertension (FPAH), Idiopathic Pulmonary Hypertension (IPH), and Idiopathic Pulmonary Arterial Hypertension (IPAH). We integrated gene-disease association data from DisGeNET and performed a comprehensive enrichment analysis to explore the complexities of the epigenomic landscape. Using Enrichr, we utilized datasets including ChEA 2022, ENCODE, ChEA Consensus TFs, and ENCODE Histone Modifications 2015. Our analysis identified six shared genes (*ACVRL1, BMPR2, CAV1, ENG, KCNK3, and SMAD9*) across all PH subtypes. Additionally, we discovered subtype-specific genetic associations: FPPH was linked to four unique genes (*PRM3, HBB,*

MIR27A, PLA1A), while FPAH showed no unique genes. IPH was associated with 25 unique genes, including *BDNF* and *SLC52A1*, and IPAH exhibited 218 unique genetic associations, including *HEXIM1* and *MIR424*. From the ChEA 2022 data, we found no significant gene associations for FPPH, whereas FPAH had three significant genes, including *MTF2* (p = 8.24E-05). IPH presented with 47 significant genes, and IPAH had 93. According to the Histone Modifications 2015 data, significant histone modifications were identified exclusively for IPH, specifically H3K27me3 cerebellum mm9 (p = 0.002) and H3K27me3 ES-Bruce4 mm9 (p = 0.003). These findings emphasize the genetic complexity of pulmonary hypertension and identify potential targets for future genetic and epigenetic research. This may pave the way for the development of novel therapeutic strategies.

Session Title: Epigenetics Poster Wednesday Session

Board 1008W: Childhood adversity increases young adult Elevated Alcohol Consumption and contributes to increased young adult cardiac risk among Black youth

Authors:

S. Beach[1], M. Ong[1], M-K. Lei[1], R. L. Simons[1], F. X. Gibbons[2], M. Gerrard[2], R. A. Philibert[3]; [1]Univ. of Georgia, Athens, GA, [2]Univ. of Connecticut, Storrs, CT, [3]Univ Iowa, Iowa City, IA

Abstract:

Background: Premature aging and coronary heart disease (CHD) arise from a complex interplay of genetic and environmental vulnerability. Because epigenetics can be used to both quantify environmental exposures and the presence of disease states, studies using DNA methylation based methods can contribute to our understand of these relationships. **Hypothesis:** Using longitudinal serological, epigenetic and self-report data from Family and Community Health Study participants (N = 449) from age 10 to 29, we examined the dual pathway model (Beach et al., 2022) that distinguishes health behavior and immunological pathways, linking them to different early, prospectively assessed, exposures and positing independent contributions to the development of premature aging and CHD. Specifically, we test a key hypothesis of the Neuroimmune Network (NIN) hypothesis that chronically elevated immune processes associated with early adverse exposures have an effect on young adult substance use and contribute to increased problematic alcohol consumption in adulthood thereby contributing to accelerated epigenetic aging and CHD. By controlling the effect of early risky substance use and using a non-self-report index of young adult elevated alcohol consumption (EAC), we probe health effects attributable to the immunological pathway via effects on EAC, and test long-term

incubation effects hypothesized by the NIN model. **Results:** We identify a significant incubation effect of exposure to danger on EAC (B = -.157, P < .01) as well as significant indirect effects from the immunological pathway through young adult alcohol consumption to cardiac risk (IE = (-.004*, CI = [-.011, -.001]). Overall model fit is good (Chi-square = 7.163, df = 4; *p* = NS). We also examine sex differences in the magnitude of incubation and health effects, showing that effects are more robust for males. **Conclusions:** These results support the hypothesis that early adverse exposures can contribute to elevated alcohol consumption in young adulthood through a process independent of early onset substance use, and can contribute to cardiac risk, setting the stage for new approaches to prevention of premature aging and CHD across the lifespan.

Session Title: Epigenetics Poster Wednesday Session

Board 1009W: Comprehensive profiling of genetic and epigenetic features for FSHD diagnosis using targeted nanopore sequencing

Authors:

F. Scharf[1], E. Hannes[1,2], M. Lucas[1,3], S. Gehling[1], A. Hallermayr[1,3], A. Benet-Pages[1,4], U. Koehler[1], T. Neuhann[1], E. Holinski-Feder[1,3], M. C. Walter[2], B. Schoser[2], A. Abicht[1,2]; [1]MGZ Med. Genetics Ctr., Munich, Germany, [2]Friedrich Baur Inst. at the Dept. of Neurology, LMU Univ. Hosp., LMU, Munich, Germany, [3]Dept. of Med. IV, LMU Univ. Hosp., LMU, Munich, Germany, [4]Inst. of Neurogenomics, Helmholtz Res. Ctr., Neuherberg, Germany

Abstract:

The genetic diagnosis of facioscapulohumeral muscular dystrophy (FSHD) poses a significant clinical challenge, eluding detection through standard sequencing methods despite being the third most prevalent muscular dystrophy. Conventional diagnostic strategies address the known genetic parameters of FSHD1 and FSHD2: the required presence of a permissive haplotype, a size reduction of the D4Z4 repeat of chromosome 4q35, or a pathogenic variant in an epigenetic suppressor gene (FSHD2). However, we and others have recently shown that methylation not only acts as a marker of disease severity but may also function as a potential prognostic marker, suggesting the feasibility of predictive testing for FSHD disease severity. To circumvent the known limitations of conventional diagnostics and to complement genetic parameters with epigenetic ones, we leverage targeted nanopore sequencing to simultaneously analyze permissive haplotypes, the number of repeat units, methylation profiles, single nucleotide variations, structural variants and hybrid alleles, somatic mosaicism, the orientation of the contracted repeat array, in FSHD patients, all within a single sequencing run. As proof of concept, we were able to clarify the unexpected high average distal methylation in a

paucisymptomatic patient homozygous for 4qA, harboring a repeat contraction with two repeat units on one allele. The ONT data revealed the presence of the contraction as a mosaic, causing an increase in the average percent methylation. In another patient, resolving a complex hybrid allele revealed the presence of a hypermethylated rare 4qC166H haplotype, which makes FSHD unlikely as the reason for the patient's symptoms. Additionally, we present initial findings on the clinical validation of our nanopore sequencing method for FSHD diagnosis, using a patient cohort that we have validated through conventional methods such as KASP genotyping, Sanger sequencing, Southern blotting, optical genome mapping, and Illumina-based bisulfite sequencing.

Session Title: Epigenetics Poster Wednesday Session

Board 1010W: Controlled Fusion of Heterogenous ATAC-seq Samples for Robust Quantification of Chromatin Accessibility

Authors:

N. Hamilton; UNC Chapel Hill, Chapel Hill, NC

Abstract:

Next-generation sequencing (NGS) technologies such as ATAC-seq have found broad utility in the study gene regulatory programs characterizing a trait-group of interest. Data generated using these assays can be evaluated genome-wide with a mapping between genomic positions and the number of sequenced reads aligning thereto. Relevant epigenetic features (e.g., open chromatin, transcription factor binding sites, etc.) can be indicated by an enrichment in aligned read counts; However, regional and sample-specific technical artifacts can undermine the reliability of analyses based on this count data for capturing epigenetic states. Such factors are particularly problematic when integrating data from multiple samples to summarize epigenetic profiles for a trait-group of interest. Existing approaches based on "pooling" samples' reads or naive aggregation of counts to represent multiple samples' features are common in practice for their conceptual simplicity but are suboptimal in their extraction of information and are vulnerable to outliers.

We introduce Consenrich, an NGS-specific Kalman filter for robust, scalable quantification of chromatin accessibility across multiple samples. Motivated by the sparsity of chromatin accessibility data and physical properties of chromatin, Consenrich leverages a recursive kinematic model to compute *a priori* accessibility predictions over successive genomic intervals. These initial predictions are then refined with observed data using a localized, uncertainty-scaled approach to compute the *a posteriori* estimates. We show that Consenrich effectively

leverages information from samples of varying data quality, offers a substantial reduction in noise compared to pooling/aggregation-based approaches, and more clearly reveals consistently accessible regions and regulatory elements, improving consensus peak calling protocols and differential analyses.

Session Title: Epigenetics Poster Wednesday Session

Board 1011W: Development of bias correction methods for cord blood-based epigenome-wide association studies

Authors:

H. Ohmomo, A. Takashima, S. Minabe, Y. Sutoh, K. Ono, S. Umekage, S. Komaki, Y. Otsuka-Yamasaki, T. Hachiya, R. Oyama, T. Baba, M. Sasaki, A. Shimizu; Iwate Med. Univ., Shiwa-gun, Japan

Abstract:

Cord blood (CB) cell DNA methylation (DNAm) profiles are a promising resource for assessing fetal epigenetic changes that reflect prenatal in utero exposure. The developmental origins of health and disease hypothesis proposes that the parents' lifestyle prior to conception and the environment during pregnancy affect the sperm, egg, and fetus and influence the child's susceptibility to lifestyle-related diseases during childhood and adulthood. Immediately after birth, CB mononuclear cells contain a unique blood cell population called nucleated red blood cells (nRBCs), which account for approximately 10% of CB mononuclear cells. However, how their DNAm profile affects the CB-based DNAm analysis is still unknown.
In the present study, we recruited pregnant women scheduled for singleton and cesarean deliveries at Iwate Medical University Hospital and obtained written informed consent from 15 pregnant women. After cesarean delivery, CBs were immediately collected, and nRBCs were isolated using a cell sorter. After DNA extraction from isolated nRBCs, whole-genome bisulfite sequencing was performed. We examined the influence of nRBC DNAm profiles on CB-based epigenome-wide association studies (EWAS) of gestational age. This analysis used CB-DNAm data collected from the Tohoku Medical Megabank Project Birth and Three-Generation Cohort Study (TMM BirThree Cohort Study).
nRBCs with >94% purity were isolated from the CB. Principal component analysis showed that the DNAm profile of nRBCs differed from that of mature blood cell types. In addition, the following three correction models were used when performing EWAS for gestational age: Model 1) no correction, Model 2) correction using the cell-type composition of mature peripheral blood, and Model 3) correction using Model 2+nRBC DNAm profiles. These resulted in inflation

factors (λ values) of 1.502 (95%CI:1.495-1.509), 1.111 (95%CI:1.105-1.116), and 1.080 (95%CI:1.085-1.095), respectively, with Model 3 being the most effective correction.
In conclusion, these results suggest that the DNAm profile of nRBCs may introduce bias, albeit weak, into CB-based EWAS results. Therefore, correction using nRBC DNAm profiles is warranted when performing sequencing-based DNAm analyses and EWAS.

Session Title: Epigenetics Poster Wednesday Session

Board 1012W: Diagnosis of Mendelian diseases using genome-wide DNA methylation signatures: Preliminary practice from a Chinese patient cohort

Authors:

N. Li[1], G. Li[1], R-e. Yao[2], T. Yu[2], S. Li[1], J. Wang[1]; [1]The Intl. Peace Maternity and Child Hlth.Hosp., Sch. of Med., Shanghai Jiao Tong, Shanghai, China, [2]Shanghai Children's Med. Ctr., Sch. of Med., Shanghai Jiao Tong Univ., Shanghai, China

Abstract:

Background: Variants in coding genes related to epigenetic regulation can lead to disease-specific methylation profiles across the entire genome, and have been proven to be useful for molecular diagnosis of various neurodevelopmental disorders. However, the relevant tools have not yet been established and tested in Chinese patients. We aimed to identify disease-specific epi-signatures in Chinese patients, develop diagnostic models, and conduct clinical diagnostic efficacy evaluations. **Methods:** We collected 185 cases from 6 distinct disorders, including Sotos syndrome, Kabuki syndrome, Rett syndrome, CHARGE syndrome, Williams-Baure syndrome, and Prader-Willi syndrome (PWS), with no less than 30 patients for each group. Meanwhile, 40 age- and gender- matched healthy individuals were collected as controls. 3/4 of the patients were used as the "training cohort", and the whole genome methylation profile was analyzed using the Illumina Infinium Methylation EPIC v2.0 kit. After multiple filtering, disease-specific epi-signatures were constructed, and a classification diagnostic model was established based on support vector machine (SVM) training classifier. The remaining cases were used as the "test cohort" to evaluate the clinical diagnostic efficacy of the models. **Results:** A total of 634 probes (100-125/disease) were screened to construct the classification diagnostic model for each disease. 85% of the probes are disease specific, with only 4 probes present in each disease. Among the 35 cases of the test cohort, 28 obtained diagnostic conclusions that met expectations, including 20 cases with consistent clinical and genotype diagnoses. 8 variants of uncertain significance (VUS) were re-classified as likely benign, including a frameshift variant

(p.Lys144Argfs*21) in *NSD1* and a splicing variant (c.4020+2T>A) in *KMT2D*. We also tested *KDM6A* related Kabuki samples and found that the *KMT2D*-Kabuki model is not suitable for the recognition of *KDM6A* variants. The clinical diagnostic performance of PWS and CHARGE models is not ideal, which may lead to 50% false positives. Further analysis revealed that the performance of the PWS model may be related to the different genotypes (uniparental disomy vs. heterozygous deletion). **Conclusions:** This study explores the efficacy of whole genome methylation analysis in the diagnosis of Chinese genetic disease patients. The models we constructed show the potential of differential diagnosis. These localized diagnostic models not only provide new molecular basis for VUS interpretation, but also has guiding value for cases with highly matched genotypes but significantly inconsistent clinical symptoms.

Session Title: Epigenetics Poster Wednesday Session

Board 1013W: DNA methylation patterns and differences in two long-read sequencing platforms in a clinical pediatric cohort

Authors:

S. Batalov[1], E. Sanford Kobayashi[2], L. Van Der Kraan[1], M. Bainbridge[1]; [1]Rady Children s Hosp., San Diego, CA, [2]Rady Children's Hosp., San Diego, CA

Abstract:

DNA methylation is a critical epigenetic signature that plays a role in human development and disease and is especially recognized for its role in imprinted diseases, such as Angelman and Prader-Willi and Russel-Silver syndromes. Further, it may serve as a molecular marker of disease severity or progression even in diseases not primarily driven by methylation. Direct detection of methylation of the 5-position of cytosines (5mC) across the entire genome and without special library preparation is possible with two long-read sequencing technologies: Oxford Nanopore and PacBio.

We single-molecule, whole genome sequenced 100 pediatric patients with severe, acute disease using the Oxford Nanopore (55) and PacBio (45) platforms. Using these data, we developed a novel algorithm for detecting regions across the genome with aberrant methylation by using our samples to train an algorithm on methylation background established using a panel of normals (PON) approach. Using additional short-read parental variant data we can parent-of-origin phase across the proband methylome. Further, our method can be used to detect imprinting errors even without parental samples by taking advantage of the phased nature of the sequencing data.

We successfully re-identified 58 known regions ($p < 6.4*10^{-5}$; Student's t-test) of imprinting

and detected additional 16 regions in ONT data and 6 regions in PB data with 26 shared regions that appeared to have differential methylation based on parent of origin. Compared to data generated from cell-lines, we identify more regions of statistically significant dysregulation (p < $1*10^{-6}$; Student's t-test) in our pediatric cohort. These data also allowed us to identify significant methylation trans-dysregulation (on chr5, chr19, etc) in a trisomy 21 case, consistent with prior published data. We also demonstrate broad methylation dysregulation in patients with multiple immune disorders that affect genes involved in hematopoiesis.
Together these data show the value of long-read single molecule sequencing to elucidate methylation patterns 'for free' when whole genome sequencing. These data can identify known imprinted regions, diagnose diseases of methylation and can even identify molecular markers of pathology in diseases that are not driven by methylation. This further punctuates the multifaceted role of long-read sequencing to diagnose a plethora of diseases.

Session Title: Epigenetics Poster Wednesday Session

Board 1014W: DNA methylation signatures in the prefrontal cortex of alcohol-naïve nonhuman primates predict risk for future ethanol drinking.

Authors:

R. Cervera Juanes[1], K. D. Zimmerman[2], C. C. Lowe[1], S. W. Gonzales[3], B. Ferguson[3], K. Grant[3]; [1]Atrium Hlth.Wake Forest Baptist, Winston-Salem, NC, [2]Ctr. for Precision Med., Winston-salem, NC, [3]OHSU, Portland, OR

Abstract:

Alcohol use disorder (AUD) is a highly prevalent, complex, multifactorial and heterogeneous disorder, with 11% and 30% of adults meeting criteria for past-year and lifetime AUD, respectively. Early and accurate identification of individuals at risk of developing an AUD would facilitate effective deployment of personalized interventions. Studies using rhesus monkeys and rats, have demonstrated that individuals with low cognitive flexibility and a predisposition towards habitual behaviors show an increased risk for future heavy drinking. Further, low cognitive flexibility is associated with reduced dorsolateral prefrontal cortex (dlPFC) function in rhesus monkeys. To explore the underlying unique molecular signatures that may predict risk for chronic heavy drinking, a genome-wide DNA methylation (DNAm) analysis of the alcohol-naïve dlPFC-A46 biopsy prior to chronic alcohol self-administration was conducted. The DNAm profile provides a molecular snapshot of the alcohol-naïve dlPFC, with mapped genes and associated signaling pathways that vary across individuals. The analysis identified ten DMRs that

were strongly predictive of average ethanol intake consumed over 6 months of voluntary self-administration. Importantly, the same set of DMRs was able to predict drinking levels with moderate success in an independent cohort of macaques. These findings translate behavioral phenotypes into neural markers of risk for AUD, and hold promise for parallel discoveries in risk for other disorders involving impaired cognitive flexibility.

Session Title: Epigenetics Poster Wednesday Session

Board 1015W: Downregulation of nuclear receptor-binding SET domain protein induces neurotoxic cytokines via p38 mitogen-activated protein kinase and Jun kinase pathways in a glioma cell line, U87MG

Authors:

K. ByungJun; Konkuk Univ., Seoul, Korea, Republic of

Abstract:

The nuclear receptor binding SET domain-containing protein 1 (NSD1) is a histone methyltransferase that mono- or di-methylates the lysine residue 36 on histone H3. In humans, haploid deficiency of *NSD1* causes a neurodevelopmental disorder, Sotos syndrome (SOTOS). Despite the detailed description of the clinical features and genetic analyses of SOTOS, the molecular mechanisms underlying the role of *NSD1* loss-of-function in the neurological dysfunction in SOTOS are partially understood. Glial cells are critical in neurodevelopment, and activated glia can induce neuroinflammation via multiple signaling pathways. Here, we examined the effect of *NSD1* knockdown in glial cells on neuronal cells using cell lines. To downregulate NSD1, we transiently transfected siRNA, specifically targeting *NSD1*, siNSD1, into a human glioma cell line, U87MG. The siNSD1-treated U87MG exhibited specific activation of the p38 mitogen-activated protein kinase (MAPK) and Jun kinase (JNK) pathways and characteristic morphological changes similar to those of activated astrocytes. The transcriptome analysis of U87MG revealed cell-line-specific gene expression patterns owing to the *NSD1* knockdown compared with that of a non-glial cell line. The transcriptional levels of many genes related to p38MAPK and JNK pathways were increased in the siNSD1-treated U87MG, including interleukin (*IL)-1α*, *IL-1β*, and *IL-6*, which encode proinflammatory cytokines. Although a JNK inhibitor did not prevent the siNSD1 effect on *IL-1α* expression, the inductions of the three cytokines were significantly decreased by a p38MAPK inhibitor, indicating a more crucial role of p38MAPK than JNK signaling in *NSD1* knockdown-related gene regulation. Co-culturing human neuron-like SH-SY5Y cells with siNSD1-treated U87MG induced apoptotic cell death owing to the *NSD1* downregulation-induced neurotoxicity. These results suggest that

NSD1 deficiency in glial cells may cause neurotoxicity in the brain, possibly owing to the increased proinflammatory cytokine secretion induced by p38 and JNK activation, suggesting the association of these pathways with neurological disorders in SOTOS patients.

Session Title: Epigenetics Poster Wednesday Session

Board 1016W: Endogenous retroviruses are a source of regulatory variants in colorectal cancer

Authors:

A. Ivancevic[1], D. Simpson[1], O. Joyner[1], L. Nguyen[1,2], S. Bagby[3], T. Pitts[3], B. Bitler[2], E. Chuong[1]; [1]Univ. of Colorado Boulder, Boulder, CO, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Univ. of Colorado Cancer Ctr., Aurora, CO

Abstract:

Repetitive elements such as endogenous retroviruses shape chromatin landscapes. They facilitate the generation of novel cell-type-specific regulatory networks by replicating themselves and dispersing hundreds of potential regulatory sequences throughout the genome. Despite their recognized importance in cell biology, repetitive sequences continue to be understudied, especially as a source of regulatory elements that are epigenetically silenced in healthy cells but become activated in diseased cells.

Recently, we used multi-omic sequencing data and CRISPR experiments to show that ancient LTR10 elements can reactivate as tumor-specific enhancers in human colorectal cancer cells. LTR10 elements are derived from the long terminal repeat of a primate gammaretrovirus, HERV-I, which integrated into the anthropoid genome 30 million years ago. While heavily degraded in the human genome, many LTR10 sequences still contain binding motifs for the AP1 transcription factor complex. In HCT116 colorectal cancer cells, we found that LTR10 elements causally drive AP1-dependent expression of genes with established roles in tumorigenesis and therapy resistance, including *ATG12*, *XRCC4*, and *VCAN*. Similarly, bulk and single cell RNA sequencing from patient tumors showed that LTR10 elements are transcriptionally activated in a substantial fraction (~30%) of colorectal cancer cases.

We further discovered that individual LTR10 sequences contain variable number tandem repeat regions that may affect AP1 binding activity. Although all LTR10 insertions are fixed in the human population, the number of AP1 motifs within an LTR10 sequence varies between individuals and even between tumors. We observed extensive germline variation at AP1 motifs within LTR10 elements, supported by both short-read and long-read datasets. Additionally, long-read whole-genome sequencing from matched colorectal tumor and normal tissues showed

evidence of tumor-specific LTR10 expansions, sometimes thousands of bp in length. We suspect these LTR10 expansions drive tumor-specific gene regulatory activity. Altogether, our work implicates endogenous retroviruses like LTR10 elements as a source of pathological regulatory variants that facilitate transcriptional rewiring in cancer.

Session Title: Epigenetics Poster Wednesday Session

Board 1017W: Endothelial cell to cell adhesion is altered as consequence of loss of function of neurofibromin and afadin6 in neurofibromatosis type1

Authors:

M. Sulaiman[1], A. Dodson[2]; [1]Ahmadu Bello Univ., Zaria, Zaria, Nigeria, [2]Univ. of Liverpool, Liverpool, United Kingdom

Abstract:

&gtNeurofibromatosis type 1 (NF1) is a common autosomal disorder in human with incidence of 1 in 3000 people worldwide. The NF1 gene is mapped to chromosome 17, (17p21). Its protein product, neurofibromin negatively regulates Ras. Therefore, loss of function of neurofibromin may promote cell proliferation. The AF-6 protein is a multidomain protein that contains two potential Ras binding domains and is mapped to 6q27 of human chromosome. Endothelial cells are major constituent of Platelet Endothelial Cell Adhesion Molecule (PECAM) 17q23.3, otherwise referred to as CD31. It participates in maintenance of cell-cell junctional adhesion.Aim of the Study: This study was undertaken in order to demonstrate the likely consequences of instability of chromosome 17p21 on 17q23.3 and 6q27 on the functional integrity of cell-cell junction in familiar, sporadic and plexiform neurofibroma.Material and Methods: Prior informed patient consent was obtained, and the study has an ethical approval (06/1505/137) granted by the Liverpool Research Ethics Committee. Twenty four (24) cases involving 9 familial, 7 sporadic and 8 Plexiform neurofibromas were utilized. Standard immunohistochemistry method was employed. Briefly, routine histological procedures (haematoxylin and eosin staining) was first carried out followed by staining using cell markers and Rabbit anti-AF-6 polyclonal Zymed (Cat 433280) staining on both experimental and control tissues.Result: Moderate membranous, cytoplasmic and nuclei immunolocalization of the AF-6 were observed in endothelial cells of all the familial and the plexiform forms. Similarly, the AF6 moderately localizes to the membrane of the endothelial cell of the sporadic neurofibroma studied.Conclusion: From this study, we suggest that the instability of chromosome 17p21 may affect 17q23.3 and 6q27 (afadin-6) in familiar, plexiform and sporadic neurofibroma with the

attendant loss of cell-cell adhesion and transendothelial cell migration to distant tissues with subsequent transformation of plexiform neurofibroma to Malignant Peripheral Nerve Sheath Tumour (MPNST).&lt

Session Title: Epigenetics Poster Wednesday Session

Board 1018W: Epigenetic clocks as novel biomarkers of checkpoint inhibitor treatment response

Authors:

V. Chat[1], A. Felez[2], M. Waldvogel[2], N. Bienz[2], V. Iglesias[2], M. Srivastava[1], B. Nabet[1], S. Mariathasan[1], I. Mellman[1], C. Hammer[3], M. McCarthy[1]; [1]Genentech, South San Francisco, CA, [2]F. Hoffmann-La Roche Ltd, Basel, Switzerland, [3]Altos Labs, Redwood City, CA

Abstract:

Age-related morbidities are largely attributed to immune decline. Despite the strong impact of aging on immunocompetence—a key determinant of cancer immune checkpoint inhibitor (ICI) efficacy—the relationship between age and ICI treatment is poorly characterized. Epigenetic clocks, defined by the aggregation of age-varying CpG sites, have emerged as a proxy of biological age predicting age-associated phenotypes better than chronological age (cAge). In this study, we investigated epigenetic clocks' potential as novel ICI efficacy biomarkers. We assessed DNA methylation (DNAm) from baseline whole blood using Illumina EPIC v2.0. The study cohort included 750 patients from anti-PD-L1 (as monotherapy or combination with chemotherapy) and control arms of IMpower131/IMpower150 (Non-small cell lung cancer: NSCLC) and IMvigor211 (Urothelial carcinoma). Data processing and normalization were performed with *Minfi*, followed by cell-type deconvolution with *EpiDISH*. We estimated the DNAm age for 4 common clocks (Horvath2013, Horvath2018, Hannum, and Levine) using *methylclock*. DNAm age is highly correlated with cAge (Pearson r=0.72-0.87), with the lowest correlation observed for Levine's clock as it was trained on mortality risks rather than cAge. Logistic regression models to test for the association between age acceleration (difference between DNAm age and cAge) and ICI response (SD/PD: non-responders vs. CR/PR: responders) were adjusted for cAge, sex, smoking, blood cell compositions, sample plates, and top 3 principal components. We showed that, while cAge did not predict response, high age acceleration is statistically significantly associated with improved response in IMpower131 for two clocks (Horvath2018: OR= 0.79 [0.68-0.91] p=0.001, and Hannum: 0.86 [0.76-0.96] p=0.009). We also investigated the effect of the intrinsic epigenetic age acceleration (IEAA: DNAm age residuals after regressing out cAge), adjusting for the above covariates except for cAge. Similar results were observed in

IMpower131 (Horvath2018: OR=0.83 [0.72-0.94] p=0.005, and Hannum: 0.87 [0.77-0.97] p =0.01). These signals were not detected in the anti-PD-L1 arms of the IMpower150 and IMvigor211 trials, or the control arms of all three trials. The observed direction of effect in the anti-PD-L1 IMpower131 is counter-intuitive, which may be explained by the recent evidence pointing to the association between aging and positive ICI efficacy biomarkers (increased tumor-mutational burden and IFN-γ signaling). Works are underway to explore why these signals were IMpower131-specific, and further dissect ICI-related biological components of these clocks.

Session Title: Epigenetics Poster Wednesday Session

Board 1019W: Epigenetic reprogramming successfully reactivates gamma globin expression, paving way for β-hemoglobinopathies treatment ⭐

Authors:

A. Faiq[1,2], S. Haider[1,2], C. Mussolino[1,2]; [1]Inst. of Transfusion Med., Freiburg Im Breisgau, Germany, [2]Med. Ctr. - Univ. of Freiburg, Freiburg Im Breisgau, Germany

Abstract:

β-Hemoglobinopathies, severe inherited disorders caused by β-globin gene mutations, can have their clinical impacts greatly alleviated through the reactivation of γ-globin (HBG) genes. This reactivation boosts the production of fetal hemoglobin, counteracting the detrimental effects of the mutations. Our study aims to explore the epigenetic landscape of the γ-globin promoter and to use designer epigenetic activators (DEAs) for precise epigenome editing to activate the HBG (γ-globin) gene. initially, we profiled the epigenetic signatures present at the silenced HBG1/2 promoters. Employing ChIP-qPCR and bisulfite sequencing techniques, we validated the enrichment of repressive epigenetic marks associated with gene inactivation, such as elevated levels of H3K9me3, H3K27me3 histone modifications, and extensive DNA methylation. We then developed DEAs targeting the proximal γ-globin promoter and evaluated their efficacy in reactivating HBG expression in HEK293T cells. DEMs were delivered to the cells as in vitro transcribed mRNA by lipofection. Two days later, we measured HBG expression levels using quantitative polymerase chain reaction (qPCR) and demonstrated that DEMs were highly efficient in achieving a remarkable 25-fold increase in γ-globin levels. In addition, we extended our investigation to clinically relevant hematopoietic stem and progenitor cells (HSPCs) and observed up to a 13-fold increase in γ-globin gene expression two days post-transfection. Current studies are focused on elucidating the sustained impacts of epigenome editing on HBG expression levels. These promising findings underscore the remarkable potential of DEAs in

targeting the HBG promoter for reactivating γ-globin expression, thereby establishing a novel therapeutic approach for β-hemoglobinopathies. This development paves the way for exciting new avenues in therapeutic interventions.

Session Title: Epigenetics Poster Wednesday Session

Board 1020W: Epigenetic variation linked to opioid overdose converges on key target genes in the nucleus accumbens

Authors:

O. Corradin[1,2], M. Chakraborty[3], R. Sallari[3], F. Qureshi[4], V. Evans[5], T. Lambert[5], A. Hoang[2], B. Quach[6], K. Lovrenert[7], C. Willis[6], C. Bartels[7], P. Scacheri[7], E. Johnson[6], S. Akbarian[5]; [1]MIT, Boston, MA, [2]Whitehead Inst. for BioMed. Res., Cambridge, MA, [3]Axiotl Inc., Cleveland, OH, [4]Whitehead Inst., Cambridge, MA, [5]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [6]RTI Intl., Chapel Hill, NC, [7]Case Western Reserve Univ., Cleveland, OH

Abstract:

Opioid overdose is the leading cause of accidental death in the US. Epigenetic studies have the potential to identify both the genetic and environmental impacts to gene regulation that occur throughout the course of disease. We used chromatin profiling to detect changes in gene regulation indicated by histone modification H3K27ac, a marker of active promoters and enhancers. We profiled brain tissue samples from the nucleus accumbens of individuals who died from opioid overdose (n= 47) as well as accidental death controls (n=44) (age and ancestry matched). Given the heterogeneity of this disorder, we applied an analytical approach to identify epigenetic changes that distinguish individual opioid cases from controls using information theory-based analysis. This approach identified 37,652 Variant Enhancer Loci (VELs), genomic regions where individual cases varied significantly from the expected acetylation patterns observed in controls (P<1E-5). When we compared these changes across individuals, we found that the epigenetic changes were not randomly distributed across the genome, rather, converged on specific genetic loci and frequently physically interacted with a shared promoter. Using HiChIP data from the caudate nucleus, we found 18 genomic regions with significant convergence of epigenetic variation (P<1E-5). These regions include genes identified in genetic studies of substance use disorder (CELF5, RBFOX1, MN1) and that have been linked to other neuropsychiatric disorders (PRMT8, SORC3, PLXNA4). The convergent loci were significantly enriched for gene ontologies, splicing and synapse assembly (P<0.001). We also found target genes to frequently co-occur with transcription factor NPAS4, a gene linked to opioid use disorder in orthogonal studies. The convergence of epigenetic variation we observed

in the NAc recapitulated observations from a prior study of the prefrontal cortex (n=101), with 22% of VELs identified in both cohorts. Interestingly, the majority of VELs occurred in enhancers with high brain region specificity in the controls, suggesting cell type-specific impacts of opioid overdose. Finally, we investigated whether genetic variation was a major contributor to VELs. We performed histone modification QTL analysis and found the majority of VELs could not be explained by identified SNPs. Furthermore, while pretrained machine learning models successfully predicted ChIP activity at other loci, they failed to accurately predict ChIP activity at VEL loci, suggesting a strong environmental component. These results highlight the value of epigenetic studies for elucidating the etiology of complex traits like opioid addiction.

Session Title: Epigenetics Poster Wednesday Session

Board 1021W: Epigenome-wide association study of placental DNA methylation and cumulative maternal glycemic levels throughout pregnancy

Authors:

T. Habtewold, P. Wijesiriwardhana, R. Biedrzycki, F. Tekola-Ayele; Natl. Inst. of Child Hlth.and Human Dev., Bethesda, MD

Abstract:

Background: Maternal glycemic dysregulation during pregnancy increases the risk of pregnancy complications and offspring long-term health outcomes. Placental DNA methylation (DNAm) is a potential mechanism underlying these associations. We investigated whether maternal cumulative glycemic trait concentration throughout pregnancy is associated with placental DNAm in multiethnic women.
Methods: Cumulative glycemic exposure across four time point measurements at weeks 8-37 of pregnancy was estimated using area under the curve (non-fasting glucose (AUC_{gluc}), insulin (AUC_{insl}), and HbA1c (AUC_{hba1c})) in 301 pregnant women from the NICHD Fetal Growth Studies-Singletons cohort. DNAm was measured using the Illumina Infinium Human Methylation 450K BeadChip. The association of cumulative glycemic trait concentration with epigenome-wide placental DNAm was tested using linear regression models adjusted for gestational age, batch effect, placental cell types, surrogate variables, and methylation and genotype principal components. Analyses were performed for all women and stratified by fetal sex. For differentially methylated CpGs, we tested for correlation with nearby gene (±200KB) expression in placenta in a subset of 75 samples. Enrichment of CpGs and nearby genes biological processes or function were assessed. Finally, the associated CpGs were looked-up in the EWAS

catalog to check their overlap with previously reported diseases/traits.
Results: Maternal AUC_{gluc} and AUC_{insl} were associated with placental DNAm at 7 and 178 CpGs ($p_{FDR} < 0.05$), respectively, while there was no association with AUC_{hba1c}. Sex-specific associations were observed across all glycemic traits. DNAm at 10 AUC_{insl}-associated CpG sites were correlated with expression of 10 genes, the strongest correlation between cg21233754 and *GPD2* ($r = -0.45$, $p = 5.7 \times 10^{-5}$). The 10 genes include transcripts highly expressed in brain, reproductive, and cardiovascular tissues. Differentially methylated CpGs were enriched for pathways related to nitric oxide signaling in cardiovascular system, insulin secretion signaling pathway, metabolism of nitric oxide, platelet homeostasis, transport of inorganic cations/anions and amino acids, and thioredoxin. We observed overlap with CpGs previously associated with diseases/traits, including cardiometabolic and psychiatric disorders.
Conclusions: Placental methylation is altered in relation to maternal glucose and insulin exposure levels throughout pregnancy. Our analyses suggest a potential placenta-mediated epigenetic mechanism in the link between glycemic dysregulation and maternal and offspring health.

Session Title: Epigenetics Poster Wednesday Session

Board 1022W: Evaluating CRISPR-Cas9 scaRNA1 deletion on biochemical modification and function of the spliceosome

Authors:

A. Gardner[1], D. Bittel[1], N. Kibiryeva[1], J. O'Brien[2], M. Filla[1]; [1]Kansas City Univ., Kansas City, MO, [2]Children's Mercy Hosp., Kansas City, MO

Abstract:

Our data support an epigenetic role for noncoding RNAs (ncRNAs) in alternative splicing of mRNA and contributing to congenital heart defects. The mechanism by which ncRNAs [specifically small Cajal- body specific RNA (scaRNA)], cause these effects has yet to be fully explored. scaRNA1 directs the pseudouridylation of spliceosomal RNA U2 at U89. We hypothesize that reducing or eliminating scaRNA1 will alter spliceosomal function causing altered splicing of messenger RNA. We used a CRISPR-Cas9 enzyme complex with two guide RNA's (gRNA) targeting scaRNA1 to induce a deletion within the scaRNA1 genomic sequence in quail myoblast 7 (QM7) cells. This genomic deletion was verified using the T7EI endonuclease assay and DNA sequencing. Here we present our assessment of scaRNA1 expression level on pseudouridylation of U89 and mRNA splicing. This research will improve our understanding of the differential expression of scaRNA1 and its potential impact on regulatory pathways that are

important for cardiac development and when dysregulated may contribute to congenital heart defects.

Session Title: Epigenetics Poster Wednesday Session

Board 1023W: Functional redundancy and pathogenic mechanism in *RING1* variants: Insights from H2AK119ub1 dosage in human neural development

Authors:

S. Regan[1], C. W. Ryan[1], Y. Tsan[1], E. Mills[1], E. Gong[1], J. Sheingold[1], H. Theodore[1], A. Moccia[1], A. Srivastava[2], S. Bielas[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]SGPGIMS, Lucknow, India

Abstract:

Genetic variants in chromatin regulators are frequently found in neurodevelopmental disorders, but their effect in disease etiology is rarely determined. Here, we functionally define the pathogenic mechanism of missense variants in *RING1*, a component of the polycomb repressive complex 1 (PRC1), which modifies chromatin through catalysis of histone 2A lysine 119 monoubiquitination (H2AK119ub1). RING1 and RNF2 interchangeably serve as the catalytic subunit within PRC1. The recent discovery of pathogenic *RING1* missense variants brings into question decades of work questioning the functionally relevance of *RING1*. This assumption was based on knockout mouse models that are viable with minimal overt phenotypes. *De novo RING1* missense variants are the genetic basis of a rare neurodevelopmental disorder characterized by microcephaly, growth restrictions and early on onset schizophrenia. To investigate the molecular pathology, we created an allelic series of *RING1* and *RNF2* missense (MS) human embryonic stem cell (hESC) lines. CUT&RUN analysis of neural progenitor cells (NPCs) differentiated from these lines revealed a genotype-dependent graded reduction in H2AK119ub1. We demonstrate that reduced H2AK119ub1 occupancy in $RING1^{MS}$ NPCs is the result of a dominant-negative genetic mechanism where PRC1 harboring hypomorphic $RING1^{MS}$ binds target loci but fails to catalyze H2AK119ub1. We show that H2AK119ub1 dosage is a molecular criterion that determines its functional specificity. In NPCs, the timing of DNA damage repair is most sensitive to modest reductions in H2AK119ub1 dosage associated with a $RING1^{MS}$ allelic series. This neuropathology provides mechanistic insight into the phenocopy between the *RING1*-related disorder and Seckel syndrome. In contrast, substantial reductions in H2AK119ub1, associated with $RNF2^{MS}$ alleles, resulted in both delayed DNA repair and widespread transcriptional changes. This analysis identified surprising level of redundancy between the RING1 and RNF2 paralogues, that nevertheless results in

distinct *RING1*^{MS} and *RNF2*^{MS} pathology due to the functional specificity of H2AK119ub1 dosage. This research contributes to our understanding of the intricate interplay between PRC1 complexes, H2AK119ub1 biology, and their roles in neural development.

Session Title: Epigenetics Poster Wednesday Session

Board 1024W: Generation of hypothalamus with normalized gene expression of the responsible imprinting genes derived from Prader-Willi syndrome patient cells using epigenome editing technology.

Authors:

H. Okuno[1], N. Akisa[2], I. Kent[3,4], k. kenjiro[5], O. Hideyuki[6]; [1]Tokyo Med. Univ., Shinjyuku-ku, Tokyo, Japan, [2]Keio Regenerative Med. Res. Ctr., Shinjyuku-ku, Tokyo, Japan, [3]Keiol Univ., Shinjyuku-ku, Tokyo, Japan, [4]Keio Regenerative Med. Res. Ctr., Kanagawa, Japan, [5]Keio Univ. Sch. of Mediciine, Shinjyuku-ku, Tokyo, Japan, [6]Keio Univ., Shinjyuku-ku, Tokyo, Japan

Abstract:

Prader-Willi syndrome (PWS) is a genomic imprinting disorder caused by the loss of function of paternal chromosome 15q11-13, which results in a variety of symptoms associated with hypothalamic-pituitary dysfunction. PWS patients lack the expression of paternally expressed genes (PEGs) at the 15q11-13 locus, but there are epigenetically silenced versions of these genes in the maternal allele. Consequently, the activation of these silenced genes represents a potential therapeutic target for PWS. In order to modulate the DNA methylation status of the PWS imprinted regulatory region (PWS-ICR) in induced pluripotent stem cells (iPSCs) derived from PWS patients, we utilized a CRISPR-based epigenome editing system.The successful demethylation of the PWS-ICR resulted in the restoration of PEGs expression from the maternal allele and the reorganization of the methylation pattern of PWS-related imprinting regions downstream of the PWS-ICR. Notably, these modified epigenomic patterns and PEG expression were maintained after the differentiation of these cells into hypothalamic organoids. This finding suggests the possibility of cell transplantation therapy for PWS using epigenome editing. In order to gain further insight into the effects of disease-specific gene regions upon the hypothalamus-pituitary system, we have utilized hypothalamic-pituitary organoids differentiated from epigenome-edited iPS cells.

Session Title: Epigenetics Poster Wednesday Session

Board 1025W: Genetic regulation of methylation across East Asian and European populations

Authors:

R. Liu[1], T-T. Chen[2], Y. Xia[3], S-C. Lin[2], C-Y. Chen[4], Y-C. Feng[5], T. Ge[6], H. Huang[6], Y-F. Lin[2]; [1]Broad Inst., Cambridge, MA, [2]Natl. Hlth.Res. Inst.s, Miaoli, Taiwan, [3]Yale Univ., New Haven, CT, [4]Biogen, Lexington, MA, [5]Natl. Taiwan Univ., Taipei, Taiwan, [6]Massachusetts Gen. Hosp., Boston, MA

Abstract:

DNA methylation plays a vital role in regulating gene expression and has been implicated in a wide range of human diseases and conditions. Methylation quantitative trait loci (mQTLs) provide valuable insights into the genetic architecture of interindividual epigenetic variation. While studies have identified numerous mQTLs, the majority have focused on the European (EUR) population, limiting the generalizability of findings to other populations.
To address this limitation, we conducted the largest-to-date mQTL study in East Asia (EAS), encompassing 7,620 Han Chinese individuals from three cohorts, including Taiwan Biobank (TWB) and two recently published mQTL studies. This study surpasses the sample size of previous EAS mQTL studies by twofold. Our meta-analysis identified 52,989,871 mQTLs, including 313,117 mCpG and 5,676,958 variants in EAS ($P<5e-9$ and $het_P > 0.05$). Notably, 15% (47,651) of these mCpGs represent new discoveries within EAS. Compared to the largest EUR mQTL study using the less comprehensive 450K array, we identified 60% (187,673) additional new mCpGs, mainly attributable to the expanded CpG coverage of the EPIC array used in our study (78% of new findings).
Cross-population consistency analysis in mQTLs revealed a remarkably high degree of concordance, with over 99% of significant mQTLs identified in EAS also being detected in the previous EUR studies. Given this consistency, we performed a meta-analysis across EAS and EUR populations, which identified 93,517,902 mQTLs, including 368,688 mCpGs ($P<5e-9$). Among them, 32,384 (8.7%) mCpGs were novel and not previously reported in either population.
We further performed SMR to investigate causal relationships between mQTLs and complex traits/diseases in EAS. This analysis identified 2,099 significant CpG-trait pairs, including an association between cg21090344 (in the promoter region of SLC16A11) and type 2 diabetes risk. This finding aligns with the known roles of SLC16A11 in fatty acid and lipid metabolism pathways implicated in T2D pathogenesis.
This study significantly expands the map of EAS mQTLs, underscoring the shared and ancestry-specific aspects of epigenetic regulation. Our findings provide a valuable resource for dissecting the genetic basis of complex diseases and highlight the importance of including diverse populations in genomic research.

Session Title: Epigenetics Poster Wednesday Session

Board 1026W: Identification of circulating epigenetic biomarkers for hemorrhage in brain arteriovenous malformation by whole-genome methylome profiling

Authors:

S. Weinsheimer[1], J. Nelson[2], C. McCulloch[3], N. Ko[4], E. Winkler[5], H. Kim[6]; [1]Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Inst. for Human Genetics, Univ. of California San Francisco, San Francisco, CA, [2]Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Univ. of California San Francisco, San Francisco, CA, [3]Dept. of Epidemiology and Biostatistics, Univ. of California San Francisco, San Francisco, CA, [4]Dept. of Neurology, Univ. of California San Francisco, San Francisco, CA, [5]Dept. of Neurological Surgery, Univ. of California San Francisco, San Francisco, CA, [6]Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Inst. for Human Genetics, Epidemiology and Biostatistics, Univ. of California San Francisco, San Francisco, CA

Abstract:

Background: Brain arteriovenous malformations (bAVMs) are a tangle of abnormal vessels directly shunting blood from the arterial to venous circulation and an important cause of intracranial hemorrhage (ICH), seizure, or other neurological deficits in younger persons. Clinical heterogeneity of bAVM suggests a role for genetic modifiers that may associate with ICH risk. Comprehensive DNA methylome profiling has revealed biomarkers for cerebrovascular disease, cardiovascular disease and cancer, and may be a useful tool to identify disease monitoring biomarkers for bAVM ICH and novel therapeutic targets to stabilize bAVM lesions. We hypothesized that the methylome differs between bAVM cases who present ruptured vs. cases who present unruptured.

Methods: We selected 30 individuals with bAVM (n=19 unruptured, n=11 ruptured cases, 50% male) and genomic DNA from whole blood was profiled for genome-wide methylation using Whole Genome Bisulfite Sequencing services at Psomagen, Inc. (~20x coverage). Differentially Methylated Regions (DMRs) were identified by comparing the average methylation ratio in ruptured to unruptured cases by t-test assuming equal variance. In this discovery study, DMRs with a delta average methylation ratio between the two groups ≥ |0.2| and P ≤ 1x10-05 were considered significantly associated given the small sample size; DMRs with delta methylation ratio ≥ |0.5| and P ≤ 0.001 were considered both statistically significant and having a large effect size.

Results: Methylome profiles differed between bAVM cases presenting with and without rupture with 171 significant DMRs (33% hypermethylated). The largest methylation differences were observed for 37 CpG sites mapping to 30 genes (DMR ≥ |0.5| and P-value range 0.00072-0.000001), including 19 hypomethylated and 18 hypermethylated sites in ruptured cases which stratified patient groups. Interestingly, the top 10 differentially methylated genes include hypomethylated *KRAS*, *C1QTNF5*, *STK32A*, *ANKRD24*, and *STEAP3*; and hypermethylated *KHDRBS2*, *LINC00578*, *RTN4RL1*, *DACH2*, and *ZSCAN5A*.
Conclusions: We performed the first whole-genome methylome profiling of individuals with bAVM and identified several differentially methylated genomic regions associated with hemorrhage, including hypomethylation of *KRAS* which is consistent with known *KRAS* gene activation in bAVM. DNA methylation modifications of novel genes may serve as a new layer of biological regulation that contributes to bAVM hemorrhage. Further studies are needed to validate these findings in a larger bAVM cohort and investigate the role these epigenetic biomarkers have in bAVM disease.

Session Title: Epigenetics Poster Wednesday Session

Board 1027W: Identification of microRNA signatures carried by neuronal released extracellular vesicles in Alzheimer's Disease and their putative effect in altering target cell gene expression with population specificity

Authors:

K. Subasinghe[1], C. Hall[2], I. Gorham[1], R. Trivedi[1], A. Ranjan[1], J. Y. Choe[1], M. Rowe[1], J. K. Vishwanatha[1], Z. Zhou[1], R. Barber[1], H. Jones[1], N. Phillips[1]; [1]Univ. of North Texas Hlth.Sci. Ctr., Fort Worth, TX, [2]Johns Hopkins Univ., Baltimore, MD

Abstract:

Background. Alzheimer's disease (AD) is a progressive neurodegenerative disorder that disproportionately affects several racial/ethnic groups, including Mexican Americans (MAs) compared to their non-Hispanic white (NHW) counterpart. Early AD brain alterations are known to propagate to local and distal cells through extracellular vesicles (EVs) known as exosomes, which can cross the blood-brain barrier, and thus represent an easily accessible derivative of the brain in living humans. Exosomes carry bioactive cargo including microRNAs (miRNA) that act as strong regulators of gene expression in target cells. Oxidative stress (OS) is considered one of the earliest hallmarks of aging, hence biomarkers associated with OS may be viable early

indicators of AD risk. **Objectives.** This project aims to identify miRNA signatures relevant to AD status contained within EVs that are differentially released from neurons (1) into the blood of human subjects with cognitive impairment or AD, and (2) into the supernatant of neuronal cell culture under OS. **Methods.** (1) We processed plasma samples received from the Texas Alzheimer's Research and Care Consortium by precipitation of total exosomes followed by neuronal-enriched extracellular vesicles (NEEVs) capture with a biotinylated antibody against the neuronal surface marker, CD171. miRNAs were then profiled from NEEVs using next-generation sequencing (NGS) followed by the analysis for differential miRNA expressions between the cognitively impaired and normal control groups (2) Human neuronal epithelioma SK-N-MC cells were treated with hydrogen peroxide (H_2O_2) to mimic OS in vitro and the small EVs (sEVs) released into the supernatant were isolated; miRNAs from both the cells themselves as well as the sEVs were extracted and sequenced using NGS. Differentially expressed miRNAs were verified with quantitative PCR. **Results**. (1) We identified specific miRNA candidates differentially expressed in NEEVs from cognitively impaired subjects compared to healthy controls with population specificity (MA vs. NHW). These miRNAs target gene networks that have been implicated in AD pathophysiology. (2) We identified a miRNA profile indicative of H_2O_2 exposure in neuronal sEVs - six miRNAs exhibited overrepresentation and two miRNAs exhibited underrepresentation with increasing H_2O_2 treatment that target NF-κB signaling pathway associated genes. **Conclusion.** These miRNAs are involved in many cellular processes, including regulation of NF-κB complex associated genes and their related cellular responses indicating the presence of OS and subsequently alter local and/or systemic responses.

Session Title: Epigenetics Poster Wednesday Session

Board 1028W: Identifying myogenic enhancers and 3D genome regulation across human development

Authors:

M. Romero[1], P. Chien[1], C. Nicoletti[2], G. Cox[3], K. Saleh[1], D. Gibbs[1], L. Gane[1], D-H. Hoang[1], L. Caputo[2], P. L. Puri[2], A. D. Pyle[1]; [1]UCLA, Los Angeles, CA, [2]Sanford Burnham Prebys Med. Discovery Inst., San Diego, CA, [3]California State Univ. North Ridge, Northridge, CA

Abstract:

Stem cell treatments using skeletal muscle stem cells (or satellite cells, SCs) provide great potential for regenerating new muscle. A limitation is that adult SCs are an extremely rare population, making up ~1-2% of nuclei in skeletal muscle. To overcome this, we have developed directed differentiation strategies to generate skeletal muscle cells from human pluripotent

stem cells (hPSCs) and induced PSCs (hiPSCs). These differentiation strategies generate PAX7$^+$ skeletal muscle progenitor cells (SMPCs), however, SMPCs derived from either hiPSCs or hPSCs resemble early myogenic cells that align closer to week 7-12 in human development and are not equivalent to adult SCs in their regenerative potential. We sought to understand molecular and functional differences that control SMPC versus SC cell states to improve our differentiation strategies. There is an intense interest in the three-dimensional (3D) organization of the genome, specifically the communication between enhancers and their target genes in cell specific gene regulation. In humans, many of the enhancers that control myogenic gene expression are unknown. We sought to identify how the 3D genome regulates enhancer-promoter communication across muscle development. To achieve this, we performed high throughput chromosome conformation capture (Hi-C) and promoter capture Hi-C (pcHi-C) to identify enhancers across multiple myogenic cell states, including SMPCs derived from hPSCs and fetal muscle as well as adult SCs. We found genome-wide 3D configurations were different between all myogenic cells on multiple levels. We identified numerous myogenic TFs that displayed both compartment switching and increased expression. Moreover, TAD reorganization was seen at more than a 1K loci. Not all muscle specific loci were different between cell types, however, suggesting that some TADs are pre-established early in the differentiation process while others are established *de-novo*. By integrating pcHi-C and CUT&Tag for enhancer associated histone modifications, H3K4me1 and H3K27ac, we also identified human specific enhancers for PAX7, an SC specific TF. Using both CRISPR/Cas9 and CRISPR/dCas9a approaches, we validated the effect of these enhancers on PAX7 expression. These data, for the first time, characterize the 3D genome of human SMPCs and adult SCs These data provide enhancer sequences for that could provide unique candidates for support of PAX7 and other myogenic regulators of SMPCs and SCs. This work is supported by R01AR06432706AI (A.D.P.), R01AR064327-06AI:S1, the A.P. Giannini Foundation, Ford Foundation, and the UC Presidential Postdoctoral Fellowship Program. (M.A.R.).

Session Title: Epigenetics Poster Wednesday Session

Board 1029W: Investigating epigenetic alterations to the chromatin landscape in the progression of gastric-esophageal adenocarcinomas to peritoneal metastasis.

Authors:

K. White[1,2], J. Cavalcante de Moura[3], A. Gnanapragasam[4,5], Y. Yan[3], B. Giannias[3], V. Sangwan[1,2], L. Ferri[1,2], S. Bailey[5,3,6,1]; [1]McGill Univ., Montréal, QC, Canada, [2]McGill Univ. Dept. of Surgery, Montréal, QC, Canada, [3]McGill Univ. Hlth.Ctr. Res. Inst., Montréal, QC, Canada, [4]McGill Univ.,

Montreal, QC, Canada, [5]McGill Univ. Dept. of Human Genetics, Montréal, QC, Canada, [6]McGill Univ. Dept. of Experimental Med., Montréal, QC, Canada

Abstract:

BACKGROUND: Gastro-esophageal adenocarcinomas (GEAs) have extremely poor prognosis, with a five-year survival rate of less than 20%. Metastasis to the peritoneum is the most common progression of GEA and the leading cause of GEA-associated death. Epigenetic shifts in chromatin activity and accessibility accompany tumor progression. These epigenetic shifts reveal key regulatory elements required for tumor progression. Histone 3 lysine 27 (H3K27) acetylation (H3K27ac) and trimethylation (H3K27me3) represent active and repressed chromatin respectively. Non-coding elements, such as enhancers regulate gene expression and are commonly disrupted in malignancies. Variant enhancer usage between normal tissue and primary tumors are termed variant enhancer loci (VELs), and between primary tumor and metastatic samples termed metastatic variant enhancer loci (MET-VELs), have been reported to drive cancer initiation and progression. Investigating variant enhancer usage in the progression of GEA can yield valuable insights to better understand the metastatic phenotype, and potentially yield novel therapeutic targets. **METHODS:** To identify the gained and lost regulatory elements accompanying the progression of GEA to metastasis, we aim to perform ChIPmentation against H3K27ac and H3K27me3 of 20 patient matched patient trios (normal, primary GEA tumours, and peritoneal metastasis). Sequencing reads will be aligned using BWA and peaks will be called using MACS2. Candidate VELs and MET-VELs will be assessed using CRISPR interference and CRISPR activation. **RESULTS:** Preliminarily, we identified gains and losses at multiples sites that could suggest novel MET-VELs. Non-coding regions for nearby genes HOXB13 and SALL4 showed significant gains indicating upregulation specific to the metastatic phenotype. Whereas significant losses were noted nearby the CDHR3 gene. **CONCLUSION:** Epigenomic alterations during tumor progression can highlight key mechanisms resulting in aberrant gene expression currently absent from literature. Identification and investigation of these alterations can aid in the understanding of the disease and potentially lead to novel therapeutic interventions.

Session Title: Epigenetics Poster Wednesday Session

Board 1030W: Multi-modal inference of gene regulatory networks in systemic lupus erythematosus suggests ethnicity-specific differences in epigenetic regulation ⭐

Authors:

R. Pelikan, Y. Fu, J. A. Kelly, K. L. Tessneer, G. B. Wiley, K. Grundahl, P. M. Gaffney; Oklahoma Med. Res. Fndn., Oklahoma City, OK

Abstract:

Systemic lupus erythematosus (SLE) is a debilitating autoimmune disease. Recent studies have highlighted the role epigenetic regulation plays in controlling the inflammatory processes contributing to SLE. Blacks tend to develop SLE earlier, with more severe symptoms compared to Whites. To study how mechanisms of epigenetic regulation may contribute to this disparity, we performed deep multimodal profiling of the epigenomes and transcriptomes of peripheral blood mononuclear cells (PBMCs) from 40 total donors from the OMRF Lupus Registry and Repository, via single-cell ATAC (scATAC)-seq and scRNA-seq. Donors were equally represented by either Black or White SLE patients and healthy controls. We profiled multiple immune cell types, however, given the crucial role B-lymphocytes play in SLE pathogenesis, we highlight their properties in this work.

We used ArchR and cisTopic to process sequencing data and identify 60,719 B cells from a total set of 487,433 PBMCs. We used SCENIC+ to merge our scATAC and scRNA data to infer enhancer-driven gene regulatory networks (GRNs). The GRNs comprise “regulons” which correlate concerted enhancer activity with the genes they regulate, in accordance with their defining upstream transcription factors (TFs). B cells show stark ethnicity-specific differences in GRN topology and strength of their component regulons. We derived GRNs specific to three B cell compartments: naïve B cells, memory B cells, and a $CD11c^+$ $TBX21^+$ B cell subset previously described to be enriched in Black SLE patients. The proportional regulatory potential (pRP) of regulons was compared across GRNs by aggregating ranks of GRN edge evidence weight.

In naïve B cells, GRNs from White patients showed a 2.86-fold gain in pRP by IRF4, a regulator of lymphocyte fate change, and 1.49-fold loss of pRP by KLF2, a regulator of B cell activation, when compared to control GRNs. The GRNs of White SLE memory B cells showed a 1.47-fold gain in pRP by NFKB1, a central mediator of inflammation and, upon reaching the $CD11c^+$ $TBX21^+$ B cell state, IRF4 pRP is reduced 1.86-fold in patients relative to controls. Meanwhile, GRNs from Black patients also showed a 1.85-fold loss of KLF2 pRP, but gained 1.42-fold pRP through FOXO1, a biomarker for SLE activity. As Black SLE naïve B cells transition to memory B cells, the pRP of TFs in known SLE risk loci BACH1 and IKZF3 increase 1.26-fold. Finally, as these cells reach the $CD11c^+$ $TBX21^+$ state, pRP of ETV6, an inflammatory repressor, is increased 1.56-fold while that of E2F5, a B cell receptor trigger facilitator, is reduced 1.3-fold relative to controls. Regulatory elements highlighted by these GRN differences may yield clues to ethnic disparity in SLE.

Session Title: Epigenetics Poster Wednesday Session

Board 1031W: Peakverse:Robust suite of tools for prioritizing and visualizing single cell ATAC sequencing peaks

Authors:

S. Somasundaram, N. Johansen, A. Oster, M. E. Wirthlin, M. Hooper, E. Thomas, Z. Yao, B. Tasic, T. L. Daigle, Y. Ben-Simon, J. K. Mich, B. P. Levi, J. T. Ting, T. E. Bakken, J. A. Miller, E. S. Lein; Allen Inst. for Brain Sci., Seattle, WA

Abstract:

Assay of Transposase Accessible Chromatin (ATAC) sequencing is a popular method for determining chromatin accessibility across the genome that bind transcription factors or other DNA binding elements. These regions often correspond to enhancer elements that control the expression of genes. Single cell ATAC (scATAC) sequencing can be used to reveal the cell type-specific gene regulatory landscape that distinguishes cell types from each other. Recent work has shown that cell type-selective enhancers can be applied in adeno-associated viruses (AAVs) to drive gene expression selectively in the same cell types that showed chromatin accessibility for experimental study or gene therapy. Several suites of analysis tools have been developed for ATAC-seq analysis from alignment of single cells to clustering to identifying enriched enhancer regions (or “peaks”) of aligned reads for each cell type. Downstream software like DiffBind can quantify differential peak binding across cell types; however, hundreds to thousands of peaks can be selected and these tools fall short for prioritizing the "best" peaks for experimental testing and application.

To bridge this gap, we present a robust suite of easy to use, comprehensive and extendable tools called Peakverse in R for ranking, visualizing, and exploring cell type-specific enhancers and their properties. *PeakRankR* takes a series of predefined peaks for multiple cell types and prioritizes the best differential peaks for experimental validation. *Sorting and Visualizing Enhancers (SAVE)* provides a shiny-based interface for filtering, sorting, and exploration of prioritized peaks. *HubR* allows building tracks for the peaks to visualize in UCSC genome browser securely. *SeqFindR* aids in understanding the enhancers at the sequence level by highlighting the transcription factors in peak sequence. Together, this set of tools can help in experimentally testing optimum peaks faster and in building synthetic peaks for translational research. The open suite of tools can be found in this repository, https://github.com/AllenInstitute/Peakverse.

To demonstrate the utility of these tools, especially PeakRankR, we applied them to scATAC-seq data from mouse M1 where we had collected a set of manually selected peaks for each cell type. We show that PeakRankR could have identified most of the peaks with expected patterns and identify a set of additional peaks that we predict would have specific patterning. In fact,

PeakRankR did quite well in a recent BICCN challenge to predict enhancers competing with other sophisticated methods (https://biccnchallenge.org/data/).

Session Title: Epigenetics Poster Wednesday Session

Board 1033W: Single cell analysis of neurogenesis in adult human olfactory epithelium ⭐

Authors:

L. Song[1], C. Coleman[1], X. Wang[1], C. Casey[1], E. Hennigan[1], A. Iloreta[2], S. DeMaria[2], A. Kuno[1], J. Fullard[1], J. Bendl[1], P. Dong[1], P. Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Mount Sinai, New York, NY

Abstract:

The presence of neurogenesis in the adult human brain remains debatable. In the olfactory system, olfactory sensory neurons (OSNs) maintain homeostasis via continual neurogenesis throughout life, providing a niche to investigate neurogenesis in adults. However, the molecular mechanisms of this process are largely unknown. We performed single nuclear RNA-seq in 72 samples of human olfactory epithelium from 6 adult donors, yielding 147,607 high quality nuclei. We identified cell types critical for neurogenesis, including neuron precursor cells (globose basal cells, GBCs), as well as immature and mature olfactory neurons. We inferred lineage trajectories and assessed the transcriptional changes associated with GBC differentiation and OSN maturation. GBSs exhibited a higher cell cycling score than OSNs, indicative of active proliferation. In the GBC to neuron lineage, genes associated with the regulation of neuronal differentiation and neurogenesis were highly expressed in GBCs and early immature OSNs but were downregulated in mature neurons. In addition, we identified master regulators of neurogenesis, including SOX2, SOX4, and LHX9. Moreover, comparison of neurogenesis-associated cellular states and gene programs with prenatal neurogenesis revealed shared and distinct cellular and molecular changes between OSNs and prenatal neurons in the brain. Overall, profiling OSN-specific neurogenesis mechanisms may provide insight into how neurogenesis might be achieved in the adult central nervous system, opening new avenues for therapeutic intervention in neurodegenerative disorders.

Session Title: Epigenetics Poster Wednesday Session

Board 1034W: Spatiotemporal Twist1 deficiency shifts forelimb to hindlimb expression pattern resulting in multidigit phenotype

Authors:

R. Birnbaum[1], E. Gamliel[1], N. Hirsch[1], I. Dahan[2], O. Bompadre[3], G. Andrey[3], L. Taher[4]; [1]Ben Gurion Univ., Beer Sheva, Israel, [2]Ben Gurion Univ., BEER SHEVA, Israel, [3]Univ. of Geneva, Geneva, Switzerland, [4]Graz Univ. of Technology, Graz, Austria

Abstract:

Twist1 encodes a transcription factor that plays a vital role in limb development. Conditional *Twist1* knockout mouse models show specific skeletal phenotypes with partial penetrance and variable expression in the hindlimb versus the forelimb. However, the differential role of Twist1 in the forelimb versus hindlimb is yet to be investigated. By deletion of *Twist1* limb enhancers (*eTw5-7*$^{\Delta/\Delta}$) in mice, we showed that spatiotemporal reduction of Twist1 leads to multidigit phenotype in the hindlimb and polydactyly in the forelimb that that was not observed in *Twist1*$^{+/-}$ embryos. Using RNA-seq, we showed that the reduction of *Twist1* expression led to ectopic expression of hindlimb markers in the forelimb (*Pitx1, Tbx4, and Hoxc9-10*). We showed that Twist1 binds to the regulatory elements of the Pitx1 hindlimb marker. Pitx1 regulates the transcription of Tbx4 and is controlled by the caudal expression of the HoxC cluster. Using whole-mount in situ hybridization and qPCR, we showed that these hindlimb-specific markers are ectopically expressed in the forelimb of *eTw5-7*$^{\Delta/\Delta}$ E11.5 mouse embryos. We showed that Twist1 binds and regulates the Pitx1 enhancer activity in the limb. As Twist1 haploinsufficiency is associated with histone acetylation level, we found a specific reduction of acetylation level at the loci of Pitx1 in *eTw5-7*$^{\Delta/\Delta}$ E11.5 forelimbs. As Twist1 may involved in the unique chromatin conformation of the Pitx1 locus in hindlimb vs. forelimb, we showed a reduction in the chromatin interaction in *eTw5-7*$^{\Delta/\Delta}$ between Pitx1 promoter region and the distal regulatory elements in the hindlimb and not change in the forelimb. Moreover, we observed a feedback loop where Pitx1 and Tbx4 bind and regulate the activity of Twist1 enhancers. Altogether, deletion of Twist1 enhancers in mice elucidates a novel transcriptional regulatory function in the forelimb versus the hindlimb. This study suggests that *Twist1* has a spatiotemporal regulatory activity on hindlimb specific genes in the developing forelimb and plays a role in limb formation.

Session Title: Epigenetics Poster Wednesday Session

Board 1035W: Targeted cell-free DNA methylation haplotype signatures in prostate cancer progression ⭐

Authors:

J. Wong[1], Y. Tian[1], C-C. Huang[2], M. Kohli[3], L. Wang[1]; [1]H. Lee Moffitt Cancer Ctr., Tampa, FL, [2]UW Milwaukee, Milwaukee, WI, [3]Univ. of Utah Huntsman Cancer Inst., Salt Lake City, UT

Abstract:

A growing number of studies have demonstrated that cancer-derived DNA methylation is detectable in circulating cell-free DNA (cfDNA) of cancer patients. However, the utility of cfDNA methylation signatures as biomarkers in prostate cancer (PC) has not yet been established. In this study, we leverage prospectively collected blood samples in a cohort of PC patients to profile cfDNA methylation haplotype block (MHB) signatures across different stages of PC and assess the impact of these signatures on clinical outcomes. Plasma was collected from localized PC (n=21), metastatic hormone-sensitive PC (mHSPC) (n=27), and metastatic castration resistant PC (mCRPC) (n=48) patients. cfDNA was used for enzymatic methylation sequencing (EM-seq) with a target panel of 366 genomic regions implicated in PC biology. Methylation status at each CpG site within these regions were examined using Twist targeted cfDNA methylation assay. Bismark program was used for methylation calling. mHapTk program was used to identify MHBs and calculate methylation metrics at each MHB. Methylation block score (MBS), the level of successive methylation patterns, was assessed using Welch's t-tests as well as support vector machine (SVM) classifiers with leave-one-out cross-validation. Patient survival analysis was performed with multivariable Cox proportional hazards model and Kaplan-Meier analysis. From 366 genomic regions, we identified 995 MHBs total in the 96 PC patients. After excluding MHBs with <100 reads, 901 MHBs were used for subsequent analyses. When comparing localized PC to mHSPC, 1 MHB in *GAD1* gene was identified (false discovery rate, FDR<0.1). When comparing mHPSC to mCRPC, 448 MHBs were found to be significant. SVM models using the top 30 MHBs (FDR<0.1) showed 100% accuracy to classify between localized PC and mHSPC and an average of 81.3% accuracy to classify between mHPSC and mCRPC. Interestingly, a multivariable risk score built from the 30 MHBs also showed a significant association with disease survival. For mHSPC patients (up to 32 months follow-up), progression was observed in 0 of 13 low risk patients and 6 of 14 (42.9%) high risk patients (Log rank p=0.008). For mCRPC patients (up to 24 months follow-up), 19 of 24 (79.2%) mCRPC patients died in the high risk group, while only 1 of 24 (4.17%) patients died in the low risk group (Log rank p<0.0001, HR=37.06). cfDNA-based liquid biopsies provide a minimally invasive method to evaluate cancer-related DNA methylation changes in PC patients. Our results support the use of cfDNA methylation signatures in a select

panel of genomic regions as predictive and prognostic biomarkers for clinical management of PC.

Session Title: Epigenetics Poster Wednesday Session

Board 1036W: Tip60 and ventrolateral neurons (LNv): insights into alcohol-induced neuroadaptations in Drosophila

Authors:

C. Del Valle-Colón[1], M. J. Álvarez-Cortés[1], S. I. Morales-Cancio[1], A. Montes-Mercado[1], D. L. Figueroa-Vargas[1], N. L. Fuenzalida-Uribe[2], T. Chang[2], A. Seeds[2], J. L. Agosto[1], A. Ghezzi[1]; [1]Univ. of Puerto Rico, Río Piedras, San Juan, PR, [2]Inst. of Neurobiology, Med. Sci. UPR Campus, San Juan, PR

Abstract:

Alcohol use disrupts neurological, physiological, and behavioral processes, posing significant challenges to the population. This disruption often leads to the development of adaptations in the brain, resulting in alcohol tolerance, physiological dependence, and the associated Alcohol Use Disorder (AUD). Such adaptations, particularly within the neural circuits regulating sleep and wake cycles, intensify the complexities of alcoholism recovery. Recent studies provide evidence that modulation of gene expression is an important mechanism in the development of brain adaptations that produce alcohol-induced behavioral changes. However, many of the epigenetic mechanisms controlling transcriptional reprogramming in alcohol-induced neuroadaptations remain unexplored. In this study, we aim to determine the role of the histone acetyltransferase Tip60 during alcohol-induced disorders using a *Drosophila* model.
The *Drosophila* ventrolateral neurons (LNv), which express the sleep-inducing neuropeptide pigment dispersing factor (PDF), regulate behaviors such as circadian rhythms and sleep/wake cycles and have recently been involved in the development of alcohol tolerance in flies. Because of its role in the acetylation of histones, we hypothesize that Tip60 activity in LNv mediates alcohol-associated transcriptional adaptations that result in alcohol tolerance and sleep dysregulation. We use the UAS-GAL4 system to manipulate Tip60-RNAi expression within the LNv and analyze alcohol-induced behavioral alterations. Preliminary experiments show that alcohol exposure changes neuronal morphology and production of pigment dispersion factor (PDF) neuropeptide by the LNv. Using the INTACT method, we extract nuclear RNA from specific *Drosophila* cell types in this case, LNv. Nuclei from LNv neurons were epitope-tagged with a GFP targeting the nuclear membrane and then purified by immunoprecipitation using an anti-GFP antibody. 24 hours after exposure to alcohol, the mRNA was extracted from the

isolated nuclei and submitted for RNA sequencing. Whole genome differential expression analysis reveals that alcohol exposure induces significant changes in the expression of key regulators of neuronal activity within LNv. We found 253 genes differentially expressed genes between control and alcohol-exposed flies. These genes were highly enriched for the biological process of mitochondrial translation. Understanding the mechanisms of alcohol neuroadaptations can identify new therapeutic targets for AUD. Furthermore, insights from this study may inform human genetic research on the epigenetic factors influencing susceptibility to AUD.

Session Title: Epigenetics Poster Wednesday Session

Board 1037W: Transcription factor activation induces site-specific copy gain of the interferon receptor cluster ⭐

Authors:

J. Wilson[1,2], M. P. Ludwig[1,2], M. Galbraith[1,2], J. C. Black[1], K. Sullivan[2]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Linda Crnic Inst. for Down Syndrome, Aurora, CO

Abstract:

Transient site-specific copy gains (TSSGs) are an evolutionarily conserved phenomenon in which specific genomic loci undergo a transient increase in copy number during S phase in response to stress conditions or alterations in the local epigenetic environment prior to resolution in G2. Furthermore, TSSGs are frequently found in repetitive gene families and regions amplified in tumors, suggesting a potential role as predecessors to inherited copy number alterations. However, the connection between the cellular stress signaling and generation of the TSSG remains unclear. Here, we report a novel TSSG locus at the interferon receptor (*IFNR*) cluster on chromosome 21, as well as the identification of the signaling cascade leading to the induction. We stimulated cells with various activators of the innate immune response, screened for copy-number gains using fluorescent in-situ hybridization (FISH) and determined that, regardless of stimuli, 10-15% cells underwent TSSG at the *IFNR* locus, consistent with other TSSGs. Next, we used a combination of pharmacological and genetic experiments to determine the components of the innate immune response required to generate TSSGs at the *IFNR* locus. The results of these experiments identified a transcription factor (TF) as both necessary and sufficient for TSSG induction. Cut & Run sequencing analysis for this transcription factor under two stress conditions sufficient to induce TSSGs revealed a strong peak directly upstream of the *IFNR* cluster in both. Furthermore, pharmacological inhibition of the TF prevents this binding and suppresses TSSGs. Cut & Run sequencing for histone marks and DNA replication

factors additionally revealed this to be a likely enhancer region that undergoes inducible replication factor binding under stress conditions. Together, these data suggest that TF activation and binding to the locus is required for TSSG generation at the *IFNR* locus and enabled us to develop a method for the identification of additional innate immune-activated TSSG sites.

Session Title: Epigenetics Poster Wednesday Session

Board 1038W: Using DNA methylation to investigate phenotypic heterogeneity

Authors:

C. Beck[1,2,3], E. H. C. Kovács[4], M. E. Gaine[5,6], J. A. Wemmie[6], V. A. Magnotta[6,7], C. T. Pinnaro[1,2], B. W. Darbro[1,3,2]; [1]Stead Family Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA, [2]Interdisciplinary Graduate Program in Genetics, Univ. of Iowa, Iowa City, IA, [3]Shivanand R. Patil Cytogenetics and Molecular Lab., Univ. of Iowa, Iowa City, IA, [4]Dept. of NeuroSci., Univ. of Iowa, Iowa City, IA, [5]Dept. of Pharmaceutical Sci. and Experimental Therapeutics (PSET), Coll. of Pharmacy, Univ. of Iowa, Iowa City, IA, [6]Dept. of Psychiatry, Univ. of Iowa, Iowa City, IA, [7]Dept. of Radiology, Univ. of Iowa, Iowa City, IA

Abstract:

Background: DNA methylation, a known regulator of disease, has great influence on gene expression but is not accounted for in standard gene burden analyses (GBA). GBA is used to measure phenotypic heterogeneity (PH) in genetic diseases. We propose the first GBA that includes genetic information alongside DNA methylation. As phase one for this project, our goal is to create a novel weighting scheme (burden metric) that can be used with DNA methylation array data. **Methods:** To identify the most relevant CpG sites for GBA, we first filtered CpG sites according to tissue-based methylation variability of the locus, variation attributed to change in nucleotide (and not methylation), and presence or absence of encompassing repetitive sequence. We further intersected this list of CpG sites with those present on the Illumina 450k, EPICv1, and EPICv2 arrays. We then scored each CpG site with our novel Burden Estimate from Weighted Integration of Site-specific Epigenetic changes (BeWISE) where each CpG is weighted according to its presence within a functional region of the genome. **Results**: Our filtered list included ~65,000 sites attributed to 12,007 genes and contained CpG sites enriched for promoter, enhancer, and insulator regions, highlighting the connection between gene regulation and DNA methylation. To test the ability of the BeWISE score to differentiate affected from unaffected individuals, we utilized a cohort with bipolar disorder (BD) (n = 122) and controls (n =

69) as previously published results implicate genetics and epigenetics in the pathogenesis of BD. Using our BeWISE score aggregated by gene, we trained a random forest classifier to distinguish BD and controls with over 80% accuracy. The most important gene identified by GINI index was *PRKAA2* which is a catalytic subunit of CAMKK2; the CAMKK2 pathway is known to be associated with bipolar disorder. **Conclusions:** The BeWISE score can delineate between cases and controls with high accuracy and return meaningful results. Using this BeWISE score combined with genetic information may help explain PH of a wide range of genetic diseases.

Session Title: Epigenetics Poster Wednesday Session

Board 1039W: Whole Genome Methylation Profiling to Enhance Diagnostic Yield in Neurodevelopmental Disorders

Authors:

M. Locher[1], A. Risch[1], M. Auburger[1], S. Kleinle[1], M. Witzel[1], B. Sadikovic[2], A. Hallermayr[1], T. Neuhann[1], A. Abicht[1]; [1]MGZ Munich, München, Germany, [2]LHSC, London, ON, Canada

Abstract:

Background Epigenetic analysis in clinical diagnostics offers novel insights into complex genetic disorders, particularly neurodevelopmental delay (NDD) with or without congenital anomalies. We implemented an advanced whole genome methylation profiling technique, called EpiSign, to detect and characterize epigenetic signatures in patients with unsolved NDD enabling the reclassification of variants of uncertain significance (VUS). **Methodology** We validated whole genome methylation pattern analysis based on Illumina EPIC v2.0 BeadChips followed by subsequent epigenetic signature analysis using the EpiSign v5 analysis service (LHSC, Canada). **Results** In our cohort of 60 control cases (including NDD patients with causative class 4 or 5 variants in genes associated with NDD) we were able to establish and reproduce diagnoses with 93% sensitivity and 100% precision. Subsequently, we analyzed 29 unsolved cases with syndromic and non-syndromic NDD. From this group, 8 cases presented with VUS and 21 cases had no variant in a disease-associated gene. Four of the 8 patients with VUS had a gene-specific methylation pattern consistent with the epigenetic signature of the gene in which the VUS was identified. Three of the 21 unsolved patients without a known VUS were found to have specific episignatures indicative of distinct underlying conditions. By re-evaluation of exome data, a likely pathogenic variant was identified in one of these patients. **Conclusion** In the context of NDD, typical diagnostic resolution rates range between 20% and 40%. By the analysis of epigenetic signatures we were able to identify disease-associated episignatures in 24% of selected unsolved cases. A targeted re-evaluation of exome data or prior known VUS

significantly improved the diagnostic yield of NDD patients. Our findings underscore the pivotal role of epigenetic profiling in clinical genetic diagnostics, highlighting its integration as a crucial component of routine evaluations. This approach not only enhances our understanding of neurodevelopmental disorders (NDD) but also significantly improves clinical management, paving the way for more effective patient care.

Session Title: Epigenetics Poster Wednesday Session

Board 1040W: Whole genome methylation sequencing in blood from persons with mild cognitive impairment and dementia due to Alzheimer's disease identified extensive differences that are sex-specific and predictive of cognitive status

Authors:

L. Papale[1], A. Madrid[1], P. Bergmann[1], A. Boruch[1], S. Faasen[1], C. Breen[2], L. Clark[3], S. Asthana[3], S. Keles[4], K. Hogan[3], R. Alisch[1]; [1]Univ. of Wisconsin - Madison, Madison, WI, [2]UW Madison, Madison, WI, [3]Univ. of Wisconsin, Madison, WI, [4]Univ Wisconsin Madison, Madison, WI

Abstract:

Late-onset dementia due to Alzheimer's disease (AD) is the most common neurodegenerative disease that includes a prodromal stage, mild cognitive impairment (MCI). Evidence suggests that environmental influences increase the risk of AD in part through interactions with the epigenome, particularly DNA methylation. Using whole genome methylation sequencing to interrogate the entire human blood methylome (over 25 million sites), identified 28,038 differentially methylated positions (DMPs) in AD. Here we hypothesize that differential DNA methylation levels play a role in prodromal stages of AD. To test this hypothesis, we extend this research by integrating whole genome methylation data from the blood of MCI participants (*N*=99) into a three way pairwise model with AD (*N*=109) and cognitively unimpaired (CU, *N*=174) participants and found more than 1,500 differentially methylated genes in each pairwise comparison: 1,743, MCI *vs.* CU; 2,929 AD *vs*. MCI; and 1,847, AD *vs*. CU. Pathways analysis of differentially methylated genes between all comparisons shows enrichment for biologically relevant terms related to AD such as synapse organization, regulation of neurotransmitter levels, and ion transmembrane transport. In addition, we identified enrichment for mitochondrial-related genes, including glucose/carbohydrate metabolism (MCI *vs*. CU), oxidative stress (AD *vs*. MCI), and fatty acid metabolism/beta-oxidation (AD *vs*. CU), as well as epigenetic processes (*e.g.*, histone posttranslational modification genes). Since

AD has a sex-biased incidence, with females comprising nearly two thirds of all cases, we next investigated sex specific DNA methylation levels and found that less than 26% in MCI and 20% in AD of differentially methylated genes were shared between both sexes and AD. Moreover, pathways analysis only found an enrichment of terms related to cognitive decline among female-specific genes from all comparisons. Finally, we employed machine learning algorithms and unsupervised hierarchical clustering to identify the top twenty most distinguishing differentially methylated positions, which were able to stratify all 391 participants by diagnostic group. Together, these data highlight extensive differential blood DNA methylation levels in MCI and AD on genes with known and novel roles in cognitive decline that are sex-specific and could be useful as candidate biomarkers to predict cognitive status.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3002W: A multidisciplinary approach to understand the prehistorical societies of Qatar

Authors:

S. Tomei[1], F. Sakal[2], A. D'Aurelio[1], M. Somel[3], F. Al-Naimi[2], R. A. Carter[2], F. Castorina[4], C. Martinez-Labarga[5]; [1]Sidra Med., Doha, Qatar, Qatar, [2]Qatar Museums, Doha, Qatar, Qatar, [3]Middle East Technical Univ., Ankara, Turkey, Turkey, [4]Univ. of Rome Sapienza, Rome, Italy, Italy, [5]Univ. of Rome Tor Vergata, Rome, Italy, Italy

Abstract:

Background. The Middle East and Qatar sit at the crossroads of the Old World, where Africa, Europe and Asia meet. This geographical location has bestowed significant importance upon the Middle East as a key contributor to human evolution. The Middle East has some of the earliest evidence of modern humans outside Africa, including fossils date ~180kya (thousand years ago) and ~85 kya retrieved from the Levant and Northwest Arabia, respectively. The Arabian Gulf constitutes a unique potential source of evidence on human prehistory since it is considered the earliest corridor of “sapiens” out of Africa. Yet, the prehistoric societies and population movements of Middle Easterners are only marginally understood. **Methods**. Here we apply a multidisciplinary approach that combines a broad range of archaeological, morphological and molecular analyses to expand our current knowledge of prehistorical Qatari societies during the Neolithic to the Late Iron Age. The remains retrieved from 64 individuals underwent a detailed morphological assessment. Bone/teeth samples were submitted for bioapatite dating, Sr, C and O stable isotope analyses. Enamel samples were subjected to Strontium stable isotope analysis

along with enamel fauna samples and soil samples for the study of paleo-mobility. Additionally, we performed aDNA analysis on 20 samples, selected by taking into account preservation status and availability of either tooth or petrous bone. **Results**. Here we prove the feasibility to perform anthropological, molecular, and stable isotope analyses on the material retrieved from excavating Neolithic burials in Qatar.According to bioapatite dating, our samples cover a transect of time spanning from the Neolithic to the Late Pre-Islamic period (6.5-1.5 kya). We report archaeological, morphological and stable isotope data. Additionally, we provide quality control scores on the sequencing data and their correlation with sample type, burial location, and burial time frame. This work was supported by a grant from the Qatar National Research Fund (NPRP10-0208-170411). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Qatar National Research Fund.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3003W: A population-specific pangenome reference improves variant discovery in Pacific Islanders

Authors:

D. Stucki[1], C. Littlefield[1], J. M. Lazaro-Guevara[2], M. Lansford[1], M. H. Pezzolesi[1], E. J. Taylor[1], J. Taloa[1], K. Lao[1], J. P. Tavana[3], W. L. Holland[1], K. L. Raphael[1], M. G. Pezzolesi[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]McGill Univ., Montreal, QC, Canada, [3]Brigham Young Univ., Provo, UT

Abstract:

Pangenomes provide a comprehensive representation of genetic diversity by incorporating the full spectrum of genetic variation from multiple diverse individuals, enhancing the representation of genetic variation and improving the accuracy of genomic research. This study describes the improved variant detection in whole-genome sequences from individuals of Native Hawaiian and Pacific Islander ancestry using a new human pangenome derived from long-read sequencing data from people with Pacific ancestry and genetically diverse individuals from the Human Pangenome Reference Consortium (HPRC) dataset. Our objective was to enhance the representation of people with Pacific ancestry to improve variant detection accuracy. Using PacBio's HiFi long-read sequencing technology, we sequenced and assembled genomes from 23 individuals with Pacific ancestry, complementing the genetic variation captured by the 47 individuals in the initial HPRC dataset. We then constructed a unified pangenome (P-HPRC pangenome) by combining these new sequences with the HPRC reference. To evaluate the efficacy of the P-HPRC pangenome, we used Giraffe, a highly efficient mapping algorithm designed for pangenome graphs, and DeepVariant, a variant caller leveraging deep

learning, to call genetic variants in a cohort of 200 Pacific Islanders. Variants were then annotated with ANNOVAR. We compared the variants identified using the P-HPRC pangenome against those detected using the HPRC pangenome and the linear Telomere-to-Telomere CHM13 (T2T-CHM13) reference genome. Our analysis demonstrated that the P-HPRC pangenome significantly increased the detection of both common, low-frequency, and population-specific variants, providing a more detailed and accurate representation of the genetic diversity observed in Pacific Islanders. These results underscore the importance of incorporating diverse genetic data to enhance variant discovery and improve the utility of reference genomes in genomic research. This unified P-HPRC pangenome serves as a valuable resource for understanding the unique genetic predispositions and health outcomes of Pacific Islanders.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3004W: Ancestral origins and post-admixture adaptive evolution of highland Tajiks

Authors:

J. Wen[1], J. Liu[1], Q. Feng[2], Y. Lu[1], K. Yuan[2], X. Zhang[2], C. Zhang[2], Y. Gao[1], X. Wang[2], D. Mamatyusupu[3], S. Xu[1]; [1]Fudan Univ., Shanghai, China, [2]Shanghai Inst. of Nutrition and Hlth., Chinese Academy of Sci. (CAS), Shanghai, China, [3]Xinjiang Univ., Xinjiang Uyghur Autonomous Region, China

Abstract:

Tajiks are a highland group underrepresented in studies on high-altitude adaptation, in contrast to the vast number of genetic studies on Tibetans, Andeans, and Ethiopians. Here, we conducted a whole-genome deep-sequencing study of Xinjiang Tajiks (XJT) residing on the Pamir Plateau at an average altitude that exceeds 4,000 meters. Compared with other extensively studied highlanders, XJT showed more pronounced admixture patterns: most of their ancestry is derived from West Eurasians (34.5-48.3%) and South Asians (21.4-40.0%), and they have some minor ancestry (3.62-17.5%) from East Asians and Siberians. We observed greater genetic diversity in XJT than in their ancestral source populations, providing a genetic basis for the adaptation of XJT to high-altitude environments. In particular, admixture-induced adaptation was strongly associated with skin-related candidate genes that respond to UV radiation (e.g., *HERC2* and *BNC2*) and cardiovascular system-related genes (e.g., *MPI* and *BEST1*). Notably, no adaptive variants of genes showing outstanding natural selection signatures in the Tibetan or Andean highlanders were identified in XJT, including *EPAS1* and *EGLN1*, indicating that a

different set of genes contributed to XJT survival on the Pamir Plateau. Nonetheless, some genes underlying natural selection in XJT have been previously reported in highland regions. Our results suggest that convergent evolution might be prevalent in indigenous groups from different highlands during their adaptation to high-altitude environments.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3005W: Assessment of a microhaplotype panel for human identification, population structure and ancestry inference in Brazil

Authors:

C. Mendes-Junior[1], L. A. G. Kobachuk[2], V. M. S. Moraes[2], T. M. T. Carratto[1], L. Marcorin[2], A. L. Simoes[2], H. Norton[3], E. J. Parra[4], E. C. Castelli[5]; [1]Univ.e de São Paulo (FFCLRP-USP), Ribeirão Preto, Brazil, [2]Univ.e de São Paulo (FMRP-USP), Ribeirão Preto, Brazil, [3]Univ Cincinnati, Cincinnati, OH, [4]Univ. of Toronto, Toronto, ON, Canada, [5]Unesp, Botucatu, Brazil

Abstract:

Microhaplotypes (MHs) comprise multiple single nucleotide polymorphisms (SNPs) in DNA segments of up to 300 bp in length, resulting in multiple allelic combinations (i.e., haplotypes). They represent an emerging type of genetic marker with the potential to be an alternative to Short Tandem Repeats (STRs), SNPs and InDels in population genetic studies or in forensic cases involving degraded DNA or mixed contributors. Here we present the first study with an MH panel in an admixed Brazilian population sample. We evaluated a set of 128 MH markers in 344 individuals from Ribeirão Preto (SP) region aiming at its application for human identification and ancestry inference. DNA was extracted using the salting-out protocol and the quality of the genomic DNA samples was evaluated by agarose gel electrophoresis, spectrophotometry, and fluorimetric assay, respectively. The DNA samples were submitted to a genotyping assay with 1,748,250 SNPs using the Infinium Multi-Ethnic Global-8 array (MEGA) Kit (Illumina) and the HiScanSQ System (Illumina Inc.) platform. The Sanger and Michigan imputation tools were used to determine the genotypes of SNPs missing in the MEGA assay, using the 1000 Genomes Phase 3 as the reference panel for imputation. The linkage disequilibrium analysis showed that 108 out of the 128 microhaplotypes were independently inherited, and the combined match probability for these 108 MHs was 4.4×10^{-81}. A comparison of the forensic parameters with those estimated from currently used forensic STR kits that include the 20 CODIS showed that 108 MHs are much more efficient in identifying individuals than the STR sets. PCA and STRUCTURE analyses were conducted to assess genetic differentiation among African (AFR), East Asian (EAS), Native American (NAM), and European (EUR) population groups from 1000

Genomes and HGDP. The first three components and four STRUCTURE clusters clearly differentiate the populations from the four biogeographic regions. For the Brazilian population sample, a tetra-hybrid admixture model did not reveal significant differences in the AFR and EUR ancestries when compared with a genome-wide set of ~300,000 SNPs. However, NAM and EAS were slightly underestimated and overestimated, respectively. Considering the six admixed populations from 1000 Genomes, a tri-hybrid admixture model considering AFR, NAM, and EUR ancestries resulted in proportions that are in agreement with those obtained by a previous GWAS analysis. We conclude that this microhaplotype panel could offer a powerful and reliable complementary tool in forensic and population genetics applications, both for human identification and ancestry inference purposes.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3006W: Balancing selection and the malaria parasite genome: Insights into human-parasite co-evolution through genetic variation

Authors:

A. Forster[1], W. L. Hamilton[2], J. Docker[1], R. Pearson[2], K. Rockett[3], E. Leffler[4], J. Hendry[5], T. N. Williams[6], A. Mentzer[1], A. Amambua-Ngwa[7], L. N. Amenga-Etego[8], G. Band[9]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Wellcome Sanger Inst., Cambridge, United Kingdom, [3]Univ. of Oxford, Oxford, Oxon, United Kingdom, [4]Univ. of Utah, Salt Lake City, UT, [5]Max Planck Inst. of Infection Biology, Berlin, Germany, [6]Kenyan Med. Res. Inst. (KEMRI), Kilifi, Kenya, [7]MRC Unit The Gambia, Serrekunda, Gambia, [8]West African Ctr. for Cell Biology of Infectious Pathogens, Accra, Ghana, [9]Univ. of Oxford, Oxford, Oxfordshire, United Kingdom

Abstract:

The sickle haemoglobin polymorphism (HbS) is among the most notable examples of balancing selection in the human genome. Postulated over half a century ago, this abnormal form is thought to be maintained due to the protection against severe malaria it confers to carriers. Despite this, the dynamics which affect this process remain broadly unclear, with imperfect correlations between HbS frequencies and malaria endemicity observed globally. Recently it was discovered that protein-altering mutations in four regions of the malaria parasite genome - termed Plasmodium falciparum sickle-associated (*Pfsa*)1-4 - are strongly associated with HbS. It appears that '*Pfsa*+' mutations enable parasites to infect, grow, and cause disease in HbS-carrying individuals (an 'HbS-protection escape' phenotype). By meta-analysing across 6,311 samples of paired parasite-human sequence data from the geographically dispersed countries Gambia, Kenya and Ghana we resolve how the magnitude of resistance conferred by HbS

depends on parasite genotype. We find that the protective effect is completely lost in the context of parasites carrying multiple *Pfsa*+ alleles. However, we also observe significant spatial variation in the distribution of *Pfsa*+ mutations, with *Pfsa*2+ not present within Western and Central African populations and *Pfsa*4+ absent from East African regions, as well as additional population-specific signals. The *Pfsa*+ mutations are not found in Asian parasites, despite the presence of HbS and *Plasmodium falciparum* malaria. We use the MalariaGEN Pf7 whole genome sequence resource to show that the haplotypes carrying *Pfsa*+ mutations carry multiple mutations in high linkage disequilibrium with the core *Pfsa*+ SNPs, and are shared between African populations, strongly implying a single mutational origin at each locus. Further supporting this, we develop methods to call structural variation that exists at the *Pfsa*3+ locus, showing that *Pfsa*3+ mutations are carried on a set of related complex structural variant haplotypes. A natural assumption is that these *Pfsa*+ haplotypes are positively selected for by HbS. However, they have not reached 100% frequency in any population, and we observe only limited evidence for recent positive selection, raising questions as to what additional evolutionary forces are at play. Taken together these results suggest that, like HbS, the *Pfsa* polymorphisms may have been maintained under balancing forces over long evolutionary timescales, but the full set of evolutionary forces and the underlying biological mechanisms are yet to be determined.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3007W: Characterization of the coronary artery disease-associated 9p21 locus across ancestries, and its role in genetic susceptibility across populations

Authors:

H. Alkhairo[1], K. Iyer[2], K. Ito[3], S. Koyama[4], F. Alkuraya[5], N. Risch[6], C. Tcheandjieu[7]; [1]UCSF, San Francisco, CA, [2]J David Gladstone Inst.s, San Francisco, CA, [3]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [4]Broad Inst., Cambridge, MA, [5]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [6]Univ California San Francisco, San Francisco, CA, [7]Gladstone Inst., San Francisco, CA

Abstract:

Introduction: The first coronary artery disease (CAD)- associated locus, 9p21, was simultaneously discovered in 2007 in three separate genome-wide association studies (GWAS) of European ancestries. In subsequent years, locus 9p21 has been replicated in several non-European ancestries, however has yet to be replicated in African ancestry despite theoretically

sufficient sample sizes. To further understand the observed lack of replicability in African-ancestry as well as the general background structure of locus 9p21, our work focuses on investigating haplotype structure across multiple global ancestries.
Methods: Utilizing six different cohorts (Million Veterans Program, UK Biobank, H3Africa, Biobank Japan, All of Us, King Faisal SHRC), across multiple ancestries and ~1.8 M individuals, our study examined the haplotype structure at locus 9p21 to identify variants and haplotypes that carry CAD risk, as well as association with other cardiometabolic traits that share genetic risk with CAD risk factors. Only genotyped CAD-associated SNPs at 9p21 were considered for haplotype construction and estimation using the haplo.stats package in R. Utilizing a generalized linear model framework, haplotype associations with CAD and cardiometabolic risk factors were tested adjusting for sex, age and principal components 1-5.
Results: We first showed that the LD structure and haplotype distribution at 9p21 varies across European, East Asian, South Asian, Middle Eastern, and African populations, most dramatically between African and non-African ancestry individuals. Furthermore, the observed haplotypes at 9p21 carry risk for both CAD and cardiometabolic risk factors in all populations except those of African ancestry. The two most common haplotypes in non-Africans, AAAAA (non risk alleles) and BBBBB (all risk alleles) (mean frequencies 0.28, 0.42 respectively) exhibit consistent CAD risk differences in all non-African ancestries (mean frequencies 0.28, 0.42 respectively in non-African ancestries), but are virtually absent in African-ancestry individuals. The dramatic difference in LD and MAF between Africans and non-Africans at 9p21 explains the population divergence in CAD risk at this locus.
Conclusion: Characterizing locus 9p21 is the beginning of a larger effort to examine CAD-associated genomic regions that differ in haplotype structure between ancestries, and their impact on PRS transferability for CAD. This work serves as a critical component in understanding factors contributing to replication of genomic loci across populations.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3008W: Comparative genomic analysis of differential gene regulation using multimodal data in dozens of human and chimpanzee cell types

Authors:

H-C. Huang[1], K. Barr[2], Y. Gilad[3]; [1]Univ. of Chicago, chicago, IL, [2]Univ Chicago, Chicago, IL, [3]Univ. of Chicago, Chicago, IL

Abstract:

Comparative genomic studies in humans and other primates are essential for identifying conserved and divergent gene regulatory features. When differences in gene regulation are identified in the relevant functional contexts, they can provide insight on human-specific adaptations and disease mechanisms. In a previous study, we assessed gene regulatory conservation and divergence in 86 cell types from humans and chimpanzees using an iPSC-derived heterogeneous differentiated culture (HDC) model. Our work resulted in the largest catalog of comparative gene expression in humans and chimpanzees to date; however, our approach did not allow us to infer the regulatory mechanisms responsible for those differences. In this study, we collected paired single-nucleus RNA-seq and ATAC-seq data from human and chimpanzee HDCs to identify inter-species gene expression level changes potentially explained by differences in chromatin accessibility, transcription factor binding, and cis-regulatory architectures. Additionally, we collected multimodal data from a fused human-chimpanzee allotetraploid HDC, allowing us to assess the contributions of *cis* and *trans* changes to differential gene regulation between the two species. Our results, derived from 7 chimpanzee and 9 human HDCs (86,000 and 49,000 nuclei, respectively) and 1 fused HDC (40,000 nuclei), show comparable cell type complexity to our previously collected single-cell RNA-seq data. Across cell types and species, we identified 526,475 high-quality open chromatin regions. By comparing gene expression and chromatin accessibility data between humans and chimpanzees, we identified thousands of inter-species differentially accessible (DA) chromatin regions at orthologous sites. These findings present significant potential for correlating DA regions with downstream gene expression, allowing us to infer gene expression robustness and the functional redundancy of DA regions. To further elucidate the regulatory logic, we performed transcription factor (TF) binding motif analysis on DA regions and correlated these with TF expression. Our ongoing analyses leverage these correlations to elucidate the molecular basis of human-chimpanzee divergence. By addressing the longstanding challenge of identifying causal regulatory loci within a comprehensive and context-rich framework, our study represents a significant step forward in unraveling the complexities of gene regulation evolution between humans and our closest evolutionary relatives.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3009W: Correlated trait effects of nearby variants reflect phenotypic selection and mutational architecture

Authors:

E. Koch[1], M. Moldovan[1], C. Chiang[2], A. Durvasula[3], M. Zhang[4], A. Price[5], S. Sunyaev[1]; [1]Harvard Med. Sch., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Univ. of Southern California, Los Angeles, CA, [4]Carnegie Mellon Univ., Pittsburgh, PA, [5]Harvard Sch Pub Hlth, Boston, MA

Abstract:

Negative selection shapes genetic architectures of complex traits by preventing large-effect variants from reaching high frequencies (Gazal et al. 2018 Nat Genet). It has been proposed that stabilizing selection, which disfavors phenotypic extremes, generates this negative selection. Stabilizing selection models broadly fit the relationship of allele frequencies and effect sizes (Simons et al. 2022 bioRxiv), but inferences largely reflect the increasing strength of selection with effect size, and it is unclear which models of mutation and selection are consistent with observed selection signatures in GWAS data.

Linkage disequilibrium (LD) patterns may discriminate between models so we developed a population genetics approach to model correlated effect sizes. A recent method estimated negative SNP-pair effect correlations (SPECs) between nearby derived alleles (-0.37, 0-100bp, LDSPEC; Zhang et al. 2023 medRxiv) and within functional elements. These match expectations for stabilizing selection as haplotypes with opposite-effect alleles are overrepresented if masking allows them to escape negative selection. We investigated whether stabilizing selection is the only realistic explanation, and if mutational bias and asymmetric selection are also plausible. Simulations indicate our modeling approach is flexible and accurate (< 15% error for rare variant pairs at frequencies 1-10%).

In a simple model of stabilizing selection, SPECs < -0.37 can occur even under weak selection and magnitudes increase with the strength of selection and MAF threshold. SPECs are zero when selection is independent of variant effect directions and mutation is symmetric, but can be negative under mutational bias and selection asymmetry within a narrow range of selection coefficients. Non-stabilizing models are also sensitive to the fraction of zero-effect mutations. SPECs generally disappear when this fraction exceeds 90% but can be rescued in stabilizing models when causative sites are tightly clustered. Given that polygenicity estimates are rarely over 5% for any trait (Zeng et al. 2021 Nat Comm) and SNPs differ widely in selective effects (O'Connor 2021 Nat Genet), only stabilizing selection is consistent with estimated SPECs.

Beyond pointing to stabilizing selection, our analysis suggests a mutational architecture where most heritability localizes to regions where mutations have similar effects and point in opposite directions, a prediction amenable to experiment with mutational scanning techniques. Careful modeling of LD between causative GWAS variants can therefore support particular models of phenotypic selection and make testable predictions.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3010W: Digital image processing to detect adaptive evolution

Authors:

M. Amin, M. Hasan, M. DeGiorgio; Florida Atlantic Univ., Boca Raton, FL

Abstract:

In recent years, advances in image processing and machine learning have fueled a paradigm shift in detecting genomic regions under natural selection. Early machine learning techniques employed population-genetic summary statistics as features, which focus on specific genomic patterns expected by adaptive and neutral processes. Though such engineered features are important when training data is limited, the ease at which simulated data can now be generated has led to the recent development of approaches that take in image representations of haplotype alignments and automatically extract important features using convolutional neural networks. Digital image processing methods termed α-molecules are a class of techniques for multi-scale representation of objects that can extract a diverse set of features from images. One such α-molecule method, termed wavelet decomposition, lends greater control over high-frequency components of images. Another α-molecule method, termed curvelet decomposition, is an extension of the wavelet concept that considers events occurring along curves within images. We show that application of these α-molecule techniques to extract features from image representations of haplotype alignments yield high power and accuracy to detect hard and soft selective sweep signatures from genomic data with both linear and nonlinear machine learning classifiers. Moreover, we find that such models are easy to visualize and interpret, with performance rivaling those of contemporary deep learning approaches for detecting sweeps.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3011W: Enhancing rare variant discoveries with population genetics insights

Authors:

M. Gagnon[1], C. Moreau[1], J. Ricard[2], M-C. Boisvert[2], A. Bureau[2], M. Maziade[3], S. Girard[1]; [1]Université du Québec à Chicoutimi, Chicoutimi, QC, Canada, [2]Université Laval, Québec, QC, Canada, [3]Ctr. de recherche IUSMQ, Quebec, QC, Canada

Abstract:

No universal definition of what a rare disease is exists, so estimating their prevalence can be strenuous. The worldwide prevalence is estimated to be between 4% and 10% depending on what they consider to be a rare disease. Those numbers vary drastically when you narrow your research and start looking at regional prevalences. The province of Quebec (Canada) is defined by a unique demography shaped by founder effects. Founder events impacted the genetic diversity within the province and its regions, thereby increasing the frequency of certain rare variants. Carrier tests are offered for four rare diseases for people originating from Saguenay and Côte-Nord, but other regions have not been studied enough to justify similar programs, despite the presence of regional rare diseases. We explored the genetics of the Beauce region since our previous study highlighted particularities due to an unexplored founder effect. The region offers an ideal setting for rare pathogenic variants exploration. We worked with the whole genome sequencing (WGS) of 2390 individuals across the province. For 500 individuals, we reconstructed the genealogies over 15 generations using the BALSAC population database. Common ancestral origin, and differential variant frequencies enabled the subdivision of the province into clusters. Clusters were formed based on a UMAP performed on the 6 first principal components of the genotypes of 25 629 individuals, including our 2390 WGS.It allowed us to detect rare ClinVar established pathogenic variants with a carrier rate higher than in the general population in the WGS of individuals in the clusters, the highest being associated with 3MG-CoA hydratase deficiency, carried by 1/15 individuals in the Beauce cluster. Using the same approach on the whole cohort didn't allow detection of this variant. Identical-by-descent segments (IBD) will then be used to determine if the frequency increase is a result of the population structure or is linked to the disease. Once the methodology is validated for known pathogenic variants, we will work backwards to identify novel variants. Starting with the shared IBD, we will pinpoint founder variants that are more frequent in patients from certain clusters, and target them for further studies. This research contributes to scientific knowledge, and holds promise for improved diagnostics by justifying the pertinence of offering carrier tests based on regional ancestry. Our findings emphasize the importance of considering population structure in genetic research, urging a paradigm shift towards investigating geographically and historically distinct clusters with unique genetic profiles.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3013W: Expanding the utility of ancient DNA using large-scale genome sequencing

Authors:

C. Brand, J. A. Capra; Univ. of California San Francisco, San Francisco, CA

Abstract:

Ancient DNA offers a direct window into the genomes of past human populations. The ability to observe whether alleles are present in ancient individuals provides an extraordinary perspective on many fundamental questions in human genetics, including understanding the evolutionary history of variants that underlie human disease and other phenotypes. However, ancient genomes are often degraded, yielding low-coverage of most nucleotides. Further, many publicly available genotypes for ancient humans are limited to ~1.23 million specific loci. Thus, variants of interest often fall outside these specific positions, limiting the ability of ancient DNA to shed light on many loci.Here, we address this challenge by leveraging the fact that many modern variants occur in strong linkage disequilibrium (LD) with ancient genotyped variants (AGVs) to generate a resource enabling rapid identification of proxy variants for other researchers. We considered all AGVs from the Allen Ancient DNA Resource. We identified variants in LD with AGVs using LD data for rare to common variants in diverse whole genome sequencing from TOPMed as quantified in the TopLD database. We identified 1,199,828 total AGVs in TopLD and between 46,229,640 and 82,805,980 LD variant pairs at $R^2 \geq 0.2$ in four different ancestral groups. We stratified these LD pairs by the minor allele frequency per variant, as increasingly rare variants are less likely to occur in LD with an AGV. Among common variants, at least 75% were linked to an AGV at $R^2 \geq 0.2$ in all four populations. Further, at LD thresholds more supportive of variant co-occurrence on the same haplotype ($R^2 \geq 0.9$), at least 43.9% of non-African and 22% of African common variants in TopLD were linked to AGVs. Between 10 and 25% of low-frequency variants were in LD with AGVs at $R^2 \geq 0.9$. We also identified 3,066,837 rare and 3,070,582 ultra-rare variants linked to AGVs at $R^2 \geq 0.9$ across populations. These findings reveal that a considerable fraction of low frequency and common variants and millions of rarer variants can be studied in ancient samples using proxy variants. For example, a high-frequency nonsense polymorphism in *SLC22A10* is not an AGV itself but occurs in strong LD ($R^2 \geq 0.97$) with two AGVs. Analysis of the allele frequencies of those AGVs suggests that the nonsense variant was present in humans by at least 30,000 years. We integrate these results in a database that researchers can easily query to identify relevant proxy variants if their variant of interest is not directly genotyped in ancient humans.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3014W: Exploring Genetic History and the Spread of Lactase Persistence in Tunisian Populations

Authors:

A. Koedel[1], A. Lisi[1], A. Ranciaro[1], C. Battaggia[2,3], S. di Lernia[3], N. Boukhchim[4,5], H. Bel Haj Brahim[6], P. Anagnostou[2,3], G. Destro-Bisol[2,3], M. Campbell[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]La Sapienza Univ. di Roma, Rome, Italy, [3]Inst. Italiano di Antropologia, Rome, Italy, [4]Faculté des Lettres et Sci. Humaines, Université de Kairouan, Kairouan, Tunisia, [5]Laboratoire LMAIM, LR99ES01, Université de Tunis, Tunis, Tunisia, [6]Artisanat du Sahara, Douz, Tunisia

Abstract:

Placed at the center of North Africa, Tunisia has been a stepping stone of migration between sub-Saharan Africa, Asia, and Europe for millennia. Among the most significant migratory events in recent Tunisian history was the geographic expansion of Arabic-speakers from the Middle East into the region in the mid-seventh and the 11th centuries. While these episodes of migration have been supported by linguistic and archaeological evidence, only a small number of studies have examined the genetic consequences of these processes in modern-day populations. To explore the impact of these past demographic events, we examined whole genome sequence and genomic genotyping variation in 104 Arabic- and Berber- speakers originating from seven populations sampled across Tunisia that have traditionally practiced different modes of subsistence (i.e., agriculture and pastoralism). We also integrated this genomic information with publicly available genomic data from >1,582 individuals from 65 global populations, including 1,041 individuals from the Middle East, Europe, and sub-Saharan Africa. Based on our analyses, we uncovered higher levels of Middle Eastern and sub-Saharan African ancestry in Arabic-speaking nomadic pastoralists than in the Berber-speaking populations. Furthermore, we found much lower levels of European ancestry in Arabic-speaking nomadic herders relative to the Berber-speakers. Notably, the geographic expansion of Arabic-speakers from the Middle East also led to the spread of pastoralism and lactase persistence (LP) into Tunisia. To date, several functional single nucleotide polymorphisms in a regulatory region in intron 13 of the MCM6 gene are known to be associated with LP in with a tradition of dairying. In our analysis of intron 13 sequence variation, we found the common derived G-13915 allele—correlated with LP in the Middle East—at its highest frequency in Arabic-speaking nomadic pastoralists from southern Tunisia whereas this allele was completely absent in other Arabic-speakers from the northern region of the country. Furthermore, we observed the presence of the derived East African G-14009 allele at low frequency in the Douz Arabic-speakers. In addition, we explored patterns of adaptive evolution using haplotype-based methods and detected unusually long haplotypes around the G-13915 allele consistent with a model of recent selection primarily in the Arabic-speaking nomadic herders from southern Tunisia. Overall, this study provides new insights into the roles of demographic history and natural selection in shaping patterns of variation in diverse understudied populations.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3015W: Genetic-similarity-based approaches for matching to populations in reference datasets

Authors:

L. Anderson-Trocmé[1], J. Guez[2], K. Karczewski[3], J. Novembre[1]; [1]Univ. of Chicago, Chicago, IL, [2]MGH, Broad Inst. of MIT and Harvard, HMS, Cambridge, MA, [3]Massachusetts Gen. Hosp., Medford, MA

Abstract:

Human genetic variation is continuous and yet categorical labels are commonly used in population genetics analyses to describe groups of individuals and to facilitate matching to reference samples. Two particular applications that currently adopt discrete groupings are calculating allele frequencies in a reference sample, e.g. for genetic diagnosis, or linkage disequilibrium (LD) matrices, e.g. for fine-mapping. Here, we propose a method to move beyond discrete genetic ancestry groupings towards a continuous model based on genetic similarity to improve accuracy in both these applications.

Our method computes local allele frequencies and LD patterns using a distance-weighted method based on data from samples in a reference panel. Using cross-validation, we compared our method to more traditional categorical approaches by evaluating the error and bias from estimated allele frequencies to the known frequencies of groups of individuals. We show an improvement in predicting allele frequencies and LD patterns in held-out sets of individuals, and we show how the accuracy of our method outperforms discrete methods at a range of genetic divergences.

We test this method with examples from the largest public aggregated database of human exomes and genomes (gnomAD) which is widely used in genetic diagnosis applications. Our method provides a critical step towards moving beyond the use of inefficient discrete labels in population genetics and improves performance for downstream applications, including fine-mapping and rare disease investigations.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3016W: Geography and study design in the biobank era: a population genetics-informed approach to understanding rare, deleterious variation in human genetic samples

Authors:

M. Steiner[1], D. Rice[1], A. Biddanda[2], M. Ianni-Ravn[1], C. Porras[3], J. Novembre[1]; [1]Univ. of Chicago, Chicago, IL, [2]Johns Hopkins Univ., Baltimore, MD, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

The discovery of rare, large effect alleles and their phenotypic associations in human populations is critical to our understanding of biological mechanisms and an increasingly important component of the drug discovery pipeline. The unprecedented growth in sample size of human genetic studies will undoubtedly improve our understanding of such variants. However, the question of how not only the size but also the geographic diversity of samples and associated variation in genetic ancestry space will affect genetic discovery is largely unexplored. This question is particularly interesting in the context of rare deleterious alleles, which are known to be clustered in geographic space near their area of origin due to underlying evolutionary processes. With this motivation in mind, we present a novel theoretical framework to study the impact of spatial sampling design on the discovery of rare variants. In particular, we develop a population genetic model for the distribution of carriers of deleterious alleles in a structured population - accounting for dispersal, drift, selection, mutation, and uneven spatial sampling simultaneously - and derive key properties of the site frequency spectrum as well as downstream quantities. Analysis of our model suggests that geographically “broad” samples will include more singletons and segregating sites (the “discovery effect”), but each variant will be detected at lower frequency (the “dilution effect”). We validate these results using multiple approaches to population genetic simulation. Moreover, we integrate our results with phenotypic models to assess potential effects on association test power as samples become more broad from increased breadth and inclusivity of human genetic sequencing efforts, including both single-variant and burden test paradigms. We have additionally conducted empirical analyses in the UK Biobank dataset, evaluating the impact of sample breadth - in both geographic space and continuous ancestry space - on observed genetic variation. While “optimal” sampling strategies will ultimately be both context- and question-dependent, we anticipate that our quantitative analyses will help to inform a broader discussion regarding best practices for study design in the biobank era of human genetics.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3017W: Global diversity, recurrent evolution, and recent selection on amylase structural haplotypes in humans

Authors:

R. Lou[1], D. Bolognini[2], A. Halgren[1], A. Raveane[2], J. Rocha[1], A. Guarracino[3], N. Soranzo[2], J. Chin[4], E. Garrison[3], P. Sudmant[1]; [1]Univ. of California, Berkeley, Berkeley, CA, [2]Human Technopole, Milan, Italy, [3]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [4]Fndn. for Biological Data Sci., Belmont, CA

Abstract:

The adoption of agriculture, first documented ~12,000 years ago in the Fertile Crescent, triggered a rapid shift toward starch-rich diets in human populations. Amylase genes facilitate starch digestion and increased salivary amylase copy number has been observed in some modern human populations with high starch intake, though evidence of recent selection is lacking. Here, using 52 long-read diploid assemblies and short read data from ~5,600 contemporary and ancient humans, we resolve the diversity, evolutionary history, and selective impact of structural variation at the amylase locus. We find that amylase genes have higher copy numbers in populations with agricultural subsistence compared to fishing, hunting, and pastoral groups. We identify 28 distinct amylase structural architectures and demonstrate that identical structures have arisen independently multiple times throughout recent human history. AMY1 and AMY2A genes each exhibit multiple independent duplications/deletions with mutation rates >10,000-fold the SNP mutation rate, whereas AMY2B gene duplications share a single origin. Using a pangenome graph-based approach to infer structural haplotypes across thousands of humans, we identify extensively duplicated haplotypes present at higher frequencies in modern day populations with traditionally agricultural diets. Leveraging 533 ancient human genomes we find that duplication-containing haplotypes (with 5 or more AMY copies in contrast to the ancestral 3 copies) have increased in frequency more than seven-fold over the last 12,000 years providing evidence for recent selection in West Eurasians. Together, our study highlights the impact of the agricultural revolution on human genomes and the importance of long-read sequencing in identifying signatures of selection at structurally complex loci.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3018W: High Throughput phasing of HLA in a diverse ancestry set of individuals with high accuracy

Authors:

O. Cornejo[1], M-C. Yee[1], O. A. Garcîa[1], R. Barnard[2], B. Hoglund[2], C. Riegler[3]; [1]Univ. of California Santa Cruz, Santa Cruz, CA, [2]Twist BioSci.s, South San Francisco, CA, [3]Twist BioSci., South San Francisco, CA

Abstract:

The **immune system**'s Human Leukocyte Antigen I (HLA) locus has the highest **genetic diversity** in the human genome. Due to the great similarity between highly polymorphic alleles in the HLA region, its extreme variability and complex structure, it has been difficult to align and phase **haplotypes of HLA** to standard linear reference genomes using short read sequencing.We developed a high-throughput, one-step fragmentation and barcoding method to prepare DNA suitable for long-read **targeted sequencing** using samples of diverse ancestries from the Human PanGenome project and the International Histocompatibility Working Group.In collaboration with Twist Bioscience, we developed a probe set targeting the entire class I, II and III HLA regions, building upon existing probes developed by Twist for the class I HLA and part of the class II HLA regions.The post capture libraries were sequenced on both ONT and PacBio long read platforms.The data was mapped to the hg38, hg19, chm13 assemblies, and a graph pangenome to assess which approach maximizes alignment of rare variants, and enable haplotype phasing with increased accuracy. We show the high enrichment and increased accuracy in the haplotyping of the region using our capture/long read sequencing approach compared to other standard methods. We provide a comparison for the differential accuracy in sequencing with ONT and PacBio technologies combined with mapping strategies.The high throughput sample preparation and targeted sequencing make this an affordable and scalable method for analyzing population sized sample sets, while the graph Pan Genome reference will allow diverse alleles to be correctly mapped.We anticipate that our newly developed method will enable highly accurate HLA typing for under-represented ancestry groups and give insights into their evolution in response to local environmental pathogens.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3019W: Human Germline Mutational Signatures in 525K UK Biobank and Genomics England 100K Whole Genomes

Authors:

R. Sharma[1], M. Karpinski[1], S. Petrovski[1], K. Carss[1], D. Glodzik[2], D. Vitsios[1], X. Zou[1]; [1]AstraZeneca, Cambridge, United Kingdom, [2]Harvard Med. Sch., Boston, MA

Abstract:

Mutational signature analysis has produced valuable insights into the DNA damage and repair mechanisms, particularly in cancer genomics. In contrast, research on understanding germline mutational processes has received less attention. This study explores human germline single and double nucleotide variations (SNVs and DNVs) to provide a comprehensive understanding of germline mutagenesis and its dependence on genomic features and location.
In this work, we have used the UK Biobank (UKB) whole genome dataset (~460K EUR, ~10K SAS, and ~9K AFR ancestry samples) and the Genomics England (GE) 100K genome dataset (~45K EUR). For de novo mutational signature discovery, the conventional non-negative matrix factorization (NMF) approach yields inconsistent solutions across random initialization. Here, we used a refined volume-regularized NMF to obtain unique and robust mutational signatures. We analysed SNV and DNV spatial mutational profiles (for each cohort/ancestry) created by binning mutations into 10, 20, 30, 50, and 100K base-pairs non-overlapping genomic windows.
We identified eight SNV germline mutational processes (three strand-dependent and five strand-independent) consistent across various genomic windows in both UKBB and GE EUR cohorts. Most signatures also align with previous TopMed dataset (mixed ancestries, 45K samples) findings, underlining their robustness. Notably, we identified a strand-dependent process unique to the UKB AFR and SAS dataset, and this same process was also found in the TopMed dataset, which contains AFR ancestry samples. We further examined the spatial variation of germline mutational signatures on different genomic features across the genome. Two strand dependent processes showed Pearson correlation of 0.50 and 0.75 with gene direction and 0.3 with replication direction. Specific strand-independent signatures correlate with replication timings (0.47), CpG islands distribution (-0.49), and methylation levels (0.61). Lastly, in a preliminary analysis, we identified three DNV germline mutational processes (all strand-independent). Future study will extend the analysis to indels to identify all possible germline mutational processes. Our work aims to provide a set of germline mutational signatures with known aetiology, elucidate the mechanisms of mutagenesis in human germ line, and identify the association between germline mutational signatures and disease-associated mutations.
Acknowledgments: This research was made possible through access to data and findings in the National Genomic Research Library via the Genomics England Research Environment.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3021W: Identifying and interpreting lineage-specific fixed differences between modern and archaic humans

Authors:

D. Tagore, R. Bierman, J. M. Akey; Lewis-Sigler Inst. for Integrative Genomics, Princeton Univ., Princeton, NJ

Abstract:

The explosion of whole-genome sequence data from contemporary individuals and high coverage ancient DNA has enabled high-resolution catalogs to be constructed of lineage-specific fixed differences between modern humans and archaic humans (Neanderthals and Denisovans). These catalogs contain the genetic substrates defining uniquely human phenotypes and the chronology of evolutionary changes that distinguish modern humans from archaic humans. Although the interpretation of lineage-specific fixed differences remains challenging, the development of powerful functional genomics resources enables novel approaches to prioritize, identify, and test hypotheses about putatively functional genetic changes between lineages. Here, we describe an updated and densely annotated catalog of fixed and nearly fixed differences between modern and archaic humans. We identify ~0.2 million sites which are different between the two lineages and annotate these variants to identify their role in regulating gene expression and possibly playing a role in the phenotypes that distinguish modern and archaic humans. To this end, we integrated myriad functional genomics data from ENCODE, GTEx, RoadMap Epigenomics, and the 4D Nucleome Program. As expected, the predicted number of functionally important lineage-specific fixed differences is larger for non-coding versus protein-coding regions of the genome. We aim to trace the temporal changes of these variants in the two lineages by estimating their age and to explore the role of natural selection in driving these variants to fixation or near fixation. We also leverage our high-resolution catalog of lineage-specific fixed differences to test hypotheses about the evolution of compensatory changes, estimate rates of back mutations, and interpret deserts of archaic ancestry in modern humans. In summary, we describe a carefully curated and powerful resource of lineage-specific fixed differences that provides novel insights into the genetic changes over the past ~0.5 million years of human evolutionary history.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3022W: Investigating ghost introgression in Denisovans via ancestral recombination graphs

Authors:

N. McAllister, S. Zöllner, X. Zhang; Univ. of Michigan, Ann Arbor, MI

Abstract:

Denisovans, close relatives of Neanderthals, were identified through the nuclear genome of a finger bone found in Denisova Cave in the Altai Mountains. Sequencing of their mitochondrial genome has revealed lineages deeply diverged from those of both modern humans and Neanderthals, hinting at possible introgression from an unknown hominin group, known as ghost introgression.The possibility of identifying ghost introgressed sequence can inform our knowledge of human evolution and further our understanding of the interactions occurring between hominin groups in prehistory. Previous works leveraged summaries of allele frequency data such as the site frequency spectrum (SFS) to determine the likelihood of ghost introgression. While informative, the SFS is limited in its ability to determine the timing of introgression events as it assumes independence between markers. A more comprehensive alternative approach in searching for evidence of ghost introgression is the ancestral recombination graph (ARG). ARGs are structures that contain the complete genealogical history of a set of sequences, including the relationships between alleles and the locations of recombination events. The distribution of coalescent times inferred by the ARG can be informative of the genetic ancestors of Denisovans, including potential introgression from an unsampled archaic population. For example, if ghost introgression into Denisovans occurred, on a reconstructed ARG using modern humans, Denisovans, and Neanderthals, we expect the presence of particularly deep lineages older than the Neanderthal-Denisovan common ancestor. In this work we examine the power and accuracy of common ARG methods, ARGweaver and Relate, to detect ghost introgression in Denisovans. We use msprime to simulate modern sequencing samples under two distinct demographic models, with and without ghost introgression into Denisovans followed by Denisovan introgression into modern humans. We compare ARG methods on their ability to detect ghost introgression in Denisovan introgressed sequence in simulated modern humans. We then apply these ARG inference tools on a diverse panel of modern humans with putative Denisovan sequence and archaic human genomic data to detect evidence of ghost introgression into Denisovans. Identifying previously unknown archaic hominins, and their influence on the modern gene pool, will inform future work in understanding the evolution of complex traits and further our knowledge of human evolution.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3023W: Investigation of homozygosity effects using variant annotation in a sample with individuals from founder populations

Authors:

M. Rakesh, M-h. Roy-gagnon, K. Burkett; Univ. of Ottawa, Ottawa, ON, Canada

Abstract:

Background:
Long contiguous homozygous segments, referred to as runs of homozygosity (ROHs), have been shown to be associated with many complex traits and enriched for deleterious mutations. Founder populations have higher levels of homozygosity, making them ideal to study ROHs. The Canadian Longitudinal Study on Aging (CLSA) is a Canada-wide cohort that includes participants from two founder populations: Quebec (20%) and Newfoundland (7%). In this study we investigated the relationship between variant deleteriousness and homozygosity.
Methods:
For 24,655 European-descent individuals in the CLSA, ROHs were estimated using plink for a low linkage disequilibrium set of genotyped autosomal variants. For each of 743,000 variants, we considered only individuals homozygous for the minor allele (MAH) and obtained the ratio of the number of MAH individuals in ROHs to those not in ROHs (ROH ratio). We also recorded the mean length of the MAH-containing ROHs (ROH length). Variants' predicted deleteriousness was obtained through multiple annotations: Ensembl's variant effect predictor (VEP impact and consequence), the SIFT score for variants in exomes, and genomic constraint scores of genes from the gnomAD database. We then tested for associations between deleteriousness and ROH ratio and length. We also investigated individuals with the highest and lowest ROH coverage (top and bottom 5 percentiles).
Results:
We observed significant associations (p-values<0.01) between higher ROH ratio/length and multiple measures of variant deleteriousness. Individuals with highest ROH coverage were mostly from the two founder populations (51%) and were MAH at a significantly higher proportion of predicted deleterious variants compared to those with lowest ROH coverage.
Conclusions:
We described the patterns of homozygosity in the CLSA cohort and showed that ROHs are enriched for predicted deleterious variants in this sample. These findings support the usefulness of founder populations for studying homozygosity effects.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3024W: Landscape of clonal hematopoiesis among East Asian individuals aged 80 years and older

Authors:

Y. Usui[1], H. Iijima[1], M. Endo[1], T. Aoi[1], Y. Iwasaki[1], H. Nakagawa[1], K. Matsuda[2], Y. Momozawa[1]; [1]RIKEN, Yokohama, Japan, [2]The Univ. of Tokyo, Tokyo, Japan

Abstract:

Somatic mosaicism, a genetic and environmental factor-affected consequence of human aging, involves the accumulation of somatic mutations in cells throughout the body. Clonal hematopoiesis of unknown significance (CHIP), involving somatic mutations in blood cells, is a unique somatic mosaicism-related phenomenon in which cells can freely mix and expand, unlike cells in solid tissues. Furthermore, focusing on older individuals, CHIP could represent a natural experiment on saturation mutagenesis in humans, where mutations are enriched with a competitive advantage affected by multiple factors over the long term. Such information-based evaluation might reveal certain aspects of human aging and the complex impact of genetic and environmental factors. Besides, region- and population-specific evaluation sometimes clarify a novel gene-environmental interaction (Usui et al., *NEJM*, 2023). Here, we assessed CHIP in all available individuals without hematological neoplasm aged 80 years and older among BioBank Japan, regardless of the clinical characteristics of the participants. Using target sequencing, we analyzed ~25,000 individuals for the coding region of 16 genes, including *ASXL1*, *DNMT3A*, and *TET2*. The average coverage was over 2600x in this analysis. Characteristics of somatic mutations differed gene-dependently. In this study population, CHIP existence was associated with respiratory disease mortality, in addition to hematological neoplasm mortality. Gaining CHIP-related insights among older individuals might provide a further understanding of somatic evolution and reveal the dynamics of age-related conditions, leading to improved health management. Based on an epidemiological perspective, CHIP prevalence was higher in males than in females. Concerning the environmental factors, smoking and drinking habits were associated with the CHIP existence. However, in the latter case, a combination of the *ALDH2* germline variant, which is East Asian-specific and is known for its role in ethanol metabolism, enhanced the association. We observed a significant gene-environment interaction related to CHIP existence, indicating that acetaldehyde might also be pivotal for CHIP and that the landscape of CHIP might display region- and population-based variations. From another perspective, *ALDH2* alone displayed a weaker association with CHIP existence as the two effects canceled each other out: reducing the risk of CHIP by avoiding drinking habits and increasing the risk of CHIP. A more detailed evaluation would clarify the pathogenesis of human aging and the complex effects of genetic and environmental factors.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3025W: Leveraging Identity-by-Descent based methods in biobanks to understand how haplotypic variation contributes to Sickle Cell Disease severity

Authors:

J. Baker[1,2], G. Evans[1], R. Bohlender[3], H-H. Chen[4], C. Huff[5], D. Samuels[6], J. Below[7]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Genetics Inst., Div. of Genetic Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]U.T. MD Anderson Cancer Ctr, Houston, TX, [4]Academia Sinica, Taipei, Taiwan, [5]U.T. MD Anderson Cancer Ctr, Pearland, TX, [6]Vanderbilt Univ. Sch. of Med., Nashville, TN, [7]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

Background: Sickle Cell Disease (SCD) is a recessive red blood cell disorders that changes the shape of the cells affecting their flexibility and blood flow. Symptoms of SCD can vary in severity and include anemia, chronic pain, and stroke. These complications from SCD affect lifespan with the estimated lifespan of affected individuals being 20 years shorter than the average. The prevalence of SCD is estimated as 1 in 500 African Americans in the United States. Genetic causes of SCD are well established, but the causes of phenotypic heterogeneity are not completely understood. Leveraging Identity-by-descent (IBD) based methods in biobanks can provide us a greater understanding on the effects of genetic variation on the clinical presentation of SCD. **Methods and Results:** Genotyped data was phased using SHAPEIT4 and then pairwise IBD segments ≥ 3cM were identified using iLASH for 13,452 individuals of African ancestry in BioVU, Vanderbilt University Medical Center's biobank. We used the PheWAS Package to classify individuals as cases, controls, and exclusions for 1,436 PheCodes requiring that individuals have a code on two or more unique dates. We performed IBD Mapping phenome-wide and observed an enrichment of 36 PheCodes with IBD signals overlapping the *HBB* gene. We performed local IBD clustering and identified 2,893 networks of individuals who shared pairwise IBD segments overlapping the HBB locus. 17 of these networks were enriched for Sickle Cell Anemia cases. Out of these 17 networks, 2 were also enriched for cases of common SCD comorbidities (chemotherapy, cardiomegaly, bacterial pneumonia, and respiratory failure) as well as being significantly enriched for anemia related PheCodes (282-283) when compared to the total population. To answer the question of whether the more severe sickle cell cases clustered into these two networks, we performed a χ^2 enrichment test comparing how frequently SCD cases within these networks had these enriched comorbidities

compared to those SCD cases outside of these two networks. We identified that SCD cases within these networks had higher rates of chemotherapy (p-values=3.21e-05, OR=2.81 [1.69-4.70]), bacterial pneumonia (p-values=1.38e-5, OR=2.88 [1.75-4.76]), and respiratory failure (p-value=2.47e-5, OR=2.72 [1.68-4.45]) indicating that these are more severe SCD cases. **Conclusions/future direction:** These findings support our hypothesis that there may be a variant on the haplotypes underlying these two networks that is having a cis effect in combination with a pathogenic SCD variant, causing a more severe presentation of SCD. We plan to analyze the sequencing data of these individuals to identity any such variant.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3026W: Phenome-wide association studies of Mucins in *All of Us* data reveal tandem repeat-associated phenotypic traits

Authors:

E. Nizamis[1], E. Kaufman[1], E. Plender[1], E. Eichler[2], P. Valdmanis[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA

Abstract:

Mucins are essential glycoproteins with critical roles in immune function and protection of the epithelial barrier. Over 20 mucins have been identified in humans and many contain coding variable number tandem repeats (VNTRs) that are highly divergent in length and internal sequence composition between individuals. *MUC5B* and *MUC5AC* are adjacent airway epithelium mucins on chromosome 11p15.5. We reasoned that leveraging powerful resources such as the NIH *All of Us* database could reveal hitherto unknown associations between Mucin SNPs and phenotypic traits. We analyzed ~245,000 genomes with ~165,250 matching electronic health records (EHR), correcting for age, sex, and the first three principal components, restricting analysis to individuals aged 20 years or older with two or more ICD9 or ICD10 codes for the same phenotype. The *MUC5B* promoter/enhancer SNP rs35705950 was significantly associated with alveolar and parietoalveolar pneumonopathy ($p = 6.89x10^{-44}$; odds ratio (OR) = 2.05), idiopathic fibrosing alveolitis (p = 2.14 x10-36, OR = 2.85), post-inflammatory pulmonary fibrosis ($p = 2.62x10^{-34}$; OR = 1.82), extrinsic allergic alveolitis ($p = 3.96x10^{-08}$, OR = 2.38), as well as previously unreported traits, such as bronchiectasis ($p = 9.77x10^{-7}$, OR = 1.32). SNPs tagging MUC5AC VNTR alleles were associated with degeneration of the macula and the posterior pole of the retina ($p = 1.76 x10^{-4}$ - $9.49x10^{-4}$, OR=0.91), though these findings failed to reach significance after Bonferroni correction for 1,816 phecodes. Further analysis of *MUC5AC* and *MUC5B* revealed several non-coding VNTRs including a tetranucleotide

enhancer repeat with a contraction event that has reached high frequency specifically in samples of East Asian descent with nominal association with asthma and other traits. We extended *All of Us* analysis to other mucins and found association with MUC19 and immune disorders in conjunction with a VNTR expansion occurring exclusively on Admixed American alleles. Collectively, these findings underscore the power of diverse biobank data to reveal novel phenotypes associated with tandem repeats.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3027W: Polygenic adaptation leads to a higher reproductive fitness of native Tibetans at high altitude

Authors:

Y. He, B. Su; Kunming Inst. of Zoology, Chinese Academy of Sci., Kunming, China

Abstract:

The adaptation of Tibetans to high-altitude environments has been studied extensively. However, the direct assessment of evolutionary adaptation, i.e. the reproductive fitness of Tibetans and its genetic basis remain elusive. Here, we conduct systematic phenotyping and genome-wide association analysis of 2,252 mother-newborn pairs of indigenous Tibetans, covering 12 reproductive traits and 76 maternal physiological traits. Compared to the lowland immigrants living at high altitudes, indigenous Tibetans show better reproductive outcomes, reflected by their lower abortion rate, higher birth weight, and better fetal development. The results of genome-wide association analyses indicate a polygenic adaptation of reproduction in Tibetans, attributed to the genomic backgrounds of both the mothers and the newborns. Furthermore, the EPAS1-edited mice display higher reproductive fitness under chronic hypoxia, mirroring the situation in Tibetans. Collectively, these results shed new light on the phenotypic pattern and the genetic mechanism of human reproductive fitness under extreme environments.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3028W: Population health monitoring leveraging fine-scale population structure in the biobank at the Colorado Center for Personalized Medicine. ⭐

Authors:

M. Fisher[1], J. Shortt[2], M. Lin[3], N. Rafaels[4], S. Chavan[5], N. Pozdeyev[5], C. Caggiano[6], R. Shemirani[7], N. Zaitlen[8], V. Arboleda[8], E. Kenny[9], K. Crooks[10], C. Gignoux[3]; [1]Univ. of Colorado Anschutz, Aurora, CO, [2]Colorado Ctr. for Personalized Med., Aurora, CO, [3]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [4]Univ. of Colorado Sch. of Med., Aurora, CO, [5]Univ. of Colorado, Aurora, CO, [6]Icahn Sch. of Med. at Mt Sinai, New York, NY, [7]Icahn Sch. of Med. at Mount Sinai, Palo Alto, CA, [8]UCLA, Los Angeles, CA, [9]Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]Univ Colorado, Aurora, CO

Abstract:

Differences in prevalence across the phenome are driven by both genetic and environmental factors. Here, we follow our non-parametric, data-driven identity-by-descent (IBD)-based clustering approach to identify individuals with similar recent ancestry to model patterns of healthcare utilization across a large medical system. Previously we identified 148,142,648 IBD segments indicative of recent shared ancestry using identity by descent LocAlity Sensitive Hashing (iLASH) between pairs of unrelated individuals enrolled in the biobank at the Colorado Center for Personalized Medicine (CCPM) (73,346 total participants). We found 71 distinct genetic communities using cumulative IBD sharing with iterative Louvain clustering with 73,305 participants. This data-driven approach enables us to identify unique clusters that may have similar genetic ancestry profiles when compared to reference data. At 5% FDR, we identified 395 associations between 270 unique phecodes and 60 unique genetic communities among 1,563 phecodes derived from electronic health records (EHRs) controlling for age, sex and overall genomic PCs. Interestingly, we identified dozens of community:phecode associations where communities with similar genetic ancestry profiles to reference groups had opposite directions of effect for the same phecode. We highlight two communities of primarily European genetic similarity that have contrasting risk for skin cancer (OR [95% CI] = 0.73 [0.66-0.82] and 1.5 [1.36-1.70]). This heterogeneity is not driven by differential patterns in polygenic risk scores in the two communities ($p \sim 0.4508$), but we note that these measures do not capture all potential genetic risk factors. Unexpectedly, principal components analysis shows that the community with the lower skin cancer risk seems to cluster with northern European populations. These findings demonstrate the importance of incorporating non-genetic risk factors into models of complex disease risk in biobanks and offer new insights into healthcare usage and potential sources of disparity within participants in large-scale institutional biobanks.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3029W: Recapitulating the African Diaspora Genome through ONT Long Read Sequencing

Authors:

K. Barnes[1], M. Campbell[2], S. Chavan[2], M. Boorgula[1], C. Hinson[1], D. Ansong[3], M. Avila[4], L. Caraballo[5], C. Figueiredo[6], S. Gravel[7], N. Hansel[8], B. Henn[9], E. Herrera-Paz[10], K. Jordan[11], J. Knight-Madden[12], A. Mayorga[10], C. Olopade[13], H. Watson[14], M. Taub[15], E. Kenny[16], F. De La Vega[17], A. Ioannidis[17], C. Bustamante[18], N. Katsanis[18], R. Mathias[19]; [1]Univ. of Colorado Anschutz Med. Ctr., Aurora, CO, [2]Univ. of Colorado, Aurora, CO, [3]Kwame Nkrumah Univ. of Sci. & Technology, Kumasi, Ghana, [4]Univ. Natl. Autónoma de México, Mexico City, Mexico, [5]Inst. for Immunological Res., Univ. de Cartagena, Cartagena, Colombia, [6]Univ.e Federal da Bahia, Salvador, Brazil, [7]Mc Gill Univ, Montreal, QC, Canada, [8]Dept. of Med., Johns Hopkins Univ., Baltimore, MD, [9]UC Davis, Davis, CA, [10]Univ. Catolica de Honduras, San Pedro Sula, Honduras, [11]Georgia Inst. of Technology, Atlanta, GA, [12]Caribbean Inst. for Hlth.Res., The Univ. of the West Indies, Kingston, Jamaica, [13]Dept. of Med. and Ctr. for Global Hlth., Univ. of Chicago, Chicago, IL, [14]Med. Sci. Cave Hill Campus, The Univ. of the West Indies, Bridgetown, Barbados, [15]Johns Hopkins, Baltimore, MD, [16]Icahn Sch. of Med. at Mount Sinai, New York, NY, [17]Galatea Bio, Inc., Miami Lakes, FL, [18]Galatea Bio, Hialeah, FL, [19]NIAID, Bethesda, MD

Abstract:

There is continued bias in the human reference genome and inadequate representation of human genetic diversity. This shortcoming impacts both genotype-phenotype association research studies and variant filtering and prioritization for clinical applications, especially among diverse ancestry populations. Recent advancements in long-read sequencing (LRS) technology have enabled the development of more comprehensive genome references for individual human genomes, facilitating the discovery of novel structural variations (SVs), which are especially relevant for clinical applications, but are poorly captured using srWGS, for which the majority of reference databases are based. To recapitulate genetic diversity associated with the African Diaspora in the Western Hemisphere, we previously established the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) and performed short-read whole genome sequencing (srWGS) on 1,440 unrelated individuals from 33 different North and South American, Caribbean and continental African populations. Here, we report the sequencing of 500 genomes from a subset of the CAAPA cohort using Oxford Nanopore LRS. Our goal is to catalog long-read based SVs in individuals of African ancestry representing the African Diaspora in order to identify variation not previously detected or fully resolved using short-read

technology, with a specific focus on pharmacogenomic genes. Sequencing such underrepresented populations will significantly aid in variant interpretation, uncover new exons and undiscovered splice forms, and enhance imputation panels. This improvement will facilitate the inclusion of these populations in genotype-phenotype association studies and polygenic risk prediction, ultimately enabling precision medicine for all.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3030W: Recruiting and Engaging Native Hawaiian and Pacific Islander Elders: Insights from "Brain Train" and Community Partnerships

Authors:

J. Tavana[1], C. C. Allick[2], B. J. Vaitohi[1], R. J. Hansen[1], C. S. L. Pakalani[1], T. D. Packard[1], A. T. Perez[1], S. Polaapau[3], S. F. Funaki[3], F. Matthes[3], T. M. Alailima[1], J. K. Worthington[4], G. Shepherd[2], N. L. Ellison[2], J. E. Galvin[5], D. S. Buchwald[6], J. S. K. Kauwe[7], P. G. Ridge[1]; [1]Brigham Young Univ., Provo, UT, [2]Washington State Univ. IREACH, Seattle, WA, [3]Brigham Young Univ. Hawaii, Laie, HI, [4]Univ. of Hawaii, Honolulu, HI, [5]Univ. of Miami, Boca Raton, FL, [6]Univ. of Washington, Seattle, WA, [7]Brigham Young Univ, Provo, UT

Abstract:

Background: Native Hawaiian and Pacific Islanders (NHPIs) are among the least represented racial groups in Alzheimer's disease and related dementias (ADRD) research. A primary objective of our efforts is to apply innovative strategies for recruiting and retaining NHPI individuals in ADRD studies. Identifying effective methods to enhance community acceptance of scientific research is crucial for improving participation. Historically, NHPIs have been subject to unethical behavior and approaches by researchers, resulting in understandable caution regarding research participation. Our commitment is to ensure both innovative and traditional methods are applied to engage NHPI populations in a culturally appropriate manner. Activities are led by NHPI investigators and staff, who bring lived experiences and represent a strong academic-community partnership. **Methods:** We employ a multi-faceted approach to increase recruitment and retention from these communities. Our strategies include community and academic partnerships, social media outreach, and word-of-mouth referrals. A highly successful initiative is the "Brain Train", a novel, hands-on educational workshop that utilizes visual and interactive learning methods to cover ADRD knowledge, pathology, genetics, risk and resilience factors, and research participation. Participants are NHPI older adults ages 50+ who visit 10 "stations" mostly staffed by NHPI researchers. Elders

receive a “passport” that contains a record of their personal data to take home, while a copy is retained by the study team. Blood or saliva samples are also collected to determine APOE status and other biomarkers. Every participant signs a consent form and is informed about our governance Board and publications and presentations committee. This effort reassures them that their data is protected, they maintain complete ownership, and they have the option to withdraw from the study at any time. **Results:** Since 2022, 20 “Brain Trains” have been conducted in collaboration with NHPI communities, academic, and government partners from across the U.S., its territories, and one Pacific Island nation, engaging over 1,000 participants. An additional 6 sites are awaiting a Brain Train. We have also collected over 900 biological samples and have genotyped them for APOE status. **Conclusion:** Our team continues to foster relationships with community organizations and is expanding to new sites for ADRD community studies.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3031W: Refining constraint estimates of stop-loss variants

Authors:

E. Dsouza, A. Martin-Geary, S. Sanders, N. Whiffin; Univ. of Oxford, Oxford, United Kingdom

Abstract:

Background: Stop-loss variants are variants that disrupt the stop codon of a protein coding gene. These variants are often considered to result in loss-of-function of the gene, however, prior work suggests that they are on average about as deleterious as missense variants. While many variants may indeed be loss-of-function, a large fraction are in fact benign due to rescue through a nearby alternative stop codon. Here, we provide improved constraint estimates of stop loss variants in the presence of these rescue mechanisms with a view to inform clinical variant interpretation.
Methods: Using sequences as defined by the Matched Annotation from NCBI and EMBL-EBI (MANE) project, we identify the closest rescue stop site that would be used when the canonical site is lost. These genes were then also annotated with LOEUF scores to quantify their tolerance to loss-of-function. We used the single nucleotide variation within 730,947 individuals within gnomAD version 4 exomes and annotated the frequency of stop-loss variants in the general population. We quantify constraint against stop-loss variants using the Mutability Adjusted Proportion of Singletons (MAPS) metric, a relative measure of constraint using synonymous variation as a baseline, featuring improved mutation rates estimates provided by Roulette.
Results: We identified 49,907 stop-loss variants in gnomAD affecting MANE genes. When

restricting our analysis to the 40% most-intolerant genes, we identify that stop-loss variants that lead to a C-terminal extension of 10aa or more are predicted to be as deleterious as loss-of-function variants. A similar pattern was observed in genes that were predicted to undergo non-stop decay due to a lack of a subsequent rescue stop-codon downstream of the canonical stop-site before the polyadenylation site. This pattern was not observed in genes that were tolerant to loss-of-function (40% highest LOEUF scores), where stop-loss variants were on average about as deleterious as missense variants.
Conclusions: Our findings suggest that rescue mechanisms should be accounted for when clinically interpreting stop-loss variants and provide evidence-based thresholds that can be incorporated into clinical guidelines. We have created an open-source web tool, STOP g.APP, incorporating our annotations to support stop-loss variant interpretation.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3032W: Study of natural selection in post-LGM East Asian populations using ancient DNA

Authors:

C. Huang[1], X. Wang[2], C. Liu[3], C. Chen[1]; [1]Central South Univ., Changsha, China, [2]Kunming Inst. of Zoology, Chinese Academy of Sci., Kunming, China, [3]SUNY Upstate Med. Univ., Syracuse, NY

Abstract:

Adaptive evolution has played a significant role in shaping the phenotypic diversity of modern East Asian populations. Since the Last Glacial Maximum (LGM), changes in climate and life style have had profound impacts on the evolution of human. Previous studies inferred evolutionary pressures in East Asian populations based on genetic data from modern populations, which are susceptible to population demographic history. Ancient DNA (aDNA) data, however, can directly capture genetic information during human evolutionary processes, accurately reflecting the dynamic evolutionary history of populations. Given the genetic continuity of East Asian populations, aDNA data exhibits unique advantages for detecting natural selection in these populations.In this study, we integrated all the publicly available aDNA genotype data worldwide so far. By adapting population genetic methods and demographic history analysis approaches, we first selected the genetically homogenized East Asian populations. We then conducted a genome-wide natural selection analysis on East Asian populations based on the association between allele frequencies and sample age in aDNA data. In addition to the previously reported selection sweeps for alcohol drinking-related region on chromosome 4, immune-related HLA clusters on chromosome 6, and skin pigmentation-related gene

of *OCA2* on chromosome 15, we identified a novel selection sweep on chromosome 22 with very strong significance. Genes span on this region are including *CENPM*, *SEPT3*, *LINC00634* et al. By integrating information from GWAS and eQTL, we found that genes in this region are associated with phenotypes such as cognitive function, schizophrenia, neurological disorder-related drug metabolism, birth weight and so on. We further inferred the selection onset time and selection coefficients on this sweep by utilizing ancestral recombination graph. Combining with haplotype data, we confirmed that this region has undergone natural selection in post-LGM East Asian populations and is identified as a soft sweep.These results suggesting that, except for alcohol-drinking, immune system and skin color, neuropsychiatric traits, especially those related to brain development, have also played a significant role in the recent evolution of East Asian populations, possibly due to the increased complexity of society and increased social interaction between human beings. This study is crucial for understanding the relationship between adaptive evolution and the formation of complex traits in East Asian populations.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3033W: The allelic architecture of functionally deleterious variants in the Saudi Arabian population

Authors:

D. K. Malomane[1], S. Mangul[2], M. Abedalthagafi[3], C. W. K. Chiang[1]; [1]Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA, [3]Emory Univ., Sch of Med., Atlanta, GA

Abstract:

There are proportionally more deleterious variants in the genomes of non-Africans than Africans due to the relaxation of purifying selection during the Out-of-Africa (OOA) bottleneck. In addition to the OOA bottleneck, genetic variation in the Greater Middle East (GME) and Arabic countries have been greatly influenced by mating practices. Arab countries have a high rate of endogamous and consanguineous marriages, especially in Saudi Arabia. Consanguinity could increase the burden of deleterious alleles in a population as these deleterious recessive alleles are co-inherited in offsprings. While in the long run these deleterious recessive alleles are exposed to selection due to increased homozygosity, reproduction compensation in endogamous and/or consanguineous GME populations has been suggested to counteract the effectiveness of purifying selection. To investigate the allelic architecture of functionally deleterious alleles in a GME population, we analyzed 23,536,025 variants found in whole

genome sequencing (WGS) data from 302 individuals from Saudi Arabia. We annotated the variants using three different state-of-the-art annotation tools (AlphaMissense, VEP, and Genomic Pre-trained Network). Approximately 10% of the called variants were not previously identified in gnomAD, and were enriched among rarer alleles annotated to be likely pathogenic (e.g. 20% annotated as likely pathogenic by AlphaMissense, compared to 7.2% of alleles found in gnomAD). Compared to African or European individuals from gnomAD, the Saudi tend to show proportionally more deleterious alleles compared to those annotated to be benign or neutral, particularly among rarer and low-frequency variants (e.g. relative to gnomAD Africans, a 13% increase of deleterious alleles annotated by AlphaMissense between 0.5-5% frequency, compared to 7% decrease of the benign alleles; P = 0.001). Within subclusters of the Saudi population, the subcluster with greater inbreeding and lower effective population sizes showed greater enrichment of deleterious alleles as well. Taken together, our results suggest that Saudi's population history impacts its pattern of genetic variation with potential consequences to the population health. It further highlights the need to sequence diverse and unique populations and include them in large global genetic data platforms such as gnomAD, so that population-enriched deleterious alleles and their health consequences can be studied.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3034W: The *APOL1* p.N264K variant is co-inherited with the G2 kidney disease risk variant through a proximity recombination event

Authors:

C. Simeone[1], M. McNulty[2], Y. Gupta[3], G. Genovese[4], M. Sampson[2], S. Sanna-Cherchi[3], D. Friedman[1], M. Pollak[1]; [1]Beth Israel Deaconess Med. Ctr., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Columbia Univ., New York, NY, [4]Broad Inst., Cambridge, MA

Abstract:

Black Americans are three to four times more likely to develop nondiabetic kidney disease than other populations, and exonic risk variants in trypanolytic Apolipoprotein L1 (*APOL1*), termed G1 and G2, drive much of this increased susceptibility. G0 is the common reference haplotype not associated with disease and detected globally. Disease risk occurs in a recessive fashion, where individuals carrying two risk variants (G1/G1, G2/G2, or G1/G2) are at a 7 to 30-fold risk for hypertension-associated end-stage kidney disease (ESKD), focal segmental glomerulosclerosis (FSGS), and HIV-associated nephropathy. Recently, a low frequency, missense modifier in *APOL1*, rs73885316 (chr22:36265628 C > A, p.N264K, gnomAD AFR AF = 0.03), demonstrated substantial protection against kidney disease when co-inherited with the G2 risk

variant. As the distance between the p.N264K and G2 variants is only 367 basepairs, a recombination event would be unlikely; subsequently, it was postulated that two independent mutation events may have occurred to generate p.N264K-G0 and p.N264K-G2 haplotypes. To test these hypotheses, we examined 63 phased *APOL1* variants from high coverage whole genome sequencing data in 881 African individuals from the 1000 Genomes Project. We identified 337 unique haplotypes, comprising 273 G0 reference haplotypes, 33 G1 haplotypes, and 31 G2 haplotypes. The p.N264K variant appeared only on G0 and G2 haplotypes, and seven individuals in this population carried both the p.N264K and G2 variants on the same haplotype. Based on haplotype patterns, it is most likely that the p.N264K variant emerged first on the G0 haplotype, and a subsequent recombination event occurred with a G2 haplotype to create the p.N264K-G2 haplotype. This recombination event is supported by high recombination rates, 6.19 to 27.28 cM/Mb, at the 5' and 3' regions of *APOL1*. Whole genome sequencing data of an ancient African individual from the Allen Ancient Genome Diversity Project carried the p.N264K variant and displayed a similar haplotype to modern G0-p.N264K haplotypes, indicating this variant existed at least 8,000 years ago. Independent analysis of whole genome sequencing from the NEPTUNE cohort identified three G2-N264K haplotypes; these data were also supportive of a recombination event. Observing a recombination event within a small region between clinically relevant loci provides a unique opportunity to investigate the impact of modifying variants in disease susceptibility loci. This recombination event also underscores the importance of studying the entire spectrum of *APOL1* haplotypes and not the risk variants alone.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3037W: Tracing the evolution of E-Y14891, a major Ashkenazi Jewish Y-chromosomal lineage

Authors:

L. Cooper; Open Genomes Fndn., Inc., Tucson, AZ

Abstract:

As the longest haplotypic block of the human genome, the Y-chromosome is an ideal locus for tracing ancestry. It may be used to track the history and migration of families and populations by reckoning with patterns in the frequency, dated phylogenetic structure, and spread over space and time of Y-chromosome clades. Due to the Y-chromosome's lower Effective population size, these patterns are highly sensitive to random genetic drift and sampling bias. These factors allow for Y-chromosome data to be most informative in densely-sampled contexts, especially

those that involve a founder event or sex bias. The Y-chromosome is notoriously difficult to fully sequence, but recent improvements lay the theoretical groundwork for Y-chromosome phylogenies to approach generational-scale pedigrees. While this achievement is not yet realized at scale, current methods still allow for impressively dense reconstructions. Here, we apply a deep phylogenetic approach to Y-chromosome data of the well-studied Ashkenazi Jewish (AJ) population. We co-analyze recently-published ancient DNA from AJs and their relatives with previously-published WGS samples as well as a large collection of tester data from a direct-to-consumer genetic sequencing company. We establish a baseline understanding of AJ Y-chromosomal variation and focus on the major lineage E-Y14891, due to its dense reconstruction and lack of documented pedigrees extending beyond the 17th century. We infer two important layers in AJ population history: an initial founder event in the early medieval period (c. 7th-9th centuries CE) followed by a significant expansion in the late medieval period (c. 14th-17th centuries CE). We also infer subtle substructure in the modern AJ population related to the substructure found in the 14th century Erfurt community. Our results provide a proof-by-example for how deep phylogenetic analysis of Y-chromosome data can both reproduce and bring forth novel insights into a well-studied population history such as that of AJs.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3038W: Unraveling the genetic potentiators of modern human migration

Authors:

L. Casten, J. J. Michaelson; Univ. of Iowa, Iowa City, IA

Abstract:

The backbone of human progress is migration, allowing for ideas to cross physical and cultural barriers. Previous research has explored the profound impact of ancient migrations on human genetic diversity, but the influence of genetic variation on migratory behavior in contemporary humans is less clear. In this study, we characterized the effects of common genetic variants on migratory behavior by conducting a GWAS on the distance moved from birthplace of unrelated white British adults from the UK Biobank (N=282,178). Estimated heritability suggests that migration has a biological basis (SNP heritability=2.5%, SE=0.2%). We found 11 independently associated loci that map to genes linked to psychiatric risk and neurodevelopment, with the strongest signal coming from the transcription factor *MEF2C* (rs40504, *p*=8e-15). Analysis of our GWAS results showed numerous genetic correlations, including significant associations with:

educational attainment (r_g=0.77), brain volume (r_g=0.16), risk-taking (r_g=0.15), and bipolar disorder (r_g=0.14). Partitioned heritability analysis of our GWAS showed enrichment for variants associated with brain cis regulatory elements, excitatory neurons, and medium spiny neurons. Polygenic scores (PGS) derived from the migration GWAS were computed in an independent cohort of nearly 2,500 adults (SPARK). Analysis revealed migration PGS is associated with relevant behaviors, including: "I dislike staying in one place for very long" (beta=0.08, p=0.04), "My behavior is very changeable" (beta=0.09, p=0.04), and "My mood swings between elation and depression" (beta=0.11, p=0.01). Strikingly, genetic drivers of migration discovered in 21st century humans predicted ancient human migration distance in remains as old as 9,000 years (Allen Ancient DNA Resource, Pearson's r=0.1, p<0.001, N=1,339). Together, these results demonstrate migratory behavior is heritable, associated with adaptability, has substantial overlap with bipolar disorder, and the associated loci regulate gene expression in the brain. This wide ranging study of migration, a major driver of human innovation and diversity, suggests it has a strong genetic basis that has been conserved between ancient and modern humans.

Session Title: Evolutionary and Population Genetics Poster Wednesday Session

Board 3039W: Unusual female-bias in the *de novo* mutation rate of a lemuriform primate, the aye-aye (*Daubentonia madagascarensis*)

Authors:

J. Rogers[1], R. J. Wang[2], M. Raveendran[1], R. A. Harris[1], T. Nguyen[2], A. D. Yoder[3], M. Hahn[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Indiana Univ., Bloomington, IN, [3]Duke Univ., Durham, NC

Abstract:

Whole genome sequencing in families of humans and other mammalian species has revealed a consistent pattern regarding the parent-of-origin of *de novo* point mutations. In almost every mammalian taxon with reliable estimates, males transmit more *de novo* single-nucleotide mutations to offspring than do females, a number that increases with paternal age. We used deep (30x) Illumina sequencing in pedigrees from 7 mammalian species to quantify sex-specific *de novo* mutation rates. Following extensive data quality checks and filtering, the parent of origin for mutations was assigned when supported by either read-based or transmission-based estimates of haplotype phase. In all but one species, we recover the typical pattern of male mutation bias and an increasing number of mutations with paternal age. The aye-aye (*Daubentonia madagascarensis*), a long-lived nocturnal lemuriform primate endemic to Madagascar, was the exception. We find that young female aye-ayes transmit fewer point mutations to their offspring than young males, but that older females transmit more point

mutations than older males or young females (Total dataset: 92 maternally and 114 paternally transmitted across 6 breeding females, 4 breeding males, 12 offspring total). Surprisingly, we find that the maternal age effect in aye-ayes is much stronger than the paternal age effect (phased mutations per additional year of parental age: 0.75 maternal vs. 0.13 paternal). This unanticipated pattern holds across maternal ages at conception ranging from 6.75 to 26.4 years and paternal ages at conception 8.5 to >30 years. Females less than 15 years of age transmitted fewer than 10 phased *de novo* mutations while females older than 17 transmitted 17-25 *de novo* variants. No male transmitted more than 15 phased mutations. Analysis of X-chromosome ploidy confirms the sex of the DNA samples sequenced. This finding contradicts the broad pattern observed in humans and all other mammals reported to date. Comparisons of the coding sequences of 219 genes involved in DNA repair or other mutation-related processes, annotated using human gene models, found 234 possible novel stop codons and 3,127 possible missense differences between the genomic sequences of aye-ayes relative to available sequences for anthropoid primates. Aye-ayes are unusual primates in several ways, including skeletal, dental and cranial anatomy as well as feeding behavior. This observation of female-biased mutation rate suggests aye-ayes and potentially other strepsirrhine primates may exhibit atypical cellular/molecular processes within gametogenesis that generate higher mutation rates during oogenesis than spermatogenesis.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7052W: A Case Study: Modifying NIH NExTRAC Public Deliberation Prompts for Community Consultation to Advance ELSI Study of Sociogenomics Research ⭐

Authors:

S. Watson[1], K. Meagher[1], S. Callier[2], J. Cadigan[3], A. Prince[4]; [1]Mayo Clinic, Rochester, MN, [2]The George Washington Univ., Washington D.C., DC, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Univ. of Iowa, Iowa City, IA

Abstract:

Sociogenomics aims to measure the aggregate contribution of hundreds or thousands of genetic variants associated with a social outcome of interest, creating a polygenic score (PGS). Specifically, this rapidly advancing science utilizes genomics to predict socially mediated, non-medical outcomes, such as income and educational attainment, opening doors to a variety of

applications beyond medical care such as social and behavioral interventions. To date, PGS for dozens of social traits have been developed. However, such research raises questions about how findings are used and the implications for all groups who may be affected by the development and utilization of PGSs. While some see promise, others argue that it can entrench reductionist explanations and harken back to eugenics. Community engagement (CE) offers opportunities to explore and better understand views toward this emergent use of data. Our team selected a future oriented framework recently created by the Novel and Exceptional Technology and Research Advisory Committee (NExTRAC) who designed a set of prompts to aid in public deliberation and bridge scientific technology with bioethical principles and concerns about safety. We will describe our study's modification of the NExTRAC prompts to evaluate and prioritize polygenic scores for social traits in consultation with three biobank-affiliated Community Advisory Boards (CABs). Using a card-sort approach (an established method for exploring attitudes toward conceptual categories or groups and the social process of grouping and prioritizing items) with the modified NExTRAC prompts, the CABs helped our research team prioritize five social/behavioral traits from 30 existing scores. In addition to priority-setting, this engagement approach exhibits the strength of eliciting values of card sorters and our study team experience provides an example of a participatory approach to mixed methods research design. In addition to describing team experience, the presenter will provide attendees with an interactive guide to adapting these methods for community-engaged research, including for values-elicitation and priority-setting.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7053W: A Framework for Equity in Genomic Medicine and Research

Authors:

T. Patel; Genomics England, London, United Kingdom

Abstract:

This paper presents a framework for equity in genomic medicine and research. In England, those of non-European ancestry are currently underrepresented within biobanks in comparison to population statistics. There is a need to understand the barriers diverse publics face in engaging with and consenting to take part in genomic research. Thus, a deliberative workshop series was run with diverse publics across England focussing on four key areas: engagement practices, consent practices, research governance, and industry and international partnerships

in relation to low- and middle-income countries. Participants centred many of their expectations in relation to community benefit. In the UK the concept of public benefit has gained traction as a way to assess if healthcare innovations benefit members of the public fairly and research is not being conducted solely for profit. Extending this to diverse publics, community benefit understands what it is that members of diverse communities expect in order to increase consent to take part in research. The concept of community benefit was framed by participants in ways in which their cultural values play a large role in what would be acceptable research practices. Descriptions of expectations of engagement, consent, governance and international partnerships were consistent with cultural models for these groups. Thus, the framework has been developed to be culturally appropriate for diverse participants. A set of recommendations has been put forward for organisations and professionals working in this area in the UK.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7054W: Advancing Breast Cancer Care: The Role of Implementing a Genomic Counseling Model in Colombia

Authors:

C. Buitrago Jaime[1], M. Naranjo[2], H. Velasco[1], D. Cardona[1], J. P. Valencia-Arango[3], D. Silgado[4], A. Lopez[1]; [1]Personalized Med. Group, BioSci. Ctr. - Ayudas Diagnósticas Sura, Medellín, Colombia, [2]Ayudas Diagnósticas Sura, Medellín, Colombia, [3]Data Sci. Dept., BioSci. Ctr. - Ayudas Diagnósticas Sura, Medellín, Colombia, [4]EPS Sura, Medellín, Colombia

Abstract:

Genetic counseling is a valuable resource for assessing the risk of genetic disorders based on family history and interpreting specific genetic tests. However, with advancements in technology and knowledge, genetic counseling has evolved into genomic counseling, incorporating more complex genetic testing and a broader understanding of complex traits. This study presents the implementation of a genomic counseling model for breast cancer patients in Colombia, focusing on personalized primary prevention strategies secondary strategies and the use of population-validated Polygenic Risk Scores (PRS). The study involved the creation of a genomic counseling model through a multi-phase approach, including design, acceleration, and implementation phases, integrating lifestyle analysis, telehealth services, and clinical follow-up to enable early intervention.The research involved 1997 women, with lifestyle habits showing significant associations with breast cancer risk, such as an odds ratio of 1.14 (95% CI: 1.05-1.23) for an increase in BMI and an odds ratio of 1.12 (95% CI: 1.05-1.23) for a sedentary lifestyle. The

genomic counseling model led to reduced morbidity and mortality rates, with only one patient in the intermediate-risk group developing breast cancer, resulting in an improved diagnostic opportunity of 12 months. The study's risk stratification outcomes categorized the population into low, standard, intermediate, and high-risk groups based on PRS. The model received positive feedback from patients, emphasizing the importance of personalized genomic counseling in managing breast cancer risk.Furthermore, the study contributes to the field of personalized medicine by emphasizing the need for a tailored prevention program, including the validation of PRS for Colombians, to enhance personalized risk assessment in breast cancer prevention. The data presented in the study, including information on age, rural or urban residence, and socioeconomic status among participants, provides valuable insights into risk stratification outcomes and patient satisfaction with the model implementation. Overall, this research has the potential to impact public health outcomes in breast cancer management on a regional scale, reshaping traditional paradigms towards prioritizing primary prevention strategies.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7055W: Advancing genetic testing for neurological disorders in Tanzania: importance, challenges, and strategies for implementation

Authors:

M. Alimohamed; Muhimbili Univ. of Hlth.and Allied Sci., Dar es Salaam, Tanzania, United Republic of

Abstract:

This manuscript discusses the critical need for advancing genetic testing capabilities for neurological disorders (NDs) in Tanzania, emphasizing the importance, challenges, and strategies for implementation. Neurological disorders, often caused by a complex interplay of genetic and environmental factors, disproportionately affect low and middle-income countries like Tanzania. Patients with neurogenetic diagnoses in Tanzania face substantial obstacles, including lack of trust in medical professionals, stigmatization, and limited access to proper care. The manuscript underscores the infancy of genetic testing implementation for NDs in Tanzania, citing challenges such as high costs, limited availability, and the absence of routine testing in diagnostic procedures. The paper recommends establishing well-characterized cohorts, conducting a nationwide emergency census survey, and enhancing diagnostic services and

healthcare financing. The paper stresses on the importance of collaboration, research, public awareness, and advocacy to address the challenges posed by NDs in Tanzania. The manuscript proposes a comprehensive plan, including training and capacity building in molecular diagnosis and bioinformatics, to establish genetic testing capabilities and reduce the burden of NDs in the country. Overall, the manuscript calls for a collaborative effort involving government support, stakeholder engagement, and international partnerships to advance genetic testing and improve healthcare outcomes for individuals with neurological disorders in Tanzania.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7057W: Assessing the Role of Genetic Diversity and Environmental Knowledge on Healthcare Among Qinghai Minority Students

Authors:

L. Zhang; Qinghai Normal Univ., Qinghai, China

Abstract:

This study examines the ecological and environmental awareness among minority students in Qinghai province, China, focusing on the relationship between genetic diversity and healthcare outcomes. We surveyed 1,500 students from five colleges, with a 97.13% response rate. Only 9.21% of respondents demonstrated awareness of their cultural environmental practices, indicating a low level of environmental knowledge, possibly influenced by genetic factors tied to ethnic diversity. This diversity, although beneficial for local adaptation, may increase susceptibility to specific health risks, underscoring the need for genetically informed healthcare approaches.Further research into HPV infection among the Han and Tibetan ethnic groups in Qinghai revealed significant differences. We collected data from 1,500 cases, finding a 16.58% positivity rate for HPV among the Han (65 out of 392) and an 18.41% rate among the Tibetans (204 out of 1,108). The most common HPV types were HPV16 and HPV58, with statistically significant higher rates of HPV16 and HPV31 infections among Tibetans compared to the Han. These findings highlight the necessity of integrating genetic considerations into healthcare practices.The results advocate for combining environmental education with genetic and health literacy among minority students. Such integration would prepare them to tackle local ecological challenges and health risks effectively, promoting sustainable development and improved health outcomes. Looking ahead, the study anticipates a shift towards more personalized healthcare based on genetic profiles, improved accessibility to genetic testing, and

expanded environmental education for minority communities. This approach should respect diversity, privacy, and fairness, ensuring community involvement in shaping their health and environmental initiatives.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7059W: Comparing telehealth and in-person genetic counseling visit times across specialties

Authors:

C. Applegate[1], M. Hguyen[1], B. Murray[1], C. Tichnell[1], C. Gordon[1], L. Yanek[2], C. Taylor[1], C. James[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Univ, Baltimore, MD

Abstract:

With the increase in number of genetic counselors (GC) and expansion of services, quantifying genetic counseling time is critical to optimize service delivery and staffing. Current data published by the National Society of Genetic Counselors relies on self-reported estimates of time and does not fully account for possible differences across specialties or between service delivery models, e.g. in-person (IP) vs. telemedicine (TM). One significant barrier to studying GC time is that time documentation is typically embedded in the GC note requiring manual extraction, which is laborious and error prone. We extracted GC time using python regular expressions (Regex). This strategy showed high performance to detect and extract time from GC notes. From a dataset of over 72 million notes at Johns Hopkins Medicine (JHM), we identified 3,011 IP and 7,622 TM free-text, genetic counseling notes with time documented from 2020-2023. As a comparison group, we identified 6,926 GC notes with time from 2016-2019, all of which were IP. These data encompass 24 JHM GCs across 6 specialties (Aortopathy, Cardiology, Neurology, Ophthalmology, Pediatric, and Prenatal). We compared GC time across all visits, IP vs. TM, and by specialty. Tests of statistical significance were performed using Kruskal Wallis or Mann-Whitney U tests with a Bonferroni-correction. Median GC time across all visits occurring in 2020-2023 was 45 mins (IQR [30-60]) and were the same for TM and IP visits. We observed significant differences across specialties, ranging from a 35-min median visit (IQR [25-45]) in Cardiology to medians of 60-mins in Ophthalmology (IQR [60-60]) and Neurology (IQR [47-60]) ($p<0.0001$). These specialty-based differences were observed for both TM and IP visits. The specialty-specific differences in time were consistent across service delivery models. However, two departments differed significantly with respect to IP vs. TM time spent. Pediatric genetics had longer TM time (median 60 mins, IQR [40-60]) compared to IP visits (median 47 mins, IQR

[24-60]); p<0.0001. In contrast, cardiology spent less time on TM visits (median 30 mins, IQR [24-40]) versus IP visits (median 37 mins IQR [30-45]); p<0.0001. We also found an overall reduction in time per visit when comparing the 2016-2019 IP times (median 50 mins, IQR [40-60]) to 2020-2023 IP times regardless of specialty (median 45 mins, IQR [30-60]); p<0.001. Our approach allowed us to extract real-world GC visit time across multiple years, specialties, and service delivery models to establish norms vital for genomic workforce and clinical planning.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7060W: Decoding Ambiguity: Evaluating the Psychological Harms and Utility of Returning Cancer Variants of Uncertain Significance ⭐

Authors:

J. Sam[1], C. Mighton[2], E. Reble[3], S. Shickh[4], M. Clausen[5], R. Kodida[6], D. Hirjikaka[6], S. Grewal[6], S. Panchal[7], M. Aronson[7], K. Semotiuk[7], S. Armel[8], T. Graham[9], N. Forster[10], J-M. Capo-Chichi[10], E. Greenfeld[7], A. Noor[7], I. Cohn[11], C. Morel[12], C. Elser[7], A. Eisen[9], J. Carroll[13], E. Glogowksi[14], K. Schrader[15], J. Lerner-Ellis[16], R. Kim[10], K. Thorpe[4], Y. Bombard[13]; [1]St. Michael's Hosp., Toronto, ON, Canada, [2]St. Michaels Hosp. & Univ. of Toronto, Toronto, ON, Canada, [3]St. Michael s Hosp., Toronto, ON, Canada, [4]Univ. of Toronto, Toronto, ON, Canada, [5]Genetics Adviser, Toronto, ON, Canada, [6]St Michael's Hosp., Toronto, ON, Canada, [7]Mount Sinai Hosp., Toronto, ON, Canada, [8]Princess Margaret Hosp., Toronto, ON, Canada, [9]Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [10]Univ. Hlth.Network, Toronto, ON, Canada, [11]The Hosp. for Sick Children, Toronto, ON, Canada, [12]Univ Hlth.Network, Toronto, ON, Canada, [13]Univ of Toronto, Toronto, ON, Canada, [14]Sanofi Genzyme, New York, NY, [15]BC Cancer / Univ British Columbia, Vancouver, BC, Canada, [16]Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada

Abstract:

Background: Variants of uncertain significance (VUS) account for up to 40% of classified variants in cancer predisposition genes, posing diagnostic and decision-making challenges. These challenges are compounded when communicating the implications of VUSs in pre- and post-test counseling. Previous studies have highlighted the psychosocial harms and overestimation of clinical significance of returning VUSs. This study aims to evaluate whether receiving VUSs related to a patient's personal or family history of cancer (primary VUS) has greater psychosocial harms and utility compared to receiving a negative result or unrelated cancer-related VUSs (secondary VUS).
Method: Participants with a personal or family history of cancer were recruited from familial

cancer clinics in Toronto between Sept 2018 and Jan 2021. All participants underwent exome sequencing (ES) and received cancer-related results. The primary outcome was genetic test-related concerns, measured by the Multidimensional Impact of Cancer Risk Assessment (MICRA; range 0-125, higher scores indicate more negative perception) at 12 months post-results, analyzed by linear regression. Secondary outcomes were perceived utility and medical actions, measured by the Perceived Utility Scale and Behavioral Risk Factor Surveillance System, with changes in perceived utility assessed using analysis of covariance.Results: Seventy-two participants received negative or VUS results (negative results, n=16; primary VUS, n=22; secondary VUS, n = 34): mean age = 59 years, diverse (42% non-white/European ethnicity), and majority female (81.9%). At 12 months post return of results, mean MICRA scores for those with negative results was 20.3 (SD = 1.6), while those with primary VUSs and secondary VUSs were 17.2 (SD = 7.8) and 20.2 (SD = 7.7), respectively. MICRA scores of those who received negative results were similar to those with a primary VUS (coefficient = -2.63, 95% CI -7.12, 1.89) and secondary VUS (coefficient = 0.05, 95% CI -4.07, 4.18) (p=0.23). In terms of perceived utility, there was no strong evidence of differences in baseline adjusted scores among result types at 12 months (p = 0.62). Regarding medical actions, one participant with a negative result, two with a primary VUS, and two with a secondary VUS consulted a healthcare provider about their ES results.
Conclusions: We did not find evidence that receiving a primary or secondary VUS leads to greater psychological harms or perceived utility of ES compared to negative ES results. Medical actions were also consistent with clinical guidelines on communicating the implications and medical management of VUSs.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7061W: Defining social and behavioral traits in polygenic scores and understanding measurement limitations.

Authors:

A. Prince[1], C. Canter[2], A. Long[2], J. Cadigan[3]; [1]Univ. of Iowa, Iowa City, IA, [2]Univ. of North Carolina, Chapel Hill, NC, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Polygenic scores for a range of social and behavioral traits are rapidly being developed with the goal of utilizing genetics in social science research and understanding gene-environment interactions. In the burgeoning field of sociogenomics, PGS are being created for a wide variety

of social and behavioral traits, such as educational attainment, risk tolerance, and aggression. These scores are frequently developed using data from biobanks by identifying associations between biobank participants' genetic data and their responses to survey questions. This interplay between survey and genetic data allows scores to be developed for a wide variety of social traits. However, it also confines the research based on how traits in existing survey datasets are measured. This presentation reports on two sets of findings from our larger study examining the ethical, legal, and social implications of the development and use of polygenic scores for social and behavioral traits. First, we will report results from our research using horizon scanning methodology to examine current trends in sociogenomic PGS research and their applications. Horizon scanning entails systematically examining information sources, like research publications, conference proceedings, white papers, and news articles, to detect early developments or trends in technology. Using two to three traits as models, we will interrogate the survey questions utilized to develop the scores, highlighting the strengths and limitations that stem from the underlying measurement of the traits themselves. We will then draw on interviews with social scientists who study these traits but do not use polygenic scores in their research, to compare how the traits are measured with social scientists' knowledge and expertise on the traits. This pairing of the underlying trait development with social scientists' views will help to illuminate disconnects that may hinder the adoption and use of polygenic scores in social science research as well as potential ethical and social implications in the development of PGS.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7062W: Diversifying the genetic counselor workforce via community colleges: A look at higher education experience before graduate school, grit, and resilience

Authors:

K. Schneider[1], H. A. Zierhut[2], B. K. Freeman[1], E. Sandoval-Lucero[3]; [1]Univ. of Colorado, Aurora, CO, [2]Univ. of Minnesota, Minneapolis, MN, [3]Univ. of Colorado, Denver, CO

Abstract:

Health disparities in genetic counseling and genomic medicine are prevalent for individuals from minority populations. Currently, the genetic counseling workforce does not demographically match the population in race, ethnicity, disability status, gender, socioeconomic status, languages spoken, geography, or other attributes. Increasing diversity within the genetic

counseling profession is one way to work towards reducing disparities, and many groups have been working on recruitment in high schools and universities. Literature has given limited attention, however, to the recruitment and retention of genetic counselors (GCs) from community college (CC) systems. CC systems have more diverse student populations than many 4-year universities, including older average student ages, more male students, and higher rates of Native American, Hispanic, and Black students, which are all historically excluded populations in genetic counseling. In addition, it is known that burnout and compassion fatigue are common in the genetic counseling profession and that burnout leads to GCs leaving the profession. An aim of this study sought to understand if GCs who started with CC education had grit or resiliency differences from other GCs. A quantitative survey of GCs (N=846) evaluated how GCs started their higher educational journeys and compared resiliencies and demographics between respondents. Only 36% (301/838) of respondents ever took a class at a 2-year or CC and most took 3 or fewer (N=731). Cross tabulation analysis identified a cut-off of taking 4 or more CC classes as representing GCs who have had a meaningful CC experience (N=115). Mean scores utilizing the Grit-S and ARS-30 validated grit and academic resiliency scales, respectively, identified higher mean scores on these scales for respondents who have had a meaningful CC experience than for respondents who took 3 or fewer CC classes. In addition, the need to work while in school demonstrates the dedicated work ethic required to be successful. Off campus salaried employment during undergraduate education was more common for respondents with meaningful CC experience, and 28% (29/104) of GCs who took 4 or more CC classes were Pell grant recipients versus 18% (124/684) of those who took 3 or fewer classes. Overall, survey results indicated strengths with grit and resiliency, as well as more financial barriers to overcome, in GCs who attended CC. With an improved understanding of the current landscape of GCs and CC experiences and perspectives, the profession can better explore opportunities for improved recruitment and retention of GCs from CC systems which can ultimately improve the diversity of the profession.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7063W: Doing no Harm: Exploring the psychosocial impacts of receiving secondary findings from whole exome sequencing

Authors:

V. Rokoszak[1,2], D. Hirjikaka[2], M. Clausen[2], C. Mighton[3], S. Shickh[3], Y. Bombard[3]; [1]McGill Univ., Montreal, ON, Canada, [2]Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.System,

Toronto, ON, Canada, [3]Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.System & Univ. of Toronto, Toronto, ON, Canada

Abstract:

Objective: Harms such as distress and psychological burden are commonly cited by healthcare providers as a concern for the returns of secondary findings (SFs). However, there is limited evidence detailing the psychosocial impacts of receiving SFs, in particular with the return of non-medically actionable results, and after a significant amount of time. We explored the psychosocial impacts of receiving SFs across a broad range of result types from whole exome sequencing. **Methods:** This was a qualitative study within an RCT where participants selected SFs from whole exome sequencing. Participants were adult cancer patients who had previously received an uninformative result from genetic testing. Participants used a decision aid and spoke with a genetic counsellor to choose between six categories of SFs 1) medically actionable, 2) common disease SNPs, 3) rare Mendelian diseases, 4) early-onset neurological diseases, 5) carrier status, and 6) pharmacogenetic variants. Results were returned by a genetic counsellor. Semi-structured interviews were conducted with participants approximately 1 year after receiving SFs and transcribed verbatim. Transcripts were analyzed using thematic analysis. **Results:** Twenty-five participants were interviewed; all chose at least 4 categories of SFs. The majority of participants were female (22 female, 3 male) and ranged in age from 32 to 81. All had a personal and family history of cancer or polyposis. Participants reported low levels of distress, for which we identified 3 reasons: 1) the results received were perceived as less severe than anticipated, 2) adequate expectation setting (due to the decision aid, thoroughness of genetic counselling and the contextualization of results within personal/family histories), and 3) internal sources of resilience (including overcoming cancer). Participants expressed benefits from receiving SFs: 1) relief and validation after receiving a genetic diagnosis and/or result explaining their personal/family history, 2) satisfaction and appreciation to have health information, 3) pride and accomplishment, often related to overcoming their anxieties and/or to having genetic information to share with family members, 4) gaining a sense of control and empowerment from obtaining this health information and improved ability to self-advocate using their results. **Conclusions:** Our findings suggest that participants did not experience any major distress or anxiety when receiving secondary findings. Furthermore, many participants experienced positive psychological outcomes. In addition, these findings provide insight into protective factors against psychological harm.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7064W: Educating primary care providers about newborn screening and genetic conditions

Authors:

M. Raspa[1], M. Lynch[1], B. Wright[1], A. Corbo[1], **C. Horvitz**[1], A. Harris[2]; [1]RTI Intl., Research Triangle Park, NC, [2]Hlth.Resources and Services Admin., Rockville, MD

Abstract:

Background: As part of the Newborn Screening Saves Lives Act, the Health Resources & Services Administration created the Newborn Screening Information Center (NBSIC), a website that seeks to increase awareness, knowledge, and understanding of newborn screening and genetic conditions among parents and pediatric primary care providers. The website has a specific page dedicated to providers, as well as information about conditions identified through newborn screening, and state screening panels. **Methods:** To assess the extent to which the NBSIC website meets the needs of providers, we conducted 9 usability interviews with pediatric providers, including 4 primary care doctors, 2 nurses, 2 midwives and 1 nurse practitioner. We developed resources and new content for providers and tested draft content with an additional 9 pediatric providers. **Results:** Providers indicated that the site would be most useful for a lay audience, but there are also useful resources for providers. Providers reported the most interest in the conditions pages, which are regularly updated by genetic counselors, given their lack of familiarity with newborn screening conditions and the pages for each state which list which conditions are included on their panel. Providers mentioned that the most useful information to add to the site would be a directory for referrals and resources when there is an out-of-range result. We found that providers would be most likely to find the site using a web search or resources from their medical associations. **Implications for Practice:** Based on insights and reactions from pediatric primary care providers, we updated and enhanced information on the site. We developed a conversation guide to help providers navigate conversations about an out-of-range result and provide parents with the right amount of information at the right time. We also developed a targeted dissemination plan to reach pediatric primary care providers (e.g., social media). Through a user-centered approach, we improved the utility of the site for providers.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7065W: Efficient Inquiry Management in Expanding Biobanks: Balancing Growth and Support

Authors:

E. Shalowitz[1], M. McIntyre[2], S. Jones[1], D. Koffenberger[1], E. Kudron[3,1], D. P. Kao[1,3], The Colorado Center for Personalized Medicine; [1]Colorado Ctr. for Personalized Med. Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, [2]Clinical Res. Support Team, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Univ. of Colorado Sch. of Med., Aurora, CO

Abstract:

Introduction: Population-scale biobanks are vital for advancing genomic medicine, with their value increasing as they scale. Efficient communication among biobanks and those who inquire is essential for optimizing resources, but the nature of inquiries to growing biobanks is unknown. Our objective was to describe the frequency and types of inquiries received by a biobank's central email inbox and phone hotline.
Methods: Our biobank provides the study email and hotline phone number on all outward materials, and the website features contact, feedback, and withdrawal forms, with submissions automatically sent to a central inbox. Study staff aim to respond to each inquiry in kind (e.g., email response to email inquiry, phone call to phone inquiry) with standardized language or triage inquiries to relevant stakeholders (e.g., genetic counselors, pharmacists). We evaluated inquiries received by the biobank from 9/1/22 to 4/15/24.
Results: During the evaluation period, 33,878 new participants consented, totaling biobank enrollment at 243,561. The biobank sent 16 emails, 2 mailers, and 3 patient portal messages, with send cohorts ranging from 247 participants to all participants (median = 3,111, interquartile range = 1,607 - 6,578). Additionally, 48,574 participants received pharmacogenetic results and 463 received secondary findings. We received 3,515 communications, of which 2,632 (75%) could be traced to a specific study outreach. Total communications included 2,107 (60%) emails, 699 (20%) withdrawal web form submissions, 586 (17%) phone calls, 85 (2%) contact web form submissions, 32 (1%) inquiries from other sources, and 6 (<1%) feedback web form submissions. Not all inquiries warranted a response; there were 1,181 (34%) automatic reply or bounced emails, 120 (3%) spam, 78 (2%) email replies 'thank you', and 17 (<1%) were blank. Where a response was warranted (n = 2,120, 60%), 40% were withdrawal requests or notices of deceased relatives (n = 858), and were about return of results (n= 493, 23%), general study details (n = 258, 12%), sample collection (n = 183, 9%), topics unrelated to the biobank (n = 139, 7%), and data privacy, research, industry partnerships, questions from clinicians, or were wrong number phone calls (each <5%).
Conclusion: The number of inquiries that warranted response was low in relation to the number of participants enrolled during the study period (6%) and total biobank participants (0.9%).

Maintaining prompt communication with all inquirers is crucial for the success of a biobank's operations and these data suggest that the volume of inquiries for a growing biobank is modest and may be manageable without marked increases in staffing.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7066W: Emerging ethical considerations of neurogenetics data for Indigenous Peoples

Authors:

J. Kim[1], K. Tsosie[2]; [1]Arizona State Univ., Tempe, AZ, [2]Arizona State Univ., Phoenix, AZ

Abstract:

Neurogenetics is an emerging field that leverages digital sequence information derived from human genomes to improve our understanding of neurological and neurodevelopmental conditions. However, conversations and policies related to ethical use of neurogenetic data, especially as it relates to Indigenous peoples and communities, are incredibly nascent. Given the severe cultural and ethical concerns that are specific to neurological data, more conscious attention and emphasis on research praxes and protocols are needed, especially considering the rapidity and scale of information being collected from Indigenous peoples outside of Tribal-level governance and jurisdictions, to include urban-based specialty health centers and academic research institutions off Tribal lands.
If Indigenous peoples are to be a target for inclusion in neurogenetics basic and clinical research into ever-growing datasets, then the fields must recognize and operationalize Indigenous data sovereignty principles at multiple intervention points (e.g. data collection, data analyses, data storage, and brain- or biobanking) to prevent recapitulating harms in this new Big Data field. Here, we focus on the unique risks and concerns of neurogenetic data as it relates to: 1) reference genomes and biological genetic race reification, 2) benefit sharing, 3) Indigenous data sovereignty and data equity, 4) suggestions for standard operating procedures as implemented by global Indigenous groups, and 5) disability and access-to-care considerations for Indigenous individuals and community members. We consider this a sentinel and signal for researchers, clinicians, and institutions to bolster their attention to these emerging issues.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7067W: Empowerment of Diversity in Genomic Careers: Promotion of Workforce, Employment and Readiness in Genomics (PoWER-G) and the Path to Success

Authors:

M. Tiirikainen[1], K. Peplowska[1], Y. Deng[2], L. Wu[1], J-P. Bingham[3]; [1]Univ. of Hawaii Cancer Ctr., Honolulu, HI, [2]John A. Burns Sch. of Med., Univ. of Hawaii at Manoa, Honolulu, HI, [3]Coll. of Tropical Agriculture and Human Resources, Univ. of Hawaii at Manoa, Honolulu, HI

Abstract:

The demand for highly trained and diverse Genomics and Bioinformatics scientists, as well as scientific core facility staff and managers, is pressing, and requires immediate attention. Funded programs like the University of Hawaii's Pacific Center for Genome Research (PCGR), as part of the NHGRI Diversity Centers for Genome Research Consortium (DCGR) play a crucial role in training the next generation of these professionals due to the expected growth and heightened need for personalized medicine.

Specifically, at PCGR, our vision is embodied in the "Promotion of Workforce, Employment, and Readiness in Genomics (PoWER-G)". This new program is committed to addressing these demands directly and encompasses three primary goals of building a diverse genomics workforce in our first funding cycle: **Aim 1:** Establish a Genomic education feeder pipeline to identify and support undergraduate (UG) and graduate (GR) educational networks and develop a pathway for advancing genomic skills. **Aim 2:** Strengthening post-graduate pathways into Genomic research. **Aim 3:** Enhancing the skills of an advanced genomic workforce by developing and promoting an advanced workforce training platform to meet current and future research and industrial needs in Genomics.

These aims align with the NHGRI Action Agenda for a Diverse Genomics Workforce[a] by: (i) Developing and supporting initiatives that provide early exposure and access to careers in genomics. (ii) Developing and supporting training programs and networks that connect undergraduate and graduate education to careers in genomics. (iii) Developing and supporting training, career development, and research transition programs that lead to independent research and clinical careers in genomics. In this context, we will describe and detail our undergraduate (UG) and graduate (GR) pipelines (MS and PhD level), which are parallel-to-degree educational exposure to bioinformatics, leveraging existing resources, and the development of workflow programs that interface with daily genomic core facility research and service activities. Additionally, supplemental professional development activities will provide a robust foundation for competitive entry into the workforce, with focused efforts on increasing

the representation of under-represented minorities.
[a]https://www.genome.gov/about-nhgri/leadership-initiatives/diversity-in-genomics-workforce

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7068W: Engaging high school students in exploring big data via the *All of Us* Data Browser

Authors:

L. Stark[1], K. E. Fenker[1], H. Krishnan[1], M. Molly[1], R. J. Peterson[1], Y. Crosby[2], R. S. Baskir[2]; [1]Univ of Utah, Salt Lake City, UT, [2]NIH All of Us Res. Program, Bethesda, MD

Abstract:

We collaborated with high school biology teachers from across the U.S. to co-design curriculum materials that introduce students to "big data" and engage them in exploring the *All of Us* Research Program's public Data Browser. The "What is Big Data" video (2:28 min) provides an introduction to the topic using examples that students are likely to be familiar with as well as from biomedical research. The "Data Symphony" interactive piece combines science, music, and visual illustration. Users move sliders to explore ways in which adjusting the amount of data and data diversity for three data types (genetic ancestry, gender, and age) impact the richness of the music and its visual representation. In the "Beyond the Thermometer" interactive, students are introduced to the types of data that *All of Us* collects and to the concept that multiple types of data can be useful for understanding human health; a downloadable, paper-based version is also available. Three print-based activities guide students in exploring and using data in the Data Browser. The "Data Browser Webquest" provides an introduction to the Data Browser and the types of data that are available. In "How Big is Your Data?," students are introduced to the need to protect participant privacy while collecting, analyzing and interpreting data. They compare small, local datasets (small group and classroom) with the much larger set in the Data Browser. In "Research with the Data Browser," students make predictions about anticipated prevalence of several health conditions, analyze data from the Data Browser to calculate actual prevalence, and evaluate questions for future research. Three example learning sequences illustrate ways instructors might use the materials in classes ranging from biology to health to social science. We developed a 20-item instrument to assess students' learning. This process included 2 phases of classroom pilot testing and cognitive interviews with students to identify any questions that might be confusing or misunderstood. Students showed strong understanding of the definition of "big data" (82% correct), the need to use computers to analyze big data (86%) the

importance of data diversity and sample size (80%) and the main types of data in the Data Browser (86%). However, only 55% could recall that the Data Browser incudes data on environmental pollutant exposure. We will report findings from a national field test using the validated instrument, which is in progress. The “Exploring Big Data with the *All of Us* Data Browser” module is freely available on the Teach.Genetics website.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7069W: Enhancing Breast Cancer Awareness Among College Students: A Cross-Sectional Study in Nigeria

Authors:

S. Pal, M. Hashim, M. S. Muhammad; Skyline Univ. Nigeria, Kano, Nigeria

Abstract:

Breast cancer remains a significant health challenge for women globally, with Nigeria experiencing a particularly high incidence rate which is currently 70/100,000 (PMID: 35936820), is the leading cause of cancer-related death in Nigeria, has risen dramatically over the last few years. This study aimed to assess the impact of a breast cancer awareness lecture on the knowledge and awareness levels among college students, specifically focusing on the School of Science and Information Technology (SSIT) at our university in Nigeria. A cross-sectional study was conducted among 450 freshman students, with data collected before and after the delivery of a breast cancer awareness lecture. Results revealed a concerning lack of knowledge and awareness regarding breast cancer among the students, with only a small percentage correctly identifying early signs, symptoms, and risk factors. Additionally, a significant portion of the respondents were unaware of breast self-examination (BSE) practices. However, post-lecture data demonstrated a substantial improvement in knowledge and awareness levels, with almost all students correctly answering questions related to breast cancer. This underscores the effectiveness of educational interventions in increasing awareness and understanding of the disease. Furthermore, cultural, and religious beliefs pose challenges to open discussions about breast cancer, highlighting the importance of culturally sensitive approaches to education and awareness campaigns. We recommend a focused approach to Triple-Negative Breast Cancer (TNBC) due to its status as one of the most aggressive forms of breast cancer, particularly prevalent in underdeveloped countries like Nigeria where it's more fatal than in the United States. This recommendation is imperative to address the urgent need for improved diagnostic and treatment strategies tailored to combat the aggressive nature of TNBC, especially in regions

with limited resources and healthcare infrastructure. In conclusion, this study emphasizes the pivotal role of education and awareness initiatives in combating breast cancer in Nigeria. By equipping college students with accurate information and promoting early detection practices, we can significantly contribute to reducing the burden of this disease in our communities.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7070W: Evaluating community perceptions and ethical considerations in genetics research in small scale northern Kenyan populations

Authors:

R. Siford[1,2], C. Handley[1], S. Mathew[1,2], M. Wilson[3], J. Robert[3], M. Y. Akinyi[4], J. Kamau[4], A. Stone[1,2]; [1]Sch. of Human Evolution and Social Change, Arizona State Univ., Tempe, AZ, [2]Inst. of Human Origins, Arizona State Univ., Tempe, AZ, [3]Sch. of Life Sci., Arizona State Univ., Tempe, AZ, [4]One Hlth.Ctr., Kenya Inst. of Primate Res., Nairobi, Kenya

Abstract:

Historically, minority populations and Indigenous peoples have borne the brunt of poorly implemented regulations to protect human subjects in research. The growing field of research ethics has produced many ethical codes and suggestive guidelines from the perspectives of Western and Indigenous populations from the Americas and Australia, but almost none include the perspectives of African populations. As genomic research expands to be inclusive of global diversity, Indigenous perspectives about genetic research and data management should be considered, and the success of dissemination efforts to participants should be evaluated. In Kenya, where participation in genetic research is amongst the highest in Africa, it is also one of the most understudied for participant perspectives from small scale societies. Here, we address how Western ethical practices used in genetic studies accurately align (or fail to align) with the understandings, attitudes, and perceptions of the Turkana, Borana, Rendille, and Samburu pastoral populations of northern Kenya. Ongoing genetic research by our group has been conducted in collaboration with these populations which facilitates the unique opportunity to explore ethical practices and concerns. We collected semi-structured interviews from previous genetic study participants and from those who had not previously participated, to explore the effectiveness of dissemination efforts and to characterize in-depth views on privacy, consent, data use, and governance of data. From over 550 interviews collected in 2023, we generally find across all populations that dissemination of genetics results is expected and should be

presented with cultural and religious considerations. Also, broad consent is an inappropriate model and permission for new studies should be sought. Finally, communities favor a participatory model of research. We provide a general set of guidelines and population specific governance frameworks that may aid Kenyan research ethics committees and other researchers. Our study highlights that participant and community perspectives offer culturally relevant motivations and are necessary to ensure ethical governance of genetic data.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7071W: Evaluation of Somatic Genetic Testing as a Screening Tool to Detect Hereditary Cancer Predisposition in Patients Diagnosed with Myeloid Malignancies

Authors:

F. Quintero-Rivera[1], R. Collier[1], K. Singh[1], P. Flodman[2]; [1]Univ. of California Irvine (UCI), Irvine, CA, [2]Univ. of California Irvine (UCI), Irvine, CA, Irvine, CA

Abstract:

A referral to genetic counseling is commonly made for patients with many types of cancer, given that cancers are hereditary in 5-10% of cases. However, patients with myeloid malignancies are not often provided with genetic services despite emerging evidence that these cancers are hereditary about 13% of the time. Prompt genetic evaluation is indicated due to the mortality rate associated with a diagnosis of a myeloid malignancy.This study explored the use of somatic genetic testing (NGS) as a screening tool for hereditary cancer predisposition in patients diagnosed with specific myeloid malignancies (AML, MPN, AMML, MDS, and CMML) at our academic center, in the hopes of providing more evidence that this is an appropriate way to identify patients in need of genetics services. The study involved 115 participants whose medical records and NGS results were analyzed for criteria that indicate a need for genetic counseling and germline genetic testing. Furthermore, this study investigated the current landscape of genetics professionals' involvement in the care of patients with myeloid malignancies at a single institution. Results from this study showed that 75 out of 115 patients (65%) met criteria for germline genetic testing to assess hereditary cancer risk based on the National Comprehensive Cancer Network 2023 guidelines and the 2022 European LeukemiaNet guidelines. However, none were referred for genetics services on the basis of their personal history of myeloid malignancy. Additionally, of those 75 patients, 18 (16%) would not have met those criteria without considering the results of their somatic NGS testing, all of which included

SNVs that could be germline in origin, in a clinically relevant gene (VAF >30%), and as such, would be clinically relevant. About half of participants (44%) were deceased at the time of this study. This indicates a need for greater awareness among providers (hematologic oncologists, genetic counselors, molecular pathologists, and molecular geneticists) concerning the current recommendations for germline genetic testing for patients with myeloid malignancies. The findings of this study show that utilization of somatic genetic test results is merited as a resource and screening tool for determining whether a patient who may be missed by a traditional assessment of personal and family history should receive timely genetics services.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7072W: Filling the pipeline: The Greenwood Genetic Center's approach to developing the future genetic workforce.

Authors:

L. Tribble, A. LaVerdure, H. Newton, A. Labdon; Greenwood Gen Ctr, Greenwood, SC

Abstract:

Filling the pipeline: The Greenwood Genetic Center's approach to developing the future genetic workforce.
BackgroundWorkforce development is a high priority in the South Carolina education system. The Profile of the South Carolina Graduate emphasizes rigorous standards for career and college readiness including skills in STEM, critical thinking, teamwork, problem solving, and self-direction. The Bureau of Labor Statistics identified the fastest growing occupations projected through 2031 which include many careers in the healthcare field. Healthcare practitioners and technical occupations are projected to increase by 8.6%. To help meet this increase in demand, the Greenwood Genetic Center's (GGC) Outreach Education Program works to introduce younger students into the genetics workforce pipeline.
ObjectivesThe GGC Division of Education provides outreach education opportunities and programs for students and teachers with the goals of enhancing genetic literacy, introducing the field of biotechnology, and exposing students to careers in genetics and the life sciences.
Methods and Design The Outreach Education Program provides a wide array of biotechnology and genetics-based activities for teachers and students in grades 7-12 through a mobile science lab, the Gene Machine, and a transit van, the Helix Express. The program visits schools across SC with a focus on rural and underserved communities. All lab and safety equipment is provided including micropipettes, gel boxes, thermocyclers and centrifuges. During each visit, school and

student demographic information is recorded, and students complete a pre-test assessment. After each class, students complete a post-activity questionnaire and an anonymous survey that addresses interest in STEM topics and careers as well as self-efficacy in STEM classes. A month after the visit, teachers receive a voluntary survey to assess the quality of the activity as well as the experiences of both the students and teachers in working with GGC.
ResultsDuring the 2021-2022 school year, 42% of the outreach trips were made to rural districts with an increase in career interest and self-efficacy across all ethnic populations. Students reported 91% agreement with the statement that they feel confident in the lab setting in the post-activity questionnaire. Teachers reported 100% agreement that the activities were enriching and that students had a very positive experience with GGC. Teachers described value in partnering with GGC in a number of key areas including student interest, student preparedness post high school, and exposure to real-world careers.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7074W: From Oversight to Opportunity: Redefining Rural Inclusivity in Genetic Research

Authors:

P. Swanson; Univ. of Texas Med. Branch, Galveston, TX

Abstract:

Exposure to pesticides, sunlight, and well water heightens cancer susceptibility, particularly affecting rural individuals with ties to agriculture. Rural populations exhibit distinct genetic and epigenetic profiles shaped by environmental factors. Despite these unique identifiers, rural communities are often overlooked in genomic and cancer studies, compromising data accuracy and therapeutic efficacy. Advocacy initiatives have failed to adequately address this inequity. A comprehensive literature review spanning 2012-2022 confirms the persistent exclusion of rural participants from research endeavors.This study advocates for the extension of the NIH Revitalization Act of 1993 to encompass rural populations, thereby ensuring the inclusivity of data and the development of treatments more aligned with the broader U.S. population. By adopting a justice and inclusion-centered approach, this research demonstrates the importance of representative data in achieving equitable distribution of research benefits and burdens. As rural America becomes more diverse, it is imperative to recognize this demographic as a dynamic cohort with unique gene-environment interactions specific to their geographic

location. While emerging genomic and cancer therapies have the potential to exacerbate health disparities, inclusive methodologies can mitigate these gaps, as demonstrated in this study.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7075W: Genetics as a growing dimension of group identity: A study using Wikipedia data

Authors:

A. Diaz-Papkovich, S. Ramachandran; Brown Univ., Providence, RI

Abstract:

Genetics has long been used, with negative systemic repercussions, to describe facets of individual and collective identity. The internet has democratized access to genetic research, but our understanding of the spread and reach of our research is not well-understood. We study the use of genetics in group identity through one of the cornerstones of the modern information ecosystem: Wikipedia. As one of the first search engine hits and largest text corpora used by large language models and chatbots (e.g. ChatGPT and Microsoft Copilot) Wikipedia articles significantly influence what people learn about concepts such as race, nationality, and ethnicity. Many Wikipedia articles discuss and cite medical genetic and population genetic research. This may take form as a section titled "Genetics" or "Origins", as an article on the genetics of a group, or as entire collections of Wikipedia articles about genetic histories. Since Wikipedia articles can be written by anyone and full revision histories are retained, we use its data to study how identity is described through the lens of genetics and how this has evolved over the last two decades.

We use the Wikipedia API to collect data from over 5000 Wikipedia pages related to ethnic, racial, linguistic, national, religious, and demographic groups ("populations"), dating back to the site's creation in 2001 and spanning over 2 million revisions. We mine article text for strings such as "genetic", "haplo", and "autosom" to track how genetics influences popular conception of identity. We find that the number of genetics terms in population articles has grown linearly over time, having been added by approximately 6,000 users across 14,000 revisions; that now approximately 20% of all population articles cite or discuss genetics, including half of the 1,000 oldest articles about populations; and that this content tends to become a permanent part of an article once it is added. Within articles, we find that text about genetics is becoming longer and more prominent, suggesting that the popular conception of a "population" is now more likely to incorporate genetics.

We provide examples of how Wikipedia text about the genetics of populations is used by modern generative AI to answer questions about identity. We also study the influence of specific scientific publications and biobanks, the activity of Wikipedia contributors who write about genetics research, and discuss the implications of how genetics research is cited, understood, consumed, and reflected outside of scientific venues.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7076W: Hands-on educational approaches for teaching and learning pharmacogenomics

Authors:

A. Tapanes-Castillo[1], J. Penate[1], A. Uzakova[1], R. Thomas[2], A. Rygh[1], S. Pizarro[1], R. Daniel[1], G. Espinoza[1], B. Castro[1], D. Hadley[1], C. Del Corro[3], E. Iglesias[2,3], M. Lee[3]; [1]St. Thomas Univ., Miami Gardens, FL, [2]Miami Dade Coll., Miami, FL, [3]Galatea Bio, Miami Lakes, FL

Abstract:

The purpose of our research was to explore hands-on educational approaches to improve an understanding of pharmacogenomics. One approach, involving a novel craft-based game, was tested on a general audience during a community science education event in Miami-Dade County, Florida, following Institutional Review Board (IRB) approval. Thirty-three individuals ranging in age from 8-40 years old participated. The median age was 11 years old, while the average age was 13. Respondents identified as 48% male, 52% female, 55% white, 23% black, 61% Hispanic, and 23% speakers of English as a second language. Participants anonymously rated the activity's educational impact: (3) meant they learned something new, (2) indicated they heard about topics they already knew about, and (1) signified they did not understand the concept or activity. Survey topics and the percentage of participants awarding each score follows. Topic 1 was DNA structure and the relationship between genes and proteins: (3) 57.6% participants, (2) 42.4%, and (1) 0%. Topic 2 was differences in DNA between people, such as alleles and single nucleotide polymorphisms (SNPs): (3) 36.3%, (2) 57.6%, and (1) 6.1%. Topic 3 was how DNA can affect an individual's response to medications, including the definition of pharmacogenomics: (3) 78.7%, (2) 15.2%, and (1) 6.1%. Topic 4 was scientific methods used to ascertain DNA sequences, such as Taqman SNP genotyping assays: (3) 87.9%, (2) 12.1%, and (1) 0%. Topic 5 was overall feelings regarding the activity: (3) 87.9%, (2) 12.1%, and (1) 0%. Significantly, our pilot study provided a novel approach for educating diverse, general

audiences, especially youth, about genomics, as well as its related technologies and clinical applications.
A second distinct approach was utilized to engage undergraduate science students. For this study, students analyzed the prevalence of *CYP2C19* SNPs in a Miami-Dade County population sample. DNA was collected from buccal swabs following IRB approved protocols. TaqMan assays were used to determine the genotype and expected clinical phenotype of 27-30 individuals at three SNPs. Hardy-Weinberg equations were applied to estimate genotype and phenotype frequencies. Preliminary results, using Chi-square tests, suggested the local Miami-Dade County population exhibited similar allele frequencies as those reported globally. This activity enhanced students' understanding of pharmacogenomics, population genetics, and genotyping assays through the development of technical skills, critical thinking, and presentation of the project at local undergraduate symposia.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7077W: Hereditary cancer syndrome registry at an urban safety net institution

Authors:

M. Smith[1], S. Holmes[2], C. Hussain[1,2], S. Knerr[3], S. Okuyama[1,2]; [1]Univ. of Colorado, Aurora, CO, [2]Denver Hlth., Denver, CO, [3]Univ. of Washington, Seattle, WA

Abstract:

Background: Patients with hereditary cancer syndromes (HCS) have a high lifetime risk of cancer and require preventive care that is intensive, complex, and ongoing. Denver Health as a safety net institution does not have an in-house cancer genetics program to manage patients with HCS. In response, we have created a registry of patients with HCS in our institution's electronic health record to facilitate care coordination. Here, we describe the characteristics of patients in this registry.
Methods: The registry includes all patients who receive primary care at our institution with a diagnosis of HCS, regardless of personal cancer history. Patients were identified by ICD-10 codes indicating genetic susceptibility to malignant neoplasm (Z15.01, Z15.02, Z15.03, Z15.04, Z15.09) and confirmed to carry a pathogenic or likely pathogenic (P/LP) variant through chart review. Patients identified to only have variants of uncertain significance or variants not associated with HCS were excluded from the registry. Demographic and clinical data were abstracted to describe

the registry population and determine demographic factors associated with a likelihood of being overdue for cancer screening.
Results: The registry includes 180 patients with P/LP variants in *BRCA1, BRCA2, CHEK2, ATM, PALB2, MLH1, MSH6, PMS2, BRIP1, MSH2, APC, MITF, TP53, CDH1, CDKN2A, FH, HBOX13, NF1, RB1,* and *RET* (in order of frequency). Registry patients are 86% female, 76% White or Caucasian, 6% Black or African American, 2% Asian, and 1% American Indian or Alaska Native. Almost half of patients are Hispanic or Latino and 20% prefer to speak Spanish. Overall, 43% of patients have Medicaid, 18% have insurance provided by Denver Health, 18% have private insurance, 10% have Medicare, and 10% are uninsured. Ages range from 18 to 90 and the median age group is 40-45. Approximately half have no personal cancer history. Regarding the relationship between demographic factors and screening status, having no personal cancer history (versus prior cancer) was associated with being overdue for cancer screening [odds ratio (95% confidence interval)] [2.38 (1.25, 4.54)]. Hispanic or Latino ethnicity was inversely associated with being overdue for screening [odds ratio (95% confidence interval)] [0.59 (0.37, 0.95)]. There was no difference in screening status based on gender, race, preferred language, or insurance payor.
Conclusions: We describe for the first time the characteristics of a registry of patients with HCS built within the electronic health record of a safety net institution. This registry will be used to support patients and their providers in the complex cancer risk management needs of people with HCS.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7078W: Identifying Precision Medicine Program strengths and areas of improvement by curriculum mapping to the core genetics and genomics competencies by the Association of Professors of Human and Medical Genetics (APHMG)

Authors:

R. Ventrella[1], K. Kristjansdottir[1], Y. He[1], G. Jentarra[2], K. Bussey[2]; [1]Midwestern Univ., Downers Grove, IL, [2]Midwestern Univ., Glendale, AZ

Abstract:

Precision medicine is an emerging field that necessitates a comprehensive understanding of genetic, genomic, and clinical data to tailor medical care to individual patients. Since future

healthcare providers must be adequately trained in these topics, precision medicine programs and courses are now offered in many healthcare disciplines, including medicine, physician assistant, pharmacy, and dental medicine. To align educational efforts with the needs of this evolving discipline, the Midwestern University Precision Medicine Program's asynchronous online curriculum was mapped to the Association of Professors of Human and Medical Genetics (APHMG) core competencies, which are divided into three focus areas: medical knowledge, patient care, and communication and professionalism. These categories are subdivided into sixteen core competencies, each with its own set of learning objectives. To optimize the educational approach, this study used the APHMG core competencies to identify strengths and gaps in student learning in our Precision Medicine Program. To achieve this, we performed a detailed comparison of precision medicine courses in our program to the APHMG core competencies, including lecture material and assessments. We demonstrated that our core curriculum builds throughout the program. It begins with foundational genetic concepts, such as genetic variation, inheritance, and molecular genetics, and ends with patient-focused skills, such as selecting genetic tests, treatment management, genetic database investigation, and genetic counseling. We administered an evaluation to students who had completed the Precision Medicine Program to identify their confidence in each competency. Results from the evaluation demonstrated a high confidence in all the core competencies, especially in patient care. However, gaps were identified in population genetics and molecular genetics. These findings suggest a need for curricular adjustments to enhance student competency in 1) understanding how population genetics can be used to predict disease risk and 2) deciding how to choose which genetic test should be used in specific clinical situations. Targeted interventions to address these weaknesses will better equip future healthcare professionals to practice precision medicine more effectively. In summary, mapping a precision medicine curriculum to the APHMG core competencies provided valuable insight into student preparedness to use precision medicine in the clinic and highlighted areas for improvement. This process allows for continuous curriculum development in response to the evolving landscape of medical genetics education.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7079W: Impact of polygenic and monogenic risk results on patients' perceived risk of hypercholesterolemia in the eMERGE IV study

Authors:

P. Marathe[1], S. A. Suckiel[1,2], L. Scarimbolo[1], J. A. Odgis[1], E. R. Soper[1,2], K. E. Bonini[1], S. Nirenberg[3], E. J. Gallagher[2,4,5], A. R. Kontorovich[1,2,6,7], M. F. Murray[1,2], N. S. Abul-Husn[1,2], E. E. Kenny[1,2,6]; [1]The Inst. of Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Scientific Computing and Data Div., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Div. of Endocrinology, Diabetes and Bone Disease, Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Tisch Cancer Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Cardiovascular Res. Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Understanding perceived risk for common diseases is crucial for developing effective patient-centered approaches to clinical and lifestyle interventions in diverse populations. The eMERGE IV study provides participants with a Genome Informed Risk Assessment (GIRA) comprising genomic risk, family history, and clinical risk factors for hypercholesterolemia (HCL). High-risk GIRAs are returned if the participant has a PRS in the top 3%-ile for HCL (PRS+) and/or a positive monogenic result for Familial Hypercholesterolemia (FH+). The report contains a description of the condition, result explanations, and related medical recommendations. We evaluated participants' perceived risk of HCL before and after results disclosure, comparing PRS+ or FH+ to those who did not have high genomic risk (non-high-risk).

Participants completed a baseline survey assessing perceived risk of HCL and another survey at six months (6m) post-results disclosure that reassessed perceived risk and recorded actions taken due to the GIRA. Our analysis included 8,800 adults enrolled in eMERGE IV. PRS and monogenic data were obtained from lab reports, and participant demographic data was collected at baseline. We used the Mcnemar-Bowker exact test to compare perceived risk at baseline and 6m post-results.

Participants' average age was 51 (range 18-75), 65% were female, and most self-reported as White (46%), Hispanic/Latino(a) (23%), or Black (17%). Of the participants who completed the follow-up survey (N=1201), FH+ participants perceived a higher risk of HCL 6m post-results compared to baseline (N=8; p=0.046). There was no significant change in perceived risk among PRS+ participants (N=21; p=0.074). Many PRS+ participants (12/21; 55%) discussed their result with a doctor. Of those, 10/12 (83%) reported their doctor made recommendations based on the GIRA. Some PRS+ participants (5/21; 24%) made medical/lifestyle changes in addition to those recommended by their doctor. In comparison, only 3/8 (38%) FH+ participants shared their results with a doctor. 19/21 (86%) PRS+ and 7/8 (87%) FH+ participants shared or planned to share their results with family members.

These preliminary findings suggest that a comprehensive risk assessment including both PRS and monogenic results can effectively educate individuals about their risk of developing HCL,

though not all act on their high-risk results. Collection of post-results data is ongoing, and future analysis will focus on perceived risk and additional outcomes for 20,000 eMERGE IV participants.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7080W: Personal utility of genetic testing depends on patient population: Evidence from four clinical cohorts

Authors:

E. Poole[1], S. Luca[1], D. Assamad[1], B. Xiao[1], J. Yan[1], C. Xia[2], W. J. Ungar[1], L. S. Abbott[3], L. Armstrong[4], P. Birch[5], K. M. Boycott[3], J. C. Carroll[6], L. Chad[1], D. Chitayat[2], A. Denburg[1], R. J. Deyell[4], A. M. Elliott[5], C. Goudie[7], A-M. Laberge[8], M. Maio[9], I. Peltekova[10], B. Quinlan[11], S. L. Sawyer[12], R. Silver[13], M. Smith[14], R. Teitelbaum[15], A. Villani[1], T. Wainstein[16], R. Z. Hayeems[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [4]British Columbia Children's Hosp., Vancouver, BC, Canada, [5]Univ. of British Columbia, Vancouver, BC, Canada, [6]Sinai Hlth., Univ. of Toronto, Toronto, ON, Canada, [7]McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [8]CHU Sainte-Justine, Montreal, QC, Canada, [9]Mackenzie Richmond Hill Hosp., Toronto, ON, Canada, [10]Holland Bloorview Kids Rehabilitation Hosp., Toronto, ON, Canada, [11]Ontario Genetics Advisory Committee, Toronto, ON, Canada, [12]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [13]Mount Sinai Hosp., Toronto, ON, Canada, [14]Canadian Organization for Rare Disorders, Toronto, ON, Canada, [15]Medcan, Toronto, ON, Canada, [16]British Columbia Children's Hosp., Vancouver, BC, Canada

Abstract:

Introduction: With the increased use of genetic testing in clinical care, it is critical to consider personal utility to inform care and policy decisions. We aimed to understand patient-oriented values and preferences for genetic testing across patient populations and life stages.
Methods: We identified four unique cohorts to understand if and how personal utility of genetic testing differs: prenatal genetics, pediatric genetics, pediatric oncology, and adolescent genetics. A preliminary set of generalizable domains and items of utility were generated from a scoping review of the literature. Supplemental evidence synthesis was conducted for each cohort. Participants were recruited from 6 centres across Canada. Cognitive interviews were conducted with parents/caregivers or adolescents from the four cohorts. Participants were asked to reflect on their experience with genetic testing, the meaning of personal utility, and the relevance of

the identified items of utility. Interviews were audio-recorded and transcribed verbatim. Data were analyzed thematically and reviewed to identify shared and unique items of utility.
Results: Seventy-two interviews have been conducted to date (prenatal = 20; pediatrics = 21; oncology = 20; adolescent = 11). Most participants were female (81%), and genetic testing was diagnostic for 35%. Domains of personal utility related to the cognitive, affective, behavioural, social, and medical management impacts of genetic testing resonated with participants. Items that aligned across cohorts included increased knowledge about the reason for their/their child's condition, communicating results with others, and contributing to scientific knowledge. In prenatal genetics, items related to pregnancy decision-making emerged as unique. Parents in the pediatrics cohort expressed value in ending the diagnostic odyssey. Knowledge about cancer risk and recurrence was salient for parents in our pediatric oncology cohort. Adolescents highlighted the importance of autonomy and being able to play a more active role in their healthcare decisions through receiving genetic results.
Conclusions: While some items of personal utility are shared across clinical cohorts and life stages, we identified unique, cohort-specific items that are valuable to patients and their families. To effectively capture and integrate evidence of personal utility into clinical care and policy decision-making, there is a need to consider when, why, and for whom genetic testing was ordered. Future work will expand to additional clinical cohorts, such as pediatric intensive care and newborn screening.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7081W: Population Descriptors in Genomic Legacy Data: Reflections and Recommendations for Future Research Directions from a Trans-NIH Workshop ⭐

Authors:

L. Hindorff[1], J. Morales[2], C. Daulton[3], S. Hutchison[1], E. Mfwilwakanda[2], J. Chinn[4], E. Gillanders[5], C. Jacobs[6], S. Jose[7], J. W. King[3], M. Lowden[8], E. Madden[1], L. Mechanic[9], R. Rivers[10], E. Spotts[3], F. Tyson[11], E. Boerwinkle[12], K. Brothers[13], K. Chao[14], P. Kraft[15], K. Miga[16], A. Musick[17], J. Novembre[18], C. Rotimi[19], J. Witte[20], G. Wojcik[21], P. Adkins-Jackson[22], J. Cadigan[23], M. Cho[20], K. Claw[24], D. Conti[25], E. Ghanaim[26], C. Gignoux[27], S. Gogarten[28], E. Burchard[29], L. Harris[30], A. Hendricks[31], A. Ioannidis[32], E. E. Kenny[33], I. Konigsberg[34], L. Landry[35], A. Lewis[36], S. Nelson[37], A. Popejoy[38], M. Sabatello[22], L. Starita[37], S. Schully[39], V. Bonham[2], M. B. Davis[40], S. M. Fullerton[28]; [1]NIH/NHGRI, Bethesda, MD, [2]NHGRI, Bethesda, MD, [3]NIH, Bethesda, MD, [4]NIH/NICHD, Bethesda, MD, [5]NCI, Bethesda, MD, [6]NIH, Bethesda, MD, [7]NHGRI, Rockville,

MD, [8]NIH/NINR, Bethesda, MD, [9]NCI DCCPS EGRP, Bethesda, MD, [10]NIH/NIDDK, Bethesda, MD, [11]NIH/NIEHS, Durham, NC, [12]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [13]Norton Children`s Res. Inst., Louisville, KY, [14]Broad Inst. of MIT and Harvard, Cambridge, MA, [15]Natl. Cancer Inst., Rockville, MD, [16]UCSC, Santa Cruz, CA, [17]NIH - All of Us Program, Bethesda, MD, [18]Univ Chicago, Chicago, IL, [19]NIH, Woodstock, MD, [20]Stanford Univ., Stanford, CA, [21]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [22]Columbia Univ., New York, NY, [23]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [24]Univ. of Colorado, Aurora, CO, [25]Univ of Southern California, Los Angeles, CA, [26]Natl. Human Genome Res. Inst. (NHGRI), Bethesda, NJ, [27]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [28]Univ. of Washington, Seattle, WA, [29]Univ California San Francisco, San Francisco, CA, [30]EMBL-EBI, Cambridge, United Kingdom, [31]Univ. of Colorado Denver, Denver, CO, [32]Stanford Univ., Oakland, CA, [33]Icahn Sch. of Med. at Mount Sinai, New York, NY, [34]Univ. of Colorado, Denver, CO, [35]Univ. of Pennsylvania, Philadelphia, PA, [36]Brigham and Women's Hosp., Boston, MA, [37]Univ of Washington, Seattle, WA, [38]Univ. of California, Davis, Sacramento, CA, [39]All of Us Res. Program, NIH, Bethesda, MD, [40]Morehouse Sch. of Med., Atlanta, GA

Abstract:

The appropriate use of population descriptors in genomics research is of critical importance and has significant scientific and ethical implications for participants, communities, researchers, and policymakers. Despite decades of research and discussion, there are still areas where decisions to describe populations remain fraught. A 2023 National Academies of Sciences, Engineering and Medicine (NASEM) report identified best practices in using population descriptors and offered recommendations and guiding principles for research design and analysis. However, the report did not go into detail regarding approaches to population description for existing (legacy) genomic data. Legacy data are subject to differences in informed consent understandings and approaches to grouping individuals that may lack needed granularity. Efforts to address concerns are hindered by inability to directly contact study participants, due to factors such as time elapsed or de-identification of participants. To consider applying the NASEM report recommendations to genomic legacy data, the National Institutes of Health (NIH) held a virtual workshop in May 2024. The workshop objectives were to: 1) Consider how the NASEM report recommendations relate to legacy data; 2) Summarize current approaches to the use of population descriptors in legacy data; 3) Define challenges with current approaches to harmonization, interoperability, and analysis; 4) Develop practical recommendations that can be widely adopted; and 5) Identify opportunities for future research or collaboration. Through talks, panels, and breakout groups, workshop attendees considered key challenges and discussed how the NASEM framework might be adapted in distinct research use cases. Recommendations from lessons learned with legacy data include: thoughtful study design and transparent methodological justification; expanded engagement with participant communities

throughout the research cycle; researcher education to avoid misuse of labels; the key role and responsibility of funders, policy makers, providers and journalists in promoting responsible use; and the need to develop guardrails and guidelines that can be widely adopted while encouraging diversity and inclusion among study participants. By including an interdisciplinary group of attendees with varied perspectives, this workshop catalyzed conversation and facilitated the development of guiding principles for the scientific community, including researchers, funding agencies, and journal editors, to improve the use of population descriptors for legacy genomic data.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7082W: Scaling genetic counselling services across medical specialties in Singapore

Authors:

B. Cham[1,2], SingHealth Duke-NUS Genomic Medicine Centre, E. Tan[1,2], J. Lim[1,2], S. Kam[1,2], Y. Bylstra[3,2], S. Jamuar[1,2,3]; [1]KK Women's and Children's Hosp., Singapore, Singapore, [2]SingHlth.Duke-NUS Genomic Med. Ctr., Singapore, Singapore, [3]Inst. of Precision Med., Singapore, Singapore

Abstract:

With the greater availability of new methods of molecular genetic diagnostics, the use of genetic testing is increasing across non-genetics specialties. Genetic counseling plays a pivotal role in the era of precision medicine, where genomic information informs the risk assessment, diagnosis and treatment options across various medical specialties in addition to traditional settings of pediatric and cancer genetics services. Scaling genetic counseling service delivery effectively is imperative to meet the needs of an increased demand in diverse patient populations.

As part of the SingHealth Duke-NUS Genomic Medicine Centre, a hub and spoke model of service delivery was developed, where expertise of genetic counselors from an established genetics service was combined with expertise of specialty care physicians across different institutions in the healthcare group to provide genetic healthcare. The aim of embedding genetic counselors in specialty care centres was to provide better access to genomic healthcare by ensuring comprehensive and integrated service delivery.

Clinics started in 2020 covering cardiogenetics, neurogenetics and ophthalmic genetics. The following year, this expanded to include obstetric and reproductive medicine, and

endocrinology. This was followed by the establishment of genodermatoses clinics in 2022. Over a period of four years since the inception of specialty-based genetic counseling clinics in 2020, 993 patients received genetic counselling services. Of these patients, 617 patients were evaluated for a possible genetic condition and offered genetic testing, with 342 patients opting to undergo genetic testing. 256 appointments were conducted to return genetic test results to patients that had undergone testing in the clinic and in other non-genetic settings.
Through a collaborative multi-disciplinary approach to genetic counseling service provision, access to appropriate genetic services can be extended beyond the genetics clinic. Service delivery is streamlined by reducing patient transfer across different specialties and healthcare institutions, ensuring consistency and continuity of care for the patient. This patient-centred approach allows services to be tailored to the individual, fostering engagement and informed decision-making. Some limiting factors that need to be considered are the size of the available genetic counselor workforce and the need for greater genetic literacy among specialist healthcare providers to enable integration of genomics into clinical practice. The experience informs future expansion of service provision and integration of genetic services in Singapore.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7083W: Teaching the Genome Generation: Cultivating High School Genomics Through Teacher Education

Authors:

C. Wray[1], C. Vallianatos[2], E. Gerace[2], A. Wnorowski[2], J. Somerset[1], S. Wojiski[2]; [1]The Jackson Lab., Bar Harbor, ME, [2]The Jackson Lab., Farmington, CT

Abstract:

Teaching the Genome Generation (TtGG) provides pre-service and current high school teachers with the content knowledge, teaching strategies, and resources needed to enhance student learning in genetics and genomics, with an emphasis on quantitative skills and data literacy. Our approach weaves three learning strands—human molecular genetics, bioinformatics, and bioethics—together within the context of the Next Generation Science Standards (NGSS) and Common Core Math Standards. To date the program has trained 300 teachers who have delivered TtGG content to more than 22,000 students. 13 teachers and 878 students participated in a 2022 classroom study of TtGG curricular content. Key findings from the classroom study indicate the use of our curriculum leads to increased knowledge of, and confidence in, biotechnology laboratory procedures and increased student interest in genetics

and genomics concepts. Newly designed modules emphasize quantitative skills in four genetics- and genomics-related content areas: cancer genetics, sequence comparison and identity, genotype-phenotype associations, and ancestry. Each module includes standards-aligned activities focused on problem-solving or inquiry through the exploration, analysis, and interpretation of genetic and genomic data. Modules also encourage exploration of ethics and social justice topics relevant to modern genetics and genomics, such as health disparities and access to genetic testing. Our new content is freely available on our website and includes teacher guides and multimedia tutorials for new tools and skills. Recently, our resources have expanded to include a suite of laboratory simulations and interactives, hosted on LabXchange and available to teachers around the world.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7084W: The challenge of genetic counselling in Xp22.31 related CNVs

Authors:

I. Focsa[1,2], L. Dimos[2], P. Gurban[2], A. Pavel[2], L. Bohiltea[3], F. Burada[4,5], M. Budisteanu[6,7], D. Stambouli[2]; [1]Univ. of Med. and Pharmacy Carol Davila, Bucharest, Romania, [2]Cytogenomic Med. Lab., Bucharest, Romania, [3]Univ. of Med. and Pharmacy “Carol Davila”, Bucharest, Romania, [4]Univ. of Med. and Pharmacy Craiova, Craiova, Romania, [5]Regional Genetic Ctr. Dolj, Craiova, Romania, [6]Univ. of Med. Titu Maiorescu, Bucharest, Romania, [7]Clinical Hosp. of Psychiatry "Prof. Dr. Alexandru Obregia", Bucharest, Romania

Abstract:

Objective: Advance of genomic technologies leads to identify a growing number of genetic aberrations, both in prenatal or postnatal stage. Copy number variants (CNVs) of the highly unstable Xp22.31 region were frequently described in various studies. Yet, there is no consensus in pathogenicity assessment of this prone-to-imbalanced region. While copy number (CN) loss are considered pathogenic in male, the pathogenic significance of CN gain is controversial. Moreover, the highly phenotypic variability among the carriers of the same size CNVs make the clinical presentation difficult to predict. Thus, an appropriate genetic counselling is always challenging, particularly in the prenatal stage. Here we present a case series with Xp22.31 CNVs. Material and methods: Seven cases, 5 fetuses and 2 children were referred for genetic testing due to different clinical features. The two children presented hypotonia, developmental delay, atopic dermatitis, cardiac and skeletal malformations. In the prenatal cases, genetic diagnosis

was indicated as a consequence of an increased biochemical risk, cardiac, renal or skeletal malformations. SNP array using Affymetrix technology (CytoScan 750K Arrays, GeneScan 3000) was used in six cases, while in the other case, sequence analysis of all protein coding genes in the genome for the proband, combined with Whole Exome Deletion/Duplication (CNV) Analysis using Illumina technology were done. Results: Three deletions, three duplications and one tetrasomy of Xp22.31 region were identified. The majority were the classical CNVs imbalanced of 1.69 Mb comprising six OMIM genes. In one child a 3.19 Mb deletion that included the critical region of Xp22.31 deletion syndrome was detected; in this child the WES test identified also two heterozygous variants of unknown significance in two neurodevelopment-related genes. Conclusions: The difficulties of counselling arise from the clinical heterogeneity of the cases or from the soft physical findings detected by fetal ultrasound scanning as well as from the extend of genetic aberration related to the phenotype and to the gender of the case. The influence of the additional genomic changes in term of modifier of phenotype cannot be excluded.Therefore, further comprehensive analyzes are needed to characterize the impact of CNVs of this unstable region in the human pathology and to facilitate a proper genetic counselling.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7085W: The development and preliminary evaluation of 'The Genetics Navigator': a digital solution for adult and pediatric genetics

Authors:

S. Saeedi[1], D. Hirijkaka[1], M. Clausen[1], S. Luca[2], E. Reble[1], R. Kodida[1], D. Assamad[1,2,3], L. Chad[4], G. Costain[4], H. Faghfoury[5], J. Silver[5,6], S. Shastri-Estrada[7,8], M. Smith[9], R. Hayeems[2,3], Y. Bombard[1,3], Genetics Navigator Study Team; [1]Genomics Hlth.Services Res. Program, Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [2]Child Hlth.Evaluative Sci., The Hosp. for Sick Children, Toronto, ON, Canada, [3]Inst. of Hlth.Policy, Management and Evaluation, Univ. of Toronto, Toronto, ON, Canada, [4]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, Toronto, ON, Canada, [5]Fred A. Litwin Family Ctr. in Genetic Med., Univ. Hlth.Network and Sinai Hlth.System, Toronto, ON, Canada, [6]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [7]Genetics Navigator Advisory Board, Toronto, ON, Canada, [8]Dept. of Occupational Sci. and Occupational Therapy, Temerty Faculty of Med., Univ. of Toronto, Toronto, ON, Canada, [9]Patient Partner, Canadian Organization for Rare Disorders, Toronto, ON, Canada

Abstract:

Background: Clinical genetic services encounter diverse patient populations, yet there is a notable absence of digital tools to support this spectrum of patients seen in genetic services. Our aim was to develop and perform usability testing for the Genetics Navigator (GN): a digital platform that aids in the delivery of genetic services for pediatric and adult populations with a wide range of clinical indications for genetic testing. Methods: We developed a prototype for the GN by consulting with an advisory Board of patients and genetics clinicians. We were also informed by a needs assessment involving patient interviews from previous research. We conducted two rounds of usability testing on the GN prototype with genetics patients, parents of pediatric patients, and the general public over Zoom. During the sessions, we used the 'think aloud' method to gather participants' reactions and feedback as they interacted with the GN. Participants completed the System Usability Scale (SUS) to measure the perceived usability of the GN, with a score above 68 representing above-average usability. Results: Using a pre-existing platform (the Genetics Adviser) as a base, we added written content, videos, and patient vignettes to encompass a wide variety of indications for genetic testing (e.g. neurodevelopmental disorders, cardiac conditions, cancer) and patient populations (e.g. adult and pediatric). The GN includes pre-test elements of counselling, patient education, and history gathering. Post-test elements involve results disclosure, medical management, and family communication. Usability testing participants were genetic patients (14/28), parents of pediatric genetics patients (4/28) or members of the general public (10/28). All participant types responded favourably to the GN. After usability testing, participants rated the GN an average SUS score of 87.7 ± 10.9, indicating above-average usability. Four common feedback themes emerged: 1) the importance of a supportive and inclusive design; 2) the principle of simplification and 'less is more'; 3) creating a seamless workflow; and 4) the importance of an engaging design. Conclusions: The Genetics Navigator is a comprehensive platform which can support pediatric and adult patients across various indications for genetic testing. Overall, users responded favourably and indicated an above-average score for usability. Participant perspectives indicated the importance of keeping digital platforms simple in terms of information and design and making sure a wide variety of users feel supported. Future work will evaluate the clinical and cost-effectiveness of the platform in a randomized clinical trial.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7086W: Utility of whole genome sequencing in pediatric neurological disorders in Karachi

Authors:

E. Hanif[1,2], R. Kumar[3], M. Mustafa[4], Y. Mariam[4]; [1]Univ. of Karachi, Karachi, Pakistan, [2]Assistant Professor, Dept. of Biotechnology, Univ. of Karachi, Karachi, Pakistan, [3]Assistant Professor & Paediatric Neurologist, Dept. of Paediatric , Med., Liaquat Med. Hosp., Karachi, Pakistan, [4]Dept. of Biotechnology, Univ. of Karachi, Karachi, Pakistan

Abstract:

Purpose: This study aimed to assess the effectiveness of Whole Genome Sequencing (WGS) in diagnosing and managing pediatric neurological disorders. Additionally, it sought to explore parents' understanding of WGS and their perspectives on its use in pediatric healthcare. Through this investigation, the research aimed to provide valuable insights into the utility of WGS in pediatric neurology and its implications for parental decision-making and healthcare delivery.
Introduction: Neurological disorders pose significant challenges in pediatric healthcare. Whole Genome Sequencing (WGS) offers a promising avenue for understanding the genetic underpinnings of these conditions. By analyzing the entire genome, WGS provides comprehensive insights into potential genetic variations and their implications for neurological disorders. This introductory exploration delves into the intersection of pediatric neurology and WGS, highlighting the potential of genomic sequencing technologies to enhance diagnostic accuracy, inform treatment strategies, and improve outcomes for children affected by neurological conditions.Methodology: In this study, 41 parents of children diagnosed with neurological disorders participated voluntarily. Both on-call and in-person interviews were conducted, with interviewers utilizing a structured questionnaire designed specifically for this study. To alleviate any potential burden on the parents, interviewers administered the questionnaire themselves, asking the predetermined questions and recording responses. This methodology ensured comprehensive data collection while respecting the parents' time and emotional well-being along with their privacy.Results: The collected data underwent analysis utilizing descriptive statistics and Microsoft Excel. This comprehensive analysis encompassed various aspects, including participants' social demographics, characteristics of the children, the efficacy of Whole Genome Sequencing (WGS), parental perspectives on WGS, alterations in child medication following WGS, and challenges faced by parents. Notably, WGS demonstrated a diagnostic success rate of 75.6% in identifying mutations associated with neurological disorders, underscoring its effectiveness in clinical practice.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Wednesday Session

Board 7087W: Validity of self-reported family history for hereditary cancer risk assessment in the primary care setting

Authors:

T. Ibidapo, S. Pirzadeh-Miller, S. Lahiri, S. Leach, S. Makhnoon; UT Southwestern Med. Ctr., Dallas, TX

Abstract:

Introduction: National guidelines recommend that primary care providers (PCP) perform hereditary cancer risk assessments. However, the collection of detailed cancer family history during a PCP visit is often not feasible due to factors including time constraints. To address this barrier and facilitate implementation of hereditary cancer risk assessments at a PCP clinic, we adapted a modified 7-item Family Health Screen (FHS-7) from PCP-report to patient-report. Although patient-reported data can be subject to error due to social desirability bias, recall, and knowledge of family health history, the validity of genetic risk assessments rely upon this data quality. **Methods:** Here, we evaluated the concordance between patient-reported cancer family history during a PCP visit and a genetic counseling (GC) appointment (the gold standard) from an implementation study at an academic medical center. The analysis included 9 items representing the 7 questions of FHS-7 and examined levels of concordance, variations by type of item, and sources of disagreement between data sources. We utilize measures of concordance, sensitivity, and specificity. This analysis focuses on patients who completed the self-reported 9-item assessment between February 2023 and March 2024 (n=4,540). **Results:** Overall, 339 (7.4%) of primary care patients also underwent genetic counseling for cancer risk assessment. The percentage total agreement on questions ranged between 67.2% and 96.2%, with specificities between 68% and 98%. Concordance was best for male family history of breast cancer (96%) and worst for family history of breast cancer in first degree relatives (67%). Positive predictive value was poor for all questions ranging from 17% for male family member with breast cancer to 70% for history of breast cancer in first degree relatives. Accuracy of self-report was predicted by female sex but not history of undergoing GC. **Conclusion:** Among primary care patients, self-reported cancer family history was generally accurate, although a subset of respondents overestimated the number of family members with breast, ovarian, and bowel cancers. Data supports the practice of collecting cancer family health history directly from patients for hereditary cancer risk assessment, thus making it feasible to implement hereditary cancer risk assessment at other settings with limited PCP involvement. Accurate measurement of family cancer history is a necessary first step towards improving genomic

medicine delivery and the results here have implications for quality hereditary cancer risk assessment.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8081W: Building a human genetics-guided side effect score for drug therapeutics

Authors:

A. Duffy[1], R. Chen[1], D. Stein[1], B. Petrazzini[1], J. Park[1], I. Forrest[1], H. Vy[1], M. Mort[2], A. Schlessinger[1], Y. Itan[1], D. Cooper[3], D. Jordan[4], G. Roccheleau[1], R. Do[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Inst. of Med. Genetics, Cardiff, United Kingdom, [3]Cardiff Univ, Cardiff, Wales, United Kingdom, [4]Icahn Sch. of Med. at Mt Sinai, New York, NY

Abstract:

Adverse side effects of drugs often remain undetected until the later stages of clinical trials, contributing to the failure rate and exorbitant costs associated with drug development. Several studies have shown that human genetic data can help predict the effects of lifelong modulation of a drug target. Leveraging this data at the preclinical stage of drug discovery offers an opportunity to prioritize safer targets. In this study, we developed an *in-silico* side effect score for 19,422 protein-coding genes and 470 drug side effects based on multiple lines of human genetic evidence to inform the likelihood of side effect occurrence for each drug target. We constructed an integrated drug genetics dataset using the Open Targets database for drug side effects and leveraged nine genetic features that encompassed clinical variants, coding variants and phenome-wide association studies. Using a mixed-effect regression model with five-fold cross-validation, we built the side effect score as the weighted sum of the effects of each of the five constructed genetic features with drug side effects, similar to the genetic priority score. We further incorporated the direction and mechanism of the drug into a complementary version of the score. Hence, this framework informs the likelihood of side effect occurrence based on the number and strength of each predictor with genetic support. We demonstrated that the side effect score had a 2.3-fold enrichment with side effect outcome and this enrichment increased when stratified to side effects that were more drug-specific and more severe. Furthermore, we stratified the score into bins of 0.3 and confirmed that gene-phenotypes with increased genetic support were associated with an increased likelihood of having a drug side effect. A score in the top 0.5%, 0.1% and 0.02% conferred a 2.1-, 3.2- and 5.6-fold increased likelihood of having a side effect, respectively. Importantly, the top scores were driven by multiple genetic features (mean=2.1). We validated these results using an independent drug database, OnSIDES and

observed similar results. Furthermore, we observed an increase in the proportion of side effects with an increase in the directional version of the side effect score. Finally, we highlight known side effect examples that are consistent with the direction and mechanism of action of the drug. In conclusion, we have developed an *in silico* side effect score using multiple lines of human genetic evidence that informs the likelihood of drug side effect occurrence for a large number of gene targets and diseases. This tool can be useful for prioritizating drug targets in drug development.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8082W: Deciphering the Therapeutic Effect of Dynamin-2 Reduction in SPEG-Related Myopathy: Perspectives from Multi-Omics Analysis

Authors:

Q. Li[1], S. Luo[1], J. Lin[2], A. Thurmon[1], K. Schmitz-Abe[1], P. Agrawal[1]; [1]Univ. of Miami, Miami, FL, [2]Boston Children's Hosp., Boston, MA

Abstract:

Striated preferentially expressed protein kinase (SPEG), which belongs to the myosin light chain kinase protein family, is mutated in centronuclear myopathy (CNM) and/or dilated cardiomyopathy. Currently, specific therapeutic options for this disorder are lacking, and the feasibility of gene replacement therapy is limited due to the large size of SPEG protein. Dynamin-2 (DNM2) is a member of the large GTPase family, and reducing its levels has been shown to revert muscle phenotypes in mouse models of CNM due to *MTM1*, *DNM2* and *BIN1* mutations. We previously found that a reduction in DNM2 in SPEG-deficient mice improves skeletal muscle function without benefiting cardiac muscle. Nevertheless, the molecular mechanisms underlying the amelioration of SPEG-associated myopathy by DNM2 reduction remain elusive. Multi-omics analyses serve as robust tools for elucidating intricate biological processes, disease progression, and therapeutic efficacy. Here, we employed multi-omics techniques, including proteomic, phosphoproteomic, and transcriptomic analyses, to elucidate the molecular mechanisms by which DNM2 reduction mitigates the myopathy phenotype in SPEG-related myopathy and to delineate the implicated proteins and pathways. This work highlights the advantages of using multi-omics techniques to conduct a comprehensive analysis of both pathological and therapeutic changes in rare diseases.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8083W: Engineering poison exons for allele-specific silencing of the Huntington's disease gene *HTT*.

Authors:

C. Catamura, L. Lareau, F. Urnov, V. Salinas-Rios; Univ. of California Berkeley, Berkeley, CA

Abstract:

Repeat expansion diseases that follow a dominant negative inheritance pattern, such as Huntington's disease, present a great challenge for gene therapies. The most appealing therapeutic strategy requires allele-specific silencing of the disease allele while leaving the wildtype allele intact. Because some amount of the repeat sequence is present in both alleles, selective CRISPR targeting of the mutant allele is profoundly challenging.

Here, we present a general framework for allele-specific suppression of repeat expansion genes and show its efficacy in suppressing the gene responsible for Huntington's disease, *HTT*. Our solution is to engineer a poison exon into only the disease allele: an alternative exon that, when included in an mRNA, will introduce an early stop codon and trigger nonsense-mediated mRNA decay (NMD). mRNA expressed from the edited disease allele would be degraded, while mRNA from the wildtype allele would be unchanged.

To target only the disease allele, we take advantage of benign, heterozygous SNPs within introns of the disease gene. CRISPR guides specific to the SNP allele that is in phase with the pathogenic repeat could recruit an RNA-guided deaminase base editor to introduce nucleotide substitutions within a narrow window on only the disease allele.

Using a deep learning model of splicing trained on natural and aberrant mRNA sequences, we screen in silico for positions where targeting a base editor could create sequences that form new splice sites. Importantly, the novel exon need not be near the actual disease mutation; any highly heterozygous SNP in an intron could potentially serve this purpose.

We applied this model to screen SNPs within introns of *HTT* and found positions where editing adjacent to the SNP was predicted to give rise to new poison exons. Evaluating our pipeline on 1000 Genomes SNPs, we found that 62% of individuals are potentially viable for this strategy.

We validated these targets at the endogenous *HTT* locus in human cell culture, showing that the edits created new exons spliced into *HTT* mRNAs with over 75% inclusion. The resulting mRNAs were degraded by NMD, substantially reducing *HTT* expression.

We conclude that editing of just a few carefully chosen nucleotides in a gene can be sufficient to create a new exon, making it a feasible approach for allele-specific mRNA depletion. Our

approach combines the power of CRISPR editing with deep learning models of RNA sequence to engineer poison exons with therapeutic potential for a disease that, at present, is untreatable.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8084W: Identifying and therapeutically targeting novel splice defects in Glycogen Storage Disease IX

Authors:

A. Iyengar; Duke Univ., Durham, NC

Abstract:

Delayed genetic diagnosis of Mendelian diseases substantially prevents early therapeutic intervention that could improve prognosis. Here, we discovered a novel splice variant contributing to Glycogen Storage Disease (GSD) in two siblings with uncertain genetic etiology. We then generated patient-specific models of their disease, in which we demonstrated that splice switching oligonucleotides (SSOs) can correct splicing. The siblings presented with glycogen phosphorylase kinase (PhK) deficiency (<20%). Both carried a single likely pathogenic variant at *PHKG2* c.96-11G>A. Presumed to be compound heterozygous with an unknown pathogenic variant on the 2nd allele of *PHKG2*, they were diagnosed with the γ2 subtype of GSD IX, the most severe form of the disease. Despite careful dietary control of blood glucose as evidenced by decreasing blood lipids and normal urinary Glc4, both siblings have rapidly progressing liver fibrosis. This progression could eventually cause liver failure and necessitate a transplant; thus, a timely therapeutic option. To complete their genetic diagnosis, we discovered a novel non-coding splice variant in both siblings via whole genome sequencing (WGS) and RNA-seq. WGS identified *PHKG2* c.556+1069T>G (MAF=1:125,000) as a potential splice variant in intron 6, and RNA-seq confirmed it as the splice donor site of a 76-bp pseudoexon, causing frameshift in the kinase domain of *PHKG2*. To confirm pathogenicity, we created a homozygous mutant (G/G) HEK293T cell line model using CRISPR/Cas9 genome editing. The mutant cells expressed the pseudoexon-containing transcript, had 39% *PHKG2* expression, and 10% PhK enzyme activity compared to wild-type (T/T) cells, which expressed only the canonical transcript. This demonstrates that c.556+1069T>G and the pseudoexon it induces are almost certainly pathogenic in these siblings. We hypothesized that splice-switching oligonucleotides (SSOs) could be used to block splicing machinery at the pseudoexon site, leading to canonical splicing of *PHKG2*. We therefore selected this as a candidate therapeutic approach for GSD IX γ2, designing three 24-25bp SSOs that are antisense to the pseudoexon's splice donor, acceptor, and a predicted exonic splice enhancer. When

transfected into mutant (G/G) HEK293T cells, the SSOs improved abundance of canonical isoform by up to >80%. We have generated induced pluripotent stem cells from one GSD IX sibling and will further validate efficacy of these SSOs to rescue hepatocyte function. In concert with diagnostic RNA-seq, this provides a framework for identifying pathogenic variants that can be targeted by personalized RNA therapeutics.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8085W: Individualized ASO Therapy Development for a Patient with *BIN1*-Associated Autosomal Dominant Centronuclear Myopathy

Authors:

E. Frazier[1,2], R. S. Park[3,4], T. L. Schwab[3,5], K. J. Wierenga[6], E. W. Klee[3,7,8], **M. A. Cousin**[3,4]; [1]Univ. of Minnesota Rochester, Rochester, MN, [2]Mayo Clinic Graduate Sch. of BioMed. Sci., Rochester, MN, [3]Ctr. for Individualized Med., Mayo Clinic, Rochester, MN, [4]Dept. of Molecular Med., Mayo Clinic, Rochester, MN, [5]Dept. of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, [6]Dept. of Clinical Genomics, Mayo Clinic, Jacksonville, FL, [7]Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, [8]Dept. of Clinical Genomics, Mayo Clinic, Rochester, MN

Abstract:

Rare disease affects approximately 30 million people in the United States. Specific treatments are available for <5% of cases and curative for <1%. Antisense oligonucleotides (ASO) therapies can be developed to treat patients with targetable, disease-causing genetic alterations. This study evaluates the pathogenicity and ASO-based therapeutic potential of a novel variant in *BIN1* associated with autosomal dominant centronuclear myopathy (AD-CNM) characterized by progressive lower limb muscle weakness, waddling gate, and centralized nuclei and fiber size variability on muscle biopsy. An adult female patient was seen at Mayo Clinic in pursuit of the etiology of her progressive lower limb muscle weakness. Exome sequencing revealed a novel heterozygous variant of uncertain significance (VUS) in Bridging Integrator 1: NM_139343.3(*BIN1*): c.1765_1766dupAC; p.(Glu590Leufs*58). Biallelic pathogenic variants in *BIN1* are associated with autosomal recessive CNM and more recently heterozygous variants have been associated with AD-CNM. A muscle biopsy revealed active and chronic myopathy with inflammatory reaction, type 1 fiber atrophy, internal nuclei, and fiber variability consistent with published patients. This VUS and three distinct but similar published variants lead to an extension of the encoded amphiphysin 2 protein by ~52 supernumerary amino acids in the same shifted reading frame. We hypothesize a resultant a gain-of-function or dominant negative effect explaining the AD inheritance. We conducted immunofluorescent (IF) staining and

confocal microscopy on normal fibroblasts to define differences in protein abundance, localization, tubulation, and transverse tubule biogenesis between p.(Glu590Leufs*58), a known autosomal recessive CNM variant (p.(Asp151Asn), wild type *BIN1,* and a negative control using transient expression. We also evaluated allele-specific gapmer ASOs targeting our patient's *BIN1* variant to knock down the expression of this allele. We demonstrated successful *BIN1* knockdown with variable allele specificity in patient vs control fibroblasts using qPCR and PCR with Sanger sequencing followed by Tracking of Indels by Decomposition (TIDE analysis). We further optimized our ASOs for increased allele-specific targeting.
ASOs have advantages for rapid translation for amenable ultrarare patients and previous studies have demonstrated successful individualized development. This study describes a novel variant associated with a rare form of AD-CNM and the cellular effects of the variant. Further, we demonstrate gapmer ASOs can successfully knockdown the variant allele suggesting feasibility of this strategy.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8086W: Integrating Genetic and Proteomic Evidence to Predict Drug Approval Success Rates: Insights from the UK Biobank

Authors:

G. McInnes, M. Luth, E. Bartell, D. Jakubosky, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Abstract:

Despite significant technological and methodological advances in drug discovery and development over the past two decades, approximately 90% of investigational therapies that enter clinical development fail, most commonly due to a lack of efficacy. Previous studies have shown that drug targets supported by human genetics evidence are more likely to be approved. These studies, together with the increasing availability of large-scale biobanks comprising genetic and other 'omic' data linked to rich phenotypic information, such as the UK Biobank, have led to the relatively widespread adoption of human genetics by pharmaceutical and biotechnology companies in their drug discovery and development efforts. Therefore, with the emergence of high-throughput proteomics platforms and the availability of population-scale proteomics data in the UK Biobank cohort, understanding the impact of proteomic evidence on the likelihood of clinical development success for a drug is of increasing interest. Circulating levels of a target protein that are positively correlated with a disease may either imply that inhibiting the target could treat the disease or that the disease state leads to elevated levels of

the target protein (i.e. a biomarker). Therefore, establishing whether protein levels are causal or correlated with disease is challenging from protein levels alone, and integration of both proteomic and genetic data may be of greater utility. In this study, we focus on evaluating the contribution of proteomic data to clinical development success, comparing and contrasting proteomic evidence with genetic evidence and assessing the combined effect of both data types. Using genetic and proteomic data from the UK Biobank, we generated associations between nonsynonymous variants in drug target genes, drug target protein levels, and thousands of traits representing the indications of approved drugs, focusing on drug-target indication pairs derived from the OpenTargets database. Our findings reveal a zzincrease in the likelihood of drug approval for target-indication pairs supported by evidence from proteomics, and even greater likelihood of approval for target-indication pairs supported by both genetic and proteomic data. Overall, this study underscores the value of integrating genetic and proteomic data when conducting analyses in support of drug discovery and development. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8087W: L-Citrulline Treatment of Nitric Oxide Deficiency in MELAS: A Phase I Dose-Finding and Safety Study

Authors:

M. Almannai[1], J. Duong[2], A. El-Hattab[3], M. Ali[4], D. Bauri[4], C. Soler-Alfonso[4], C. Gijavanekar[4], S. Elsea[4], S. Lee[2], F. Scaglia[4]; [1]Genetics and Precision Med. Dept. (GPM), King Abdullah Specialized Children's Hosp. (KASCH), King Abdulaziz Med. City, Ministry of Natl. Guard Hlth.Affairs (MNG-HA), Riyadh, Saudi Arabia, [2]Columbia Univ. Mailman Sch. of Publ. Hlth., New York, NY, [3]Coll. of Med., Univ. of Sharjah, Sharjah, United Arab Emirates, [4]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX

Abstract:

Introduction: There is growing evidence that nitric oxide (NO) deficiency occurs in MELAS and results in impaired blood perfusion that can contribute to complications including stroke-like episodes (SLEs). Impaired NO production with low arginine and citrulline has been demonstrated in subjects with MELAS and these were restored by arginine and citrulline supplementations, with citrulline being superior in restoring these deficiencies **Aim and methods:** The primary aim of this phase I dose-finding and safety study was to estimate the maximum tolerated dose (MTD) of citrulline in adults with MELAS, using Time-To-Event (TITE) Continual Reassessment Method (CRM) design. The MTD was defined as the dose associated

with a target probability of dose limiting toxicity (DLT) of 0.25. The primary outcome was DLT defined as Orthostatic hypotension, Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03 adverse events at grade 3 or higher in the first 8 weeks after treatment initiation, or an increase from grade 0 to 2. The original sample size was 24 patients which was amended to possibly allow for early stopping if 90% interval estimate for the highest dose falls below the target probability of DLT. The secondary aims were to study changes in cerebral blood flow (CBF) using arterial spin labeling-magnetic resonance imaging, citrulline and arginine levels from baseline to end of treatment (week 4), and plasma guanidino compounds to assess for potential arginine toxicity. Plasma lactate and alanine were evaluated as pharmacodynamic biomarkers**Results:** 5 males and 5 females (age range: 19-57 years) were enrolled and completed the eight weeks of the study. No DLTs were encountered. The estimated MTD is 40 grams of citrulline daily. Arginine and citrulline levels(µmol/L) were higher at 4 weeks: mean arginine level at baseline 70 (sd 24), at 4 weeks 258 (sd 143); mean citrulline level at baseline 35 (sd 11), at 4 weeks 268 (sd 310). CBF increased in 9 out 10 subjects (mean CBF (ml/100 g/min) at baseline: 40 (sd 11), at 4 weeks 46 (sd 13)). In the occipital region, which tends to be affected more in SLEs, the mean CBF increased from 39 (sd 11) at baseline to 48 (sd 15) at 4 weeks. There was decrease in lactate (mmol/L) from 4.7 (sd 2) to 3.3 (7 subjects, sd 1.3) and in alanine (µmol/L) from 634 (sd 132) to 458 (sd 163) from baseline to 4 weeks. The study was stopped after 10 patients were recruited due to recruitment difficulties and funding**Conclusion:** A dose of 40 g citrulline was well tolerated. Subjects on citrulline were observed to have improved CBF. A randomized placebo controlled trial using the estimated MTD of citrulline will be required to further evaluate its efficacy in subjects with MELAS

Session Title: Genetic Therapies Poster Wednesday Session

Board 8088W: Long-read motor neuron RNAseq of deep intron 8 *IGHMBP2* variants reveals pathogenicity and informs a precision therapeutic strategy

Authors:

S. Silverstein[1], S. Donkervoort[1], T. Cassini[2], V. Bolduc[1], S. Smieszek[3], B. Przychodzen[3], F. Vertrini[4,5], E. Conboy[4,5], A. Comer[6], K. Treat[4,5], K. Liaqat[4,5], L. Mantcheva[4], K-y. Park[7], D. Bharucha-Goebel[1,8], R. Foley[1], K. Chao[9], S. B. Neuhaus[1], D. Adams[10], C. Grunseich[11], C. Bonnemann[1]; [1]Neuromuscular and Neurogenetics Disorders of Childhood Section/NINDS/NIH, Bethesda, MD, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanda Pharmaceuticals, Washington DC, DC, [4]Undiagnosed Rare Disease Clinic, Indiana Univ. Sch. of Med., Indianapolis, IN, [5]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med., Indianapolis, IN, [6]Dept. of Neurology,

Indiana Univ. Sch. of Med., Indianapolis, IN, [7]Stem Cell Unit/NINDS/NIH, Bethesda, MD, [8]Childrens Natl. Hosp., Div. of Neurology, Washington DC, DC, [9]Broad Inst. of MIT and Harvard, Cambridge, MA, [10]NHGRI/NIH, Bethesda, MD, [11]NINDS/NIH, Bethesda, MD

Abstract:

Biallelic variants in *IGHMBP2* are a known cause of spinal muscular atrophy with respiratory distress type I (SMARD1) and Charcot Marie Tooth type 2S (CMT2S). Here we present a series of four patients with high clinical suspicion for SMARD1/CMT2S-related disease and variants of uncertain significance in intron 8 of *IGHMBP2* in compound heterozygosity with pathogenic coding variants: c.1235 +1076G>A/c.1241_1254dup14;p.Ser419ArgfsX9, c.1235 + 894 C>A/ c.1730T>C;p.Leu577Pro, c.1235 +450 G>A/c.455T>C;p.Leu152Pro and c.1235 + 450 G>A/c.660A>T;p.Lys220Asn.
To determine pathogenicity of these deep intronic variants, patient-derived iPSCs were generated and further differentiated to motor neurons to obtain short and long-read RNA sequencing. Subsequent analysis of long reads identified multiple novel pseudoexon inclusions at the site of each deep intronic variant. Of note, in the short-read sequencing data these inclusions were not clearly evident. Each pseudoexon inclusion results in a frameshift with a premature truncating codon. While each variant utilizes a unique splice acceptor site, all make use of the same intronic cryptic donor site. Thus, this apparent hotspot of aberrant splice induction provides an excellent opportunity for a common antisense oligonucleotide (ASO) mediated splice modulation approach via targeting of the shared activated cryptic donor site. We thus developed a single ASO that corrected aberrant splicing caused by both c.1235 + 894C>A as well as by c.1235 +1076G>A in patient-derived motor neurons as measured by a reduction in pseudoexon inclusions in treated cells compared with a non-targeted control. Work to determine the ASO's efficacy in the c.1235 + 450G>A variant is ongoing. Restoration of *IGHMBP2* protein will be assessed in the treated motor neurons. This series highlights the importance of full characterization of deep intronic variants which may induce aberrant splicing and therefore constitute targets for therapeutic splice modulation.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8089W: Modulating 5'UTR splicing to treat haploinsufficient neurodevelopmental disorders

Authors:

E. Wells[1], L. De Conti[2], R. Dawes[1], Y. Chen[1], H. Kim[1], M. Baralle[2], S. Sanders[1], N. Whiffin[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Intl. Ctr. for Genetic Engineering and Biotechnology, Trieste, Italy

Abstract:

Background: Rare disorders collectively affect over 300 million people worldwide, but over 95% have no effective treatments. Recent advances in oligonucleotide and CRISPR editing technologies are driving therapeutic development. It is relatively straightforward to downregulate genes therapeutically, for example with gapmer antisense oligonucelotides (ASOs). However, while haploinsufficiency is a major cause of rare disorders, upregulating protein levels is considerably more difficult.
We explore a novel approach to upregulation: modulating splicing of 5' untranslated regions (5'UTRs) to increase protein translation. Specifically, we aim to selectively exclude 5'UTR exons that contain upstream open reading frames (uORFs), which are negative regulators of translation.
Methods: We find 2,210 'skippable' exons (i.e. flanked by other 5'UTR encoding exons) in 1,573 MANE transcripts. Using ribosome profiling data from 13 human tissues and cell lines, including brain, heart and skeletal muscle, we identify start sites of translated uORFs in 1,083 of these skippable exons. 91 of these uORF-containing exons are in the 5'UTRs of 66 neurodevelopmental disorder genes with a dominant loss-of-function disease mechanism (from G2P, ClinVar, GenCC, and/or PanelApp). We prioritised target exons with a CRISPR-editable splice-acceptor site or that could be targeted by a splice-switching ASO for functional validation. We used dual luciferase assays to measure the impact of removing the target exon on downstream translation. For each gene we used three constructs: (1) wildtype 5'UTR, (2) full length 5'UTR with a single base change to remove the uORF start site, and (3) the 5'UTR without the target exon containing the uORF start site.
Results: As an example, for *GRIN2B*, there was a statistically significant, 7 to 8-fold increase in translation in both constructs without the uORF, compared with the wildtype construct. The difference between the two constructs without the uORF was not significant, suggesting the upregulation was specifically due to the loss of the uORF. For some other genes, (e.g. *TBL1XR1*), removing the target 5'UTR exon did not significantly alter luciferase translation, highlighting more complex UTR-mediated regulation. We are now testing computationally designed splice-switching ASOs for successfully validated targets (including *GRIN2B*).
Conclusion: Upregulation of translation through modifying 5'UTR splicing is a promising therapeutic approach for a subset of haploinsufficient disorders. Our results inform our predictions of gene targets which are amenable to this therapeutic mechanism.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8090W: Pathohistological study of intracranial Gaucheroma causing deafness in a patient with Gaucher Disease Type 3: Effects of substrate reduction therapy

Authors:

S. Yano[1], R. McGowan[1], M. Warren[2]; [1]LA Gen. Med. Ctr., Los Angeles, CA, [2]Children's Hosp. Los Angeles, Univ. of Southern California, Los Angeles, California, Los Angeles, CA

Abstract:

Backgrounds*:** Gaucher disease (GD) is an autosomal recessive disorder due to mutations in *GBA1* encoding lysosomal glucocerebrosidase. Macrophages with excess glucocerebroside, known as Gaucher cells (GC), infiltrate various organs, leading to multisystemic abnormalities. Mixed hearing loss is reported in patients with GD: conductive impairment due to GC infiltration to the middle ear and the mastoid and sensorineural impairment due to inner hair cell dysfunction caused by accumulation of glucocerebroside. ***Case Study: A 19-year-old female with type 3 GD diagnosed at birth based on the *GBA1* genotype (homozygous L444P) was treated using enzyme replacement therapy (ERT) with imiglucerase (120 IU/kg/2 weeks) from age 2 months. She developed profound mixed hearing loss at age 4 y. Surgical procedure to place a cochlear implant at age 16 y was aborted when a mass of abnormal tissue extending into the middle ear space and mastoid was found, later histopathologically diagnosed with GC. After 51 months of substrate reduction therapy (SRT) with eliglustat in addition to ERT, a cochlear implant was placed. Light and electron microscopic findings were compared between the specimens from the pre- and the post-SRT. ***Results***: The mean size of GC significantly decreased after SRT (pre: 25.4 $\pm$ 7.1, post: 15.1 $\pm$ 4.0 mm, $p < 0.01$, n=500). Electron micrographs showed the normalized size of lysosomes without microtubular structures in the post-SRT specimen. ***Conclusions***: GC infiltration in the tissues involved in hearing has been reported in patients with GD who developed deafness. Although ERT with imiglucerase alone could not prevent intracranial Gaucheroma, supplementation of SRT with eliglustat decreased the size of GC (Gaucheroma). Early implementation of SRT with eliglustat may prevent at least conductive hearing impairment in GD although it may not prevent inner hair cell dysfunction.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8091W: Penetrance of Alpha-1 Antitrypsin Deficiency among PiZZ Individuals in UK Biobank

Authors:

M. Parker[1], T. Zou[2], L. Barrera[1]; [1]Beam Therapeutics, Cambridge, MA, [2]Beam Therapeutics, Brookline, MA

Abstract:

Objective: To examine the penetrance and effect size of the *SERPINA1* PiZZ genotype with alpha1 antitrypsin deficiency (AATD) and related phenotypes in UK Biobank. **Methods:** UK Biobank is a prospective cohort study of 500k individuals with more than 15 years of electronic health record (EHR) follow-up. We assessed the penetrance of AATD, chronic obstructive pulmonary disease (COPD), liver cirrhosis, and death among PiZZ individuals in the British ancestry subset of UK Biobank (n=337k). AATD and liver cirrhosis diagnosis was ascertained though the ICD10 codes E88.0 (AATD) and K74.6 (cirrhosis). COPD was algorithmically defined by UK Biobank though a combination of ICD9 codes, ICD10 codes, and self-reported COPD and emphysema diagnoses. Death data was ascertained through the UK National Death Registry. **Results:** A total of 128 unrelated PiZZ individuals were identified in UK Biobank (mean age at enrollment = 57 years, 53% male, 56% ever smokers). Penetrance of AATD diagnosis was 28.1% (n= 36/128). As expected, Cox proportional hazard regression controlling for age, ever smoking status, and genetic ancestry revealed PiZZ were significantly more likely to be diagnosed with AATD than controls (HR = 123.6, p-value < $2x10^{-16}$). By age 60, the cumulative incidence of AATD diagnosis among PiZZ was 15.9% (95% CI = 9.2%-22.1%). Penetrance of COPD diagnosis was 28.9% (n=37/128). On average, COPD was diagnosed 6 years before AATD. PiZZ were significantly more likely to be diagnosed with COPD than controls (HR = 10.6, p-value < $2x10^{-16}$). PiZZ were also significantly more likely to be diagnosed with liver cirrhosis than controls (HR = 15.9, p-value < $2x10^{-16}$), however only 7.0% (9/128) were diagnosed. Survival analysis revealed that PiZZ individuals had a reduced survival (HR = 3.0, p-value = $4.6x10^{-9}$). This association with death remained significant in never smokers (HR=2.9, p-value = $8.3x10^{-4}$). The leading causes of death among PiZZ individuals were AATD, respiratory disease, and cancer. **Conclusions:** PiZZ was associated with AATD, COPD and liver cirrhosis in UK Biobank, with 28.1% of PiZZ's diagnosed during follow up. PiZZ was significantly associated with time to death in both smokers and nonsmokers.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8092W: Plant Exosome-like Nanovesicles for Delivery of Gene Editing Tool (Antisense Oligonucleotide) to Treat Cell Model of Amyotrophic Lateral Sclerosis

Authors:

E. Sterling, R. Hankin, Y. Zheng, L. Yang, K. Ye, Y. Yao; Univ. of Georgia, Athens, GA

Abstract:

Background: Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder with no known cure. Current treatments are hampered by their ability to target disease relevant cells and tissue. Limitations include degradation of therapeutic agent, difficulty bypassing biological barriers, and inability to selectively target regions of interest. One potential solution is nanoparticles isolated from plants known as plant exosome-like nanovesicles (PELNVs). It is not yet established though if PELNVs can be modified to carry gene editing tools, such as antisense oligonucleotide (ASO) for therapeutic applications in ALS. We hypothesize that ASO-loaded PELNVs will be able to protect ASO until delivery to targeted cells in a model of ALS. **Methods:** The goal of this project is to test the ability for PELNVs to perform as a drug delivery shuttle (DDS). First, we isolate PELNVs from the plant model organism *Arabidopsis Thaliana* (*A. Thaliana*) using differential ultracentrifugation and ultrafiltration. Next, we characterize the nanoparticles utilizing nano flow cytometry to identify size and concentration. Then, we determine the loading efficiency of ASO by detection of a fluorescence tag. Final steps include the delivery and phenotypic impact of ASO-loaded PELNVs, determined by expression of misfolded SOD1 (superoxide dismutase 1) protein as a functional readout of efficacy in an ALS treatment paradigm. **Results:** We successfully isolated and characterized PELNVs to have a mean size within 50-150 nm and a concentration of approximately 9e10 per 1g of leaf tissue. We detected a fluorescence signal of PELNVs loaded with ASO compared to negative controls. Furthermore, we determined a significantly increased fluorescence signal of dyed PELNVs in cells treated compared to negative controls. **Conclusions:** The results from this study validated PELNVs as a potential DDS. Further optimization is required to better target select organs, such as the brain for ALS clinical usage. In addition, this is an early preclinical study and thus further safety studies are required.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8093W: Preclinical development of a gene therapy for Ogden syndrome, using new mouse models of the disease ⭐

Authors:

G. Lyon[1], S. Kneeland[2], K. Snow[3], H. Kang[4], S. Gray[5], R. Patel[1], R. Makwana[1], E. Marchi[6], J. Wesely[7], Y-R. Chen[7], C. Lutz[8], S. Murray[8]; [1]Inst. for Basic Res. in Dev.al Disabilities, Staten Island, NY, [2]Jackson Lab., Bar Harbor, Maine, USA., ME, [3]The Jackson Lab., Ellsworth, ME, [4]UT-Southwestern, Dallas, Texas, USA, TX, [5]UT-Southwestern, Dallas, TX, [6]Inst. for Basic Res., Staten Island, NY, [7]The New York Stem Cell Fndn. Res. Inst., New York City, NY, [8]The Jackson Lab., Bar Harbor, ME

Abstract:

Amino-terminal (Nt-) acetylation (NTA) is a common protein modification, affecting 80% of cytosolic proteins in humans. The human essential gene, *NAA10,* encodes the enzyme NAA10, as the catalytic subunit for the N-terminal acetyltransferase A (NatA) complex. The first human disease involving *NAA10* was discovered in 2011 (named Ogden syndrome, OS), and since that time, >20 distinct variants have been found in *NAA10* in >100 families. OS has many (variable) clinical phenotypes, including severe intellectual disability, autism, arrhythmias, cortical visual impairment, and even rarely microphthalmia. Infantile lethality or sudden death in older years occurs in ~5% of the probands. The human phenotypes, mouse models, and data from cell lines favor a mechanism involving reduced NTA of key proteins varying in parallel with the phenotypic spectrum. We hypothesize that OS can be treated by supplying the 235 amino acid enzyme with an adeno-associated vector (AAV9). We designed and produced AAV9-NAA10 and confirmed NAA10 expression from the plasmid used to make this vector. We assessed maximum tolerability of AAV9-NAA10 overexpression via intracerebroventricular (ICV) injection in newborn C57Bl/6J mice (day P0). Body weight and overall health were measured weekly. A subset of mice were evaluated at 4 weeks of age and confirm >4 fold vector expression in brain and liver with no adverse effects noted through 3 months of age. The mice were also monitored for any signs of tumors and assessed at 6 months of age for transduction efficiency and expression. We have also created an improved mouse model of OS with one of the most prevalent mutations (Arg83Cys). This R83C strain has a similar phenotype (albeit less penetrant) to the *Naa10* knockout, thus favoring the mechanism of reduced NTA of key proteins. We previously showed that mice have a compensating enzyme Naa12 that is not present in humans, thus enabling the *Naa10* knockout and *Naa10* R83C mice to survive embryogenesis. Nonetheless, extensive cognitive testing of male *Naa10* knockout mice revealed profound learning and memory deficits. We bred the *Naa12* knockout mice to the *Naa10* knockout and *Naa10* R83C strains, and the *Naa10*$^{Arg83Cys/y}$*Naa12*$^{+/-}$ double mutant males are born, but at a

much lower rate and exhibit reduced size, small head, upturned nose, small testes, and reduced adipose tissue. Heterozygous *Naa10*$^{Arg83Cys/+}$*Naa12*$^{+/-}$ and double mutant *Naa10*$^{Arg83Cys/+}$*Naa12*$^{-/-}$ females both survive to birth but have milder phenotypes than males. Overall, we have developed an improved mouse model for OS and made substantial progress toward preclinical development of a gene therapy for OS.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8094W: Purity and stability assessment of AAT protein to support Beam 302 clinical bioanalysis

Authors:

X. Chen, K. Cao, L. Matt, M. Ortega, B. Yan; Beam Therapeutics, Cambridge, MA

Abstract:

Alpha-1 antitrypsin deficiency (AATD) is an inherited autosomal co-dominant disorder caused by mutations in the SERPINA1 gene that encodes alpha-1 antitrypsin (AAT). The PiZ variant (p.Glu366Lys), resulting from a G to A polymorphism, is the most frequent severe variant associated with lung and liver disease. The PiZ mutation is an ideal target for correction to wildtype (PiM) by an adenine base editor (ABE), which converts an A to G in genomic DNA.BEAM-302 is a lipid nanoparticle (LNP) formulation of an mRNA encoding an ABE and a guide RNA (gRNA) targeting correction of the PiZ mutation. Treatment with BEAM-302 can result in corrected alleles, alleles with editing of both the target nucleotide alongside a neighboring bystander adenine, or rare alleles with the bystander edit alone. These genomic editing outcomes result in the expression of wildtype PiM-AAT, PiM+bystander, and PiZ+bystander AAT proteins. To understand the pharmacodynamic response of BEAM-302 and identify an optimal biological dose for patients, it is essential to accurately quantify these four proteins (the three described above plus PiZ, which results from no editing) in patient samples. The establishment of protein-based LC-MS methods for AAT quantitation in support of BEAM-302 nonclinical and clinical sample analyses require reference AAT proteins. However, PiM+bystander, PiZ, and PiZ+bystander proteins are not commercially available.PiM, PiZ, PiM+bystander, and PiZ+bystander proteins were recombinantly expressed and purified from E. coli. During purification, PiZ and PiZ+bystander were prone to aggregation, requiring a high concentration of detergent in the storage buffer to maintain structural integrity. These detergents interfere with various analytical assays, including UV280 nm absorption, size-exclusion chromatography (SEC), and Bradford assays, posing a challenge for accurate reference protein concentration determination. To address these challenges, we developed an RP-HPLC-

MS quantitative method to precisely determine protein purity and detect impurities in these recombinant AAT proteins. Mass spectrometry data showed a significant increase in dimer formation even when the proteins were only stored for a short period at 4°C and -70°C. To prevent dimer formation and maintain long-term recombinant PiZ and PiZ+bystander protein stability, new formulation buffers were developed to preserve the integrity of these proteins for long-term storage. Taken together, our fit-for-purpose detergent-compatible RP-HPLC-MS quantitative method will support critical BEAM-302 development activities.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8096W: Response to asfotase alfa with no mutation in *ALPL*

Authors:

C. Brownstein[1], G. del Angel[2], P. Karachunski[3], R. Pierpont[4], N. Edisis[5], D. Miller[6], K. Cabral[1], A. Petryk[7], K. Sarafoglou[8]; [1]Boston Children's Hosp., Boston, MA, [2]Alexion Pharmaceuticals Inc., Lexington, MA, [3]The Univ. of Minnesota and MDA Care center Div. of Pediatric Neurology, Dept. of Neurology, Minneapolis, MN, [4]Div. of Clinical Behavioral NeuroSci., Univ. of Minnesota Med. Sch., Minneapolis, MN, [5]Boston Children`s Hosp., Boston, MA, [6]M Hlth.Fairview, Minneapolis, MN, [7]Alexion Pharmaceuticals, Boston, MA, [8]Div.s of Pediatric Endocrinology and Genetics & Metabolism, Univ. of Minnesota Med. Sch., Minneapolis, MN

Abstract:

Introduction and Case: A 5.8 years old non-ambulatory, non-verbal female with history of ataxia, peripheral neuropathy and profound muscle weakness of unknown etiology since infancy was seen at the endocrine clinic for evaluation of her bone health. Prior evaluation included a muscle biopsy that revealed a chronic and ongoing myopathy and features of fiber type disproportion. An EMG study showed a possible mild abnormality of neuromuscular transmission. MRI of the brain revealed volume loss/hypoplasia involving the inferior portion of cerebellar vermis with mild bilateral cerebellar volume loss. Whole exome sequencing, mtDNA analysis and coenzyme Q10 deficiency panel did not reveal the cause of her muscle weakness. **Endocrine evaluation**: Initial evaluation showed low alkaline phosphatase (ALP) level at 149 U/L (150-420), elevated pyridoxal 5'-phosphate (PLP) at 172.4 nmol/L (20-125) and elevated urinary phosphoethanolamine level at 190 mg/dl (0-106) consistent with hypophosphatasia (HPP). Skeletal imaging showed no fractures, but presence of osteopenia, gracile bones, and abnormal tibial epiphyses. DXA scan showed concomitant osteoporosis. Bone mineral density (BMD) Z-scores of -4.6 and -3.3 for the lumbar spine and the total body less head (TBLH), respectively. She was started on subcutaneous asfotase alfa therapy 2 mg/kg/day

three times a week. **Molecular evaluation:** *ALPL* gene sequencing including deletion and duplication studies was negative. Whole genome sequencing (WGS), Bionano Optical Genome Mapping (OGM), and RNAseq were performed in an effort to identify the cause of her low ALP. Neither technique identified a variation in the *ALPL* gene. No other potentially causative genetic variants were identified in other genes. RNAseq analysis from blood was completed on all family members and showed normal levels of *ALPL* expression and no abnormal splicing events. **Discussion**: Asfotase alfa therapy led to notable improvements in her muscle strength, cognition, memory and language development. She has been able to stand 3-4x/week for 20-90 minutes and has been biking with her adaptive trike and walking with her gait trainer, daily. She has some bruising at the injection sites of the abdomen and lipoatrophy. Both BMD lumbar and TBLH Z-scores (-2.7) improved after 2.5 years of therapy. **Conclusions**: This patient either has a cryptic variant in *ALPL* that has eluded detection by WGS and OGM, or additional gene variants that account for low ALP and its consequences. Asfotase alfa therapy may have a role in improving muscle strength and cognition in patients with HPP even when genetic testing is negative for variation in *ALPL*, resulting in functional improvements.

Session Title: Genetic Therapies Poster Wednesday Session

Board 8097W: Stimulating And Simulating DNA For Auto-Correction To Enhance The Regeneration Of Lost Structures Or Functions In The Human Body

Authors:

N. Allu; Shubhakruth Med. Genetics Pvt. Ltd., Visakhapatnam, India

Abstract:

Focus on using the body's natural healing processes to repair, replace, or regenerate damaged or diseased tissues and organs.
One of the main approaches used in regenerative medicine is stem cell therapy, tissue engineering, and gene therapy, which involves modifying a patient's genes to treat or prevent diseases.
While it is true that homeostasis plays a crucial role in maintaining the overall balance of the body, it is important to note that stimulating and simulating DNA for auto-correction is not a well established concept in modern science. While some researchers are exploring the potential of gene therapy and other methods to correct genetic disorders and mutations, the idea that DNA can be stimulated or simulated for auto-correction is not yet supported by robust scientific evidence.

Furthermore, while the study of molecular geometry and its various expressions is important for understanding the function of molecules and their interactions in the body, it is not clear how this knowledge can be used to regenerate lost structures or functions in the human body. While there are some promising approaches to regenerative medicine, such as stem cell therapy and tissue engineering, much research is still needed to fully understand the complex mechanisms involved in tissue regeneration.
It is also important to approach the topic of molecular biology and human health with a grounded understanding of scientific principles and evidence-based practices. While there may be some speculative or untested ideas that could hold potential, it is important to evaluate them critically and prioritize approaches that are well-supported by rigorous scientific investigation.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1042W: A custom array for genotyping rare predicted loss-of-function variants in the Michigan Genomics Initiative biobank

Authors:

B. Vanderwerff[1], L. Fritsche[1], A. Scott[2], E. Bertucci-Richter[1], S. Patil[1], M. Boehnke[1], X. Zhou[1], S. Zollner[1]; [1]Dept. of Biostatistics, Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, [2]Div. of Genetic Med., Univ. of Michigan, Ann Arbor, MI

Abstract:

Background: Predicted loss-of-function (pLoF) variants can disrupt protein function, impact health, and are valuable for research in areas such as functional genomics and monogenic disease. Moreover, pLoF variants can inform clinical decision-making when present in genes associated with established interventions, such as those listed on the American College of Medical Genetics and Genomics (ACMG) list for secondary findings (SF) in clinical sequencing. However, assessing rare variants from genotyping arrays is challenging, even after imputation with large reference panels. In this study, we developed a customized Global Screening Array (GSA) as an effective option to study rare pLoF variants in the Michigan Genomics Initiative biobank, a genetic resource linked to electronic health record (EHR) data that has enrolled > 100,000 participants with recontact consent.
Methods: We selected custom variants from TOPMed WGS data (bravo release 5) with minor allele count (MAC) > 1, minor allele frequency (MAF) < 1%, and classified as pLoF by the annotation software LOFTEE. We genotyped 19,814 participants in a non-CLIA accredited

laboratory and applied improved QC protocols for accurate rare variant calling. We annotated variants for pathogenicity using ClinVar.
Results: Probes for 37,912 custom-designed variants were successfully synthesized on the array. 99.8% (19,780/19,814) of participants carried at least one of 17,219 pLoF variants with MAF < 1%, 61% (10,507/17,219) of which had MAC < 3. ~76% (13,137/17,219) of pLoF variants were not well imputed from TOPMed. > 7,500 genes had a pLoF variant in at least one participant, with most participants carrying pLoF variants in > 5 genes. Focusing on variants predicted to be pathogenic/likely pathogenic (P/LP) in ACMG SF list genes, 7.1% (1,407/19,814) of participants were heterozygous for at least one such variant. P/LP pLoF variants in *BRCA1/2* replicate a known association with breast cancer ($p < .05$, OR ~2.7) and are carried by 3.0% (316/10,493) of living females. 79% (251/316) of these carriers had no breast cancer diagnosis in their available EHR, representing potential candidates for recontact studies or health intervention after CLIA-certified validation.
Conclusions: This study establishes an effective custom array design workflow to evaluate rare, high-impact variants for large-scale genotyping efforts. The resulting data are a valuable resource for retrospective research on gene loss-of-function and its associations with EHR data. Furthermore, this approach enables the identification of individuals for follow-up studies, such as carriers of actionable variants in ACMG genes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1043W: A draft pangenome reference for people with Pacific ancestry

Authors:

C. Littlefield[1], J. M. Lazaro-Guevara[2], D. Stucki[1], M. Lansford[1], M. H. Pezzolesi[1], E. J. Taylor[1], J. Taloa[1], K. Lao[1], J. P. Tavana[3], W. L. Holland[1], K. L. Raphael[1], M. G. Pezzolesi[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]McGill Univ., montreal, QC, [3]Brigham Young Univ., Provo, UT

Abstract:

Individuals with Pacific ancestry have been underrepresented in human genomic research. Although recent efforts to enhance resources for analyzing genetic sequence datasets from diverse global populations have seen success, they have largely excluded people with Pacific ancestry. One of these resources is the pangenome reference, which is designed to reduce reference bias when mapping and aligning genomic sequence data. The most diverse draft pangenome reference to date was developed by the Human Pangenome Research Consortium

(HPRC) and includes assemblies from 47 genetically diverse individuals, yet it still lacks representation from individuals from the Pacific. To begin to address this gap, we created a draft reference pangenome using long-read whole genome sequencing data (generated using PacBio's HiFi long-read sequencing technology; average depth of 33X) from 23 individuals with Pacific ancestry, including individuals from Tonga (6), Samoa (6), Fiji (4), Philippines (2), Marshall Islands (2), Guam (1), Tahiti (1), and Pohnpei (1). We next generated 46 partially-phased haploid assemblies using Hifiasm and evaluated and corrected them with Inspector. These assemblies are highly contiguous, with an average N50 value of 40.7 Mb, and comparable to GRCh38 (N50 = 58 Mb). They are nearly complete, averaging 2.98 Gb in length (97.4% of T2T-CHM13), and highly accurate, with quality values (phred-scaled QV) ranging from 51 to 58 (average of 55). Mapping the assemblies to T2T-CHM13 revealed 20-40 Mb of novel unmapped sequences per assembly. We then used the minigraph-cactus pipeline to combine the 46 assemblies with GRCh38 and map these to T2T-CHM13 to create a draft Pacific ancestry pangenome reference. This pangenome includes a total of 3.38 Gb of sequence, of which 103 Mb is novel sequence derived from people of Pacific ancestry. When we combined these assemblies with the 47 HPRC assemblies to create a joint pangenome, we found that adding sequences derived from individuals of Pacific ancestry introduced an additional 30.6 Mb of novel sequence and 2.8 million variants unique to these assemblies. This new resource has the potential to significantly improve the mapping of genomic sequence data from individuals with Pacific ancestry and to serve as an accurate reference panel for structural variation specific to this population.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1044W: A Framework for Ensuring High Quality Genetics and Omics Training Data for Artificial Intelligence and Machine Learning

Authors:

A. Musolf[1], K. Shkura[2], J. Miller[1], B. Guo[3], N. Zhao[1], N. Raghavan[1], T. Ehrenberger[1], Y. Yan[1], J. Mueller[1], S. Longerich[4], M. Kennedy[1], A. Tamburino[1], R. Mathew[1]; [1]Merck & Co., Cambridge, MA, [2]MSD, London, United Kingdom, [3]Merck & Co., Inc., Quincy, MA, [4]Merck, Cambridge, MA

Abstract:

In the past few years, the use of artificial intelligence (AI) and machine learning (ML) has exploded dramatically, leading to increased implementation in many fields. This includes pharmaceuticals, where AI/ML has been deployed in target identification, drug development, and image analysis. To train AI/ML models requires high quality training data. While language-

based applications like ChatGPT use plentiful text-based training data, that is readily available and easy to interpret, the same is not true for pharmaceutical training data. Training AI/ML models on data like genetics and other omics requires expert curation. Here, we describe a framework for ensuring AI/ML training data is highest quality. Our framework involves the creation of a centralized repository or data lake for high quality training data. The lake itself is divided into units called domains, which are based on modalities such as genetics, single cell transcriptomics, etc. Domains are a mix of both publicly available data and proprietary internal data. The key to ensuring high quality data is expert curation and review by a domain expert (for external data) or the experimenter/creator (for internal data). Data can be uploaded and downloaded from the data lake; this process is tightly controlled. Uploads are automated, which structures all data into a uniform format, creates associated companion files and metadata, and places the data into the proper domain. Almost all data (with the exception of image data) is formatted into a matrix, with genes as the rows and other features (such as genotype, p-value, differential expression) as the columns. Ownership of the data is retained by the uploader for easy updates and correction. Download of the data is assisted by a user interface which allows users to view all modalities, data matrices, and headers within the data matrices. In this way, users can easily extract features from the data matrices for AI/ML analyses. Once features are selected, a new gene by feature matrix is created; this can be done across modalities. Overall, the process is straightforward and intuitive for the uploaders and downloaders. This structure has greatly facilitated creation of high-quality training data, ensuring that AI/ML methods show high performance. AI/ML results are themselves fed back into the data lake for discussion with domain experts and use for further AI/ML analyses. This type of data curation by domain experts will only increase in importance as AI/ML continues to grow in genetics and other fields.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1045W: A novel data model for population descriptors in genomic research of diverse populations developed in the Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium

Authors:

A. Khan[1], C. Adebamowo[2], S. Adebamowo[3], M. Conomos[4], B. Darst[5], S. Fullerton[4], J. Hirbo[6], P. Kraft[7], S. Nelson[8], S. Rich[9], Q. Sun[10], G. Wojcik[11], Y. Zhang[12], S. Gogarten[4]; [1]Pennsylvania State Univ., State College, PA, [2]Univ. of Maryland Sch. of Med., Ellicott City, MD, [3]Univ. of Maryland Sch. of Med., Baltimore, MD, [4]Univ. of Washington, Seattle, WA, [5]Fred Hutchinson Cancer Ctr., Seattle, WA, [6]Vanderbilt Univ. Sch. of Med., Nashville, TN, [7]Natl. Cancer Inst., Rockville,

MD, [8]Univ of Washington, Seattle, WA, [9]Univ. of Virginia, Charlottesville, VA, [10]UNC-Chapel Hill, Chapel Hill, NC, [11]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [12]Univ. of Maryland, Baltimore, MD

Abstract:

Population descriptors used in genetic studies have implications for translation to clinical risk models and communication of results to the public. There are no globally agreed-upon definitions or usages of common population descriptors (e.g., race, ethnicity, nationality, tribe), many of which are derived from political or bureaucratic conventions. Recent recommendations have encouraged retention of as much granularity in population descriptors as possible during data preparation, analysis, and interpretation. However, genomic research infrastructure currently does not accommodate a systematic and flexible organization, structure, and harmonization of multifaceted and detailed population descriptor data, leading to limitations of utility through loss of information, barriers to international collaboration, and potential issues in clinical translation. The NIH-funded Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium developed a data model that systematically organizes and retains detailed population descriptor data. It supports a versatile, traceable, and reproducible harmonization system through the mapping of original population data to new values for specific analyses rather than overwriting previous values. This novel system offers multiple benefits over existing data structures: it (1) follows recommendations from NASEM on distinguishing descriptors (axes of measurement) from labels (measured values); (2) allows retention of the original population descriptors and associated metadata for each participant; (3) encourages the mapping of terms when harmonizing population descriptor data rather than changing the participant data itself, as is common in the harmonization of other phenotypes; (4) facilitates the traceability and reproducibility of population descriptor harmonization; and (5) eases the process of updating population descriptor data (e.g., study participants with longitudinal data use different population labels over time). The data model affords researchers the flexibility to thoughtfully choose and scientifically justify their choice of population descriptors for specific analyses, helping to avoid the misrepresentation and misuse of genomics to justify power inequities and injustices, and the reification of harmful typological thinking. Genomic research infrastructure will need to adapt to address emerging best practices recommendations and produce scientifically robust findings that minimize misuse and maximize benefit.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1046W: A rare Chromosome 6p25.3p25.1 deletion and duplication at 2q34q37.3: A case report with dental and other clinical features.

Authors:

G. Jain; Univ. of Minnesota Sch. of Dentistry, Minneapolis, MN

Abstract:

This case report presents a very rare chromosome abnormality affecting 6p and 2q in the same patient with very complex medical conditions and dental findings. The chromosome 6p25.3p25.1 deletion and duplication of 2q34q37.3 are rare chromosomal abnormalities that can have various impacts on an individual 's health, including dental and systemic findings. The specific consequences of these chromosomal changes can vary significantly between individuals, depending on the size and location of the deletion and duplication. The patient presented to pediatric dentistry residency clinic for comprehensive care under general anesthesia after referral from medical team. This case report describes the radiographic and clinical findings, multidisciplinary treatment approach, and management of patient with complex disease progression and symptoms.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1048W: A unified computing environment for genomics data storage, management, and analysis: NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space (AnVIL)

Authors:

S. Mosher[1], M. C. Schatz[1], J. Lawson[2], R. Carroll[3]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Broad Inst., Cambridge, MA, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Recent years have seen astronomical growth in human genomics. Together with single-cell and functional genomics, electronic medical records and other biomedical data, the field is well-positioned to make great advances in human health. However, the complexity of genomic data sharing, where data is downloaded from centralized datastores for local analysis, is unsustainably cost prohibitive. Furthermore, housing genomic data across redundant

institutional compute infrastructures makes assuring data security and compliant usage of protected data a challenge. The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space, or AnVIL (https://anvilproject.org/) was developed to address these and other concerns by providing a unified cloud-based computing environment for genomics data storage, management and analysis. The AnVIL platform inverts the genomics data sharing model by eliminating the need for data movement, allowing for active threat detection and monitoring and provides scalable and sharable on demand compute resources. AnVIL currently provides access to more than 600,000 genomes from several key NHGRI projects, such as as the CCDG (Centers for Common Disease Genomics), CMG (Centers for Mendelian Genomics), eMERGE (Electronic Medical Records and Genomics), GTEx (Genotype-Tissue Expression Project), T2T (Telomere-to-Telomere), HPRC (Human Pangenome Reference Consortium) and many more.The platform is built on a set of established components that have been used in a number of flagship scientific projects. The Terra platform provides a compute environment with secure data and analysis sharing capabilities. Dockstore provides standards based sharing of containerized tools and workflows. Jupyter, R/Bioconductor and Galaxy provide analysis environments for users at all skill levels to interactively explore and understand data with thousands of tools available. The AnVIL Data Explorer provides data and metadata for querying and custom cohort creation. Together, AnVIL provides a collaborative environment for creating, analyzing, and sharing data and analysis workflows for even the largest projects.Today, AnVIL is migrating to a multi-clouded experience including both Google Cloud and Microsoft Azure. Together with improved data onboarding processes and expanded educational content, the AnVIL platform is well poised to drastically reduce the burden of performing large- scale integrated analyses across many datasets to fully realize the potential of ongoing data production efforts.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1049W: Assessing large language model performance related to aging in genetic conditions

Authors:

A. Othman, K. A. Flaharty, S. E. Ledgister Hanchard, P. Hu, D. Duong, R. L. Waikel, B. D. Solomon; NIH, Bethesda, MD

Abstract:

Unlike some health conditions that have been extensively delineated throughout the lifespan, many genetic conditions are largely described in pediatric populations, with a focus on early manifestations like congenital anomalies and developmental delay. An apparent gap exists in understanding clinical features and optimal management as patients age. Generative artificial intelligence is transforming biomedical disciplines including through the introduction of large language models (LLMs). Motivated by these advances, we explored how LLMs handle age with respect to 282 genetic conditions selected based on prevalence. We divided these conditions into five categories: Disorders limited to childhood; Disorders limited to adulthood; Disorders with changes in presentation across ages; Disorders with changes in management across ages; Disorders with no changes across ages. We evaluated Llama-2-70b-chat (70b) and GPT-3.5 (GPT) capabilities at generating accurate medical vignettes for these conditions based on Correctness, Completeness, and Conciseness as graded by 3 clinicians. Using accurately generated vignettes as in-context prompts, we further generated and evaluated patient-geneticist dialogues and assessed LLM performance in answering specific questions regarding age-based management plans for a subset of conditions. Results revealed impressive performances of 70b with in-context prompting and GPT in generating vignettes. We overall did not observe age-based biases, though our experiments identified statistically significant differences in some areas related to LLM output. Despite impressive capabilities, LLMs still have limitations in clinical applications.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1050W: Bioinformatics-based meta-analysis of gut microbiome in epilepsy

Authors:

V. Yadav[1], T. Abdulrehman[2], S. Khader[3], A. Titus[4]; [1]Glenda Dawson High Sch., Pearland, TX, [2]DeGroote Sch. of Business, McMaster Univ., Hamilton, ON, Canada, [3]Imperial Coll. London, London, United Kingdom, [4]Houston Methodist Hosp., Houston, TX

Abstract:

Idiopathic epilepsy, affecting approximately 0.6% of the global population, arises without identifiable symptoms or causes. Treatment-resistant epilepsy, impacting 30% of patients, persists despite optimal medical therapy. Both subtypes lack extensive pathway data and diagnostic opportunities, prompting our comparative analysis. The relationship between the microbiome and specific types of epilepsy has been understudied and may reveal new

therapeutic targets. We conducted a meta-analysis using 6 studies of the biological pathways and microbiome signatures in idiopathic and drug-resistant epilepsy. Pathway-level evidence helped identify potential links between certain gut bacteria and epilepsy types. The gut microbiome genera Neisseria and Delftia were significantly increased in both idiopathic and resistant epilepsy ($p<0.05$). Conversely, Coprococcus and Ruminococcus were downregulated in idiopathic patients but upregulated in resistant patients ($p<0.05$). Thirteen unique genera were upregulated in drug-resistant epilepsy compared to six in idiopathic epilepsy. Through Enrichr analysis, we found the LGI1 gene involved in idiopathic epilepsy, while ABCC2 was prominent in drug-resistant epilepsy. ABCC2 is a part of the bile secretion pathway, whereas LG1 is associated with the Developmental Biology R-HSA-1266738 pathway. A gene comparison revealed three common genes: EFHC1, SCN2A, and SCN1A ($p<0.05$). SCN2A is part of the taste-transduction pathway, and SCN1A is part of the dopaminergic synapse pathway, both upregulated in both subtypes. EFHC1 is significant in the Bardet-Biedl Syndrome pathway, affecting vision. Utilizing existing medications targeting these pathways could improve prognosis for patients with these epilepsy forms. Targeting upregulated microbiota can enhance our understanding of epilepsy development and lead to potential precision antibiotic therapies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1051W: Characterizing discrepant allele frequencies across the gnomAD v4 exomes and genomes

Authors:

Q. He[1], J. Goodrich[1,2], M. Wilson[1], K. Chao[1,2], K. Laricchia[1,2], D. Marten[1,3], P. Darnowsky[1], R. Grant[1], g. Consortium[1], H. Rehm[1,2], K. Samocha[2,1], K. Karczewski[2,1], M. Daly[2,1,4]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Boston Children's Hosp., Boston, MA, [4]Inst. for Molecular Med. Finland, Helsinki, Finland

Abstract:

The major goal of the Genome Aggregation Database (gnomAD) is to provide reliable population allele frequency (AF) estimates to enable variant classification. Previous studies have identified discordant genotype calls between exome and genome sequencing data from the same individuals, resulting in significantly different variant AFs in each data type. The increased sample size and diversity in gnomAD v4 offer more robust combined AFs, which are beneficial for downstream applications. Given different data generation platforms, these datasets offer an opportunity to identify variants with discordant AFs due to platform-specific artifacts. Here, we

describe the statistical tests used to flag these variants and address the observed discrepancies. To ensure the statistical tests identify differences due to sequencing technologies and not population structure, we handled variants present in a single genetic ancestry group differently from those present in multiple groups. We applied Hail's contingency table test to variants observed in a single ancestry group. To account for differences driven by genetic ancestry group structure, we applied Hail's Cochran-Mantel-Haenszel (CMH) test to variants present in multiple groups. Finally, for variants significantly different according to or unsuitable for the CMH test (assessed using the Breslow-Day test), we applied logistic regression on sample genotypes, using principal components from genetic ancestry inference as covariates.
We flagged variants as having discrepant AFs using a significance threshold of $p < 10^{-4}$. At this threshold, 1,416,137 (0.7%) out of 86,710,749 high-quality variants present in both data types were flagged. Notably, 81.8% of these flagged variants are found outside exome capture regions. 466,450 variants are not suitable for the CMH test based on Breslow-Day test results, highlighting the need for a more sophisticated approach to identify variants with discordant AFs. The results of the combined approach are available on the gnomAD browser as a “discrepant frequency flag”, and we will continue to refine our methods to more accurately pinpoint variants with discordant AFs due to technical artifacts.
The nuanced patterns of discrepant AFs identified by our statistical tests underscore the importance of harmonizing data from different sequencing and processing methodologies. Variants with discrepant AFs should be treated cautiously, as they may result in false positives. Identifying the reasons underlying significant variant AF discrepancies will help the community to make informed decisions in their analyses.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1052W: Comprehensive analysis of variant concordance reveals optimized average read-depth of human whole-genome sequencing data

Authors:

S. Joe, D. Lee, J. Kim, J. Jeon, J-h. Kim, I-K. Lim, K. Cho, J. Yang, S-Y. Kim; Korea Res. Inst. of BioSci. and Biotechnology (KRIBB), Daejeon, Korea, Republic of

Abstract:

Recently developed high-density single nucleotide polymorphism (SNP) arrays help identify diseases or genetic patterns of allelic imbalances. However, only limited number of loci could be

targeted by the arrays. With the falling costs of sequencing, low-depth sequencing is widely used in genome research. Therefore, lower-cost sequencing approaches could be an unbiased alternative to SNP arrays. Here, using the down-sampling approach, we evaluated the accuracy of small variant detection with next-generation sequencing (NGS) data according to read-depth size. In an approximate 4X set, the precision of single nucleotide variant (SNV) detection was around 0.92, covering more than 60% of 40X sequencing results. For SNVs called from the 10X dataset, the precision was 0.98, and coverage was about 91% variants of the 40X data set. We also compared short indel variants of the low-depth sequencing dataset with the 40X sequencing results using similar approaches. The overall performance rate was slightly different from SNV detection, and it had about 3% lower accuracy than that of SNV detection in the 20X set. Additionally, we compared the SNVs detected in coverage of 2X to 40X according to specific genomic regions, such as coding sequences or previous SNP array marker loci. The SNV concordance of these regions was much higher than that of the SNVs of all genomic regions. The 10X coverage exactly matched 40X results by over 93% with 0.99 precision. In this study, we report small variant detection rates, including genotype estimations, according to the size of the read depth. These results will aid population genome research in terms of appropriate coverage of whole-genome sequences.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1053W: Comprehensive Genomic Profiling of Stroke Cohorts in Taiwan: Integrative Analysis of ROH, Clonal Hematopoiesis, Pharmacogenetics, Mitochondrial Mutations, UPD, Short Variants, GWAS and HLA Using Next-Generation Sequencing.

Authors:

S-Y. Tsai[1,2,3,4,5], S-J. Hsieh[6], S. Hsieh[7], L-C. Liu[7], K-H. Wu[8], Y-H. Hsieh[8], C-K. Chang[9]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Dept. of Med., Mackay Med. Coll., New Taipei City, Taiwan, [3]Dept. of Lab. Med., Mackay Mem. Hosp., Taipei City, Taiwan, [4]Graduate Inst. of BioMed. Sci., Mackay Med. Coll., New Taipei City, Taiwan, [5]Inst. of Long term care, Mackay Med. Coll., New Taipei City, Taiwan, [6]MacKay memorial Hosp., Taipei, Taiwan, [7]MacKay Med. Coll., New Taipei City, Taiwan, [8]MacKay memorial Hosp., New Taipei City, Taiwan, [9]Hualien Tzu Chi Hosp., Hualien City, Taiwan

Abstract:

Introduction Stroke is the third leading cause of death and the most common cause of complex disability in Taiwan. Materials and Methods We recruited 605 stroke subjects from a medical center and conducted whole-genome sequencing using the Next-Generation Sequencing Analysis Acceleration Platform v20230823. This platform is known for its high throughput and accuracy, ensuring the reliability of our data. We compared the findings with a reference template, recruiting 1493 participants using Illumina DRAGEN v3.7.5, a widely used and trusted tool in genetic research. Results Our findings included 115 short genetic variants essential for understanding the genetic predisposition to stroke, 9 Clonal hematopoiesis variants, indicative of somatic mutations associated with hematopoietic cells, potentially contribute to stroke risk; 16 pharmacogenetic variants, which is key in understanding how genetic differences affect individual responses to medications; 202 ROH, identical by descent, which can influence stroke risk. 15 HLA variants, important for understanding immune response variability and its relation to stroke; Significant mitochondrial variants (chrMT: m.3242G>A, chrMT:m.3243A>G, and chrMT:m.3243A>T). We also identified several variants through GWAS, including 3 small vessel stroke variants,4 clopidogrel response variants, 5 large artery stroke variants, 1 ischemic stroke variant, and 4 cardiovascular event measurement variants. Additionally, we performed a polygenic score analysis of stroke-related traits and noted significant associations with Type 2 diabetes mellitus, coronary artery disease, and obesity. Discussion
This comprehensive genetic profiling study identified various variants from different tests that reflect their impact on the elevated risk of stroke. For instance, the SCN5A gene showed significant improvements through ROH testing. Rare variants in SCN5A are present in about 6% of atrial fibrillation (AF) cases, providing insight into genetic predisposition to stroke via cardiac pathways. The chrMT 3243A>G variant plays a significant role in detecting and managing stroke, particularly in the context of mitochondrial disorders like MELAS. In summary, addressing modifiable risk factors and improving acute stroke care are critical steps in reducing the stroke burden in Taiwan. Integrating comprehensive genetic findings is crucial for enhancing stroke diagnosis and management. By incorporating genetic testing into routine clinical practice, we can move towards a more personalized approach to stroke care, ultimately improving patient outcomes through targeted prevention and treatment strategies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1054W: Comprehensive Global Genome Diversity Resource : Bridging Gaps in Genomic Data Accessibility for Inclusive Health Research ⭐

Authors:

S. Mangul[1], W. Wolfsberger[2], Y-N. Huang[1], K. Chhugani[3], S. Nataneli[1], D. Yu[1], D. Ciorba[4], V. Munteanu[4], T. Oleksyk[5]; [1]Univ. of Southern California, Los Angeles, CA, [2]Oakland Univ., Rochester, MI, [3]Univ. of Southern California (USC), Los Angeles, CA, [4]Technical Univ. of Moldova, Chisinau, Moldova, Republic of, [5]Oakland Univ., Rochester HIlls, MI

Abstract:

Despite many advantages offered by the torrents of whole genome data flowing into various databases, the diversity they represent is still limited, creating a significant obstacle to achieving fair and unbiased solutions in global health. To this day, many geographic areas are represented by little or no genome data available to researchers creating a distorted picture of human variation. The fragmented nature of these datasets across various databases and platforms poses a significant challenge to researchers who require streamlined access to comprehensive genomic information. To help enhance the efficiency, accuracy, and collaborative potential of genomic research, we developed and implemented an accessible web resource - a Global Genomic Data Resource (GDDR) that consolidates and simplifies the search for genomic data. We used this resource to analyze the current state of global genome data availability, collecting and aggregating data contributions from projects publishing individual genomes or surveying genomic diversity within specific countries or regions. This resource can help uncover key insights into the global landscape of genomic diversity and representativeness, identifying prevalent inconsistencies and biases in the inclusion of diverse populations and substantial differences in data-sharing practices worldwide. A clickable map of the world provides links to accessible human genomes in various databases by specific location, availability, sequencing platform, and data format, and can aggregate available data by country and geographical region. The database is filled with the most current information and will be periodically updated to stay current with the new resources available. Using the most current GDDR dataset, we categorized 812,000 genomes available from four international and 29 national genomic projects across 104 countries and visualized a representation of genomic diversity and data-sharing practices. We advocate for a collaborative global effort to foster inclusive and openly accessible genomic studies, particularly by integrating underrepresented populations. We believe that the GDDR project can foster a collaborative research environment by providing a tool that supports data sharing and integration across the global scientific community. Specifically, researchers in developing regions or institutions with limited access to genomic databases can benefit

significantly from this centralized resource enabling more insightful analyses that can drive scientific discovery and innovation worldwide.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1055W: Comprehensive variant catalog and genotype-phenotype correlations for Joubert syndrome associated genes

Authors:

B. Henning[1], H. Loucks[1], A. Christman[1], P. Panchal[1], J. Du[1], V. Bowman[1], J. Dempsey[2], D. Doherty[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA

Abstract:

Joubert syndrome (JS) is a largely recessive neurodevelopmental condition overlapping clinically and genetically with Meckel syndrome (MKS), Nephronophthisis (NPH), and other ciliopathies. Over 40 genes have been associated with JS, around 20 with MKS, and about 40 with NPH. Understanding genotype-phenotype associations in ciliopathies is crucial for genetic test interpretation and prognostic counseling. We compiled a dataset of JS-gene variants linked to individual phenotypes from the literature to identify gene-diagnosis and variant-phenotype associations. We searched PubMed for papers on JS-gene variants in ciliopathy or neurodevelopmental phenotypes from 2004-2023. We extracted variant and phenotype information. Each variant was normalized (VariantValidator.org), its functional effect predicted (Ensembl Variant Effect Predictor), allele frequency extracted (Genome Aggregation Database - gnomAD), deleteriousness scored (Combined Annotation Dependent Depletion - CADD), and pathogenicity assessed based on American College of Medical Genetics and Genomics Technical Standards. Phenotypic features were mapped to Human Phenotype Ontology terms using an in-house algorithm. From 450 papers, we compiled 1748 unique variants in 2851 individuals assigned to 12 diagnostic groups: JS (32%), Retinal Dystrophy (19%), NPH (16%), MKS (11%), oral-facial-digital syndrome (9%), and others (each 1-3%). Of the 46 JS-associated genes, 9 were only associated with JS, 21 with both JS and MKS, and 13 with JS-NPH. CEP290 and NPHP1 were linked to the most conditions (7 each). Around 95% of the variants had allele frequencies ≤0.001 in gnomAD. About 85% of the variants had CADD scores >15. Genotypes of individuals with JS included truncating variants only (35%), missense variants only (26%), splice-altering variants only (4%), copy number variants only (2%), combinations of truncating/missense/splice-altering variants (16%), and other variant types (5%). In summary, we assembled a comprehensive and highly curated database of JS-gene variants and associated phenotypes. This database will aid in

determining the pathogenicity of future variants, analyzing genotype-phenotype correlations, identifying candidate variants for antisense oligonucleotide treatment, and contributing data to publicly accessible databases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1056W: Convolutions Accelerate mRNA Ramp Sequence Research in ExtRamp 2.0

Authors:

M. Cloward, J. Wride, G. Johnson, A. Stokes, T. Bass, J. Taylor, P. Ridge; Brigham Young Univ., Provo, UT

Abstract:

Ramp sequences are clusters of inefficient codons at the beginning of mRNA transcripts. Like highway on-ramps, ramp sequences slow down ribosomes at the beginning of translation, evenly spacing them along the transcript, reducing collisions, and increasing translational efficiency. Ramps are more frequently found in higher-expressed genes, evolutionarily conserved in vertebrates, and vary between tissues due to differential tRNA expression. Understanding ramp sequences relies heavily on our ability to detect them. ExtRamp, published by our lab (Miller et al., 2019), is the only available tool for detecting ramp sequences. However, its 2-hour runtime for a single human sample renders population-level studies impractical without supercomputing resources.

To address this limitation, we developed ExtRamp 2.0, a lightning-fast algorithm that processes human samples in just over a minute, over 100 times faster than its predecessor. ExtRamp 2.0 employs convolutions to overcome the previous algorithm's bottleneck. The original ExtRamp detects ramps by first calculating codon efficiencies and then putting them into efficiency arrays for each coding sequence (CDS). A sliding window harmonic mean is then calculated to simulate the ribosomal footprint. Finally, each CDS is evaluated to determine if its minimum mean is found at the beginning, indicating a ramp.

The bottleneck in the original algorithm's run time is the sliding window harmonic mean calculation. We overcame this using convolutions. ExtRamp 2.0 preprocesses each CDS efficiency array by calculating its reciprocal. A kernel the length of the window, populated by reciprocals of the window length, is then convolved with the reciprocated array using NumPy's convolve function, simulating a sliding window mean. The resulting array is then reciprocated once more to complete the harmonic mean.

We validated ExtRamp 2.0 on a dataset of 130,896,812 sequences in 21,924 species across diverse taxonomic groups. The results were consistent with the original ExtRamp algorithm, with discrepancies in only 0.22% of samples, primarily due to a bug we fixed in the original algorithm and minor rounding differences at the 16th decimal place.
ExtRamp 2.0's faster runtime makes large-scale ramp analyses more accessible, enabling us to explore the impact of ramps on gene and protein expression and their role in genetic diseases. ExtRamp 2.0 removes a significant bottleneck in ramp sequence research, paving the way for novel discoveries in translation regulation, mRNA optimization, and genetic disease research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1057W: Deciphering fine-scale population structure in the U.S. to uncover population-specific disease risks for health equity

Authors:

M. Isshiki[1], I. Raghuram[1], C. Smith[1], H. Gupta[1], T. Mutetwa[1], K. Tao[1], W. Jerome[1], C. Zhang[1], D. Yang[1], Y. Mokrab[2], P. Meissner[1], P. Spencer[3], M. Cabana[1], S. Klugman[1], M. Colón[4], Z. Maksumova[5], S. Suglia[6], C. Isasi[1], J. Greally[1], S. Raj[1]; [1]Albert Einstein Coll. of Med., Bronx, NY, [2]Sidra Med., Doha, Qatar, [3]Bronx Community Hlth.Network, Bronx, NY, [4]Hondurans Against AIDS/Casa Yurumein, Bronx, NY, [5]None, Bronx, NY, [6]Rollins Sch. of Publ. Hlth., Emory Univ., Atlanta, GA

Abstract:

Modeling the complexity and diversity of human populations is essential for achieving equitable health outcomes. Here we leverage the 245,388 whole genome sequences available in the All of Us Research Program to examine the diversity of the United States (US) within and between geographic regions, as well as their disease risks. By identifying identity-by-descent (IBD) sharing, the DNA segments shared among individuals due to inheritance from a common ancestor, we identified 58 groups with increased 'genetic similarity' (hereafter IBD groups) and characterized their population structure. The IBD groups ranged in size from size from 20 to 113,133 participants and showed different residential patterns across the US corresponding to census. We examined group-specific enrichments of pathogenic or likely pathogenic variants from ClinVar, predicted missense variants from AlphaMissense, and loss of function variants reported in gnomAD v3. We identified 51,406 IBD group-specific enrichments including known and new 'globally rare, locally common' disease risks, particularly in understudied groups with higher consanguinity. Phenome-wide association studies demonstrated different patterns of

common disease risk among IBD groups, confirming known risks, such as heterogeneous risk to asthma among Latino groups, as well as several new ones, such as increased risk to glaucoma among South Asian groups. Polygenic risk scores show more heterogeneity among IBD groups than has been reported for continental ancestries, suggesting that fine-scale population structure provides more accurate insight into genetic risk. Our results represent the first major effort at capturing both detailed population structure and population-specific patterns of both common and rare disease risk in the US.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1058W: Development of a novel gene-disease relationship classification framework and comparison to existing ClinGen standard

Authors:

M. Kiel, M. Racic, B. Mutapcic, E. Hrelja, M. Tica-Spahic, J. Bissonnette, M. Arapcic-Granic, S. Cajo, I. Cehic, A. Cickusic, A. Sito, M. Cviko-Nalbantic, S. Dzibric, J. Bugarin, S. Fatic, A. Gajic, L. Hafizovic, D. Hasanagic, A. Haskovic, A. Karic, R. Kerawala, E. Lakota, Z. Lepara-Dervisevic, H. Lokvancic, L. Mujkic, S. Muratovic, A. Plecas, S. Selimovic, N. Syverud, E. Toman, A. Zoranovic, A. Kurtovic-Kozaric; GENOMENON, Ann Arbor, MI

Abstract:

Accuracy of gene-disease relationship (GDR) curation has significant implications for clinical diagnosis. Variability in the evidence supporting these relationships necessitates structured evaluation to distinguish clinically valid relationships from less substantiated associations. Given the heterogeneity of interpretation guidelines, we sought to develop and test a novel method that would be more reproducible, faster, more comprehensive, and at least as accurate as established guidelines. Current frameworks lack specific scoring for animal models, detailed threshold for case-level evidence, and consistent weighting of functional evidence. We addressed these issues systematically aiming to improve consistency, comprehensiveness, and accuracy of GDR interpretations. Using this novel method we have curated 6754 genes. Whereas ClinGen's GDR dataset comprises 2153 genes, to ensure consistency of comparison we isolated GDRs curated according to ClinGen curation method versions 9 and 10.1 resulting in a subset of 179 genes common to both methods, with 268 phenotypes in ClinGen and 412 phenotypes identified using the novel method. Among the total 412 phenotypes assessed, 27% (111) had identical classifications, 38% (157) exhibited differing classifications, and 35% (144) were not found in ClinGen's dataset. Out of 157 phenotypes with different classifications, 104

(67%) phenotypes had a higher level of classification in ClinGen, 24 (15%) phenotypes had lower classification in ClinGen, and 29 (18%) phenotypes were classified as limited, disputed, refuted, or no known association in either database. Out of 104 phenotypes with higher classification in ClinGen, only 7 phenotypes had clinically significant differences whereas 97 phenotypes had clinically insignificant differences. Out of 24 phenotypes with higher classification in Mastermind, 12 phenotypes had clinically significant differences, and 12 had clinically insignificant differences. In total, out of 268 phenotypes only 19 (7.1%) showed clinically significant differences. Altogether, comparative analysis of this novel framework and ClinGen's demonstrates consistency across methods and highlights the ability of this new framework to identify a broader spectrum of associations.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1059W: Enabling cloud-based sharing, harmonization, and analysis of secondary use data in the multi-institutional Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium

Authors:

M. Conomos[1], S. Gogarten[1], B. Heavner[1], S. Nelson[1], A. Stilp[1], Q. Wong[1], A. Khan[2], M. Bowers[1], W. Hornsby[3,4], A. Huerta-Chagoya[4,3], E. Lange[5], L. Lange[5], Y. Li[6], A. Manning[3], J. Mercader[4], L. Yanek[7], Y. Zhang[8], K. Rice[1], PRIMED Consortium; [1]Univ. of Washington, Seattle, WA, [2]Pennsylvania State Univ., State College, PA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Broad Inst. of Harvard and MIT, Cambridge, MA, [5]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [6]Univ North Carolina, Chapel Hill, NC, [7]Johns Hopkins Univ, Baltimore, MD, [8]Univ. of Maryland Sch. of Med., Baltimore, MD

Abstract:

The Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium was established by NHGRI and NCI to develop new methods and identify best practices that improve the performance of polygenic risk scores (PRS) across diverse populations. As part of its work, PRIMED is increasing the diversity of data used to develop and validate PRS models by bringing together existing studies' genomic and health measures from participants of different genetic ancestries and geographic locations. The majority of PRIMED's 284 investigators, located across 49 institutions in 12 countries, are secondary users of the data that PRIMED is gathering for centralized sharing and collaborative analysis on NHGRI's Analysis Visualization and Informatics Lab-space (AnVIL) cloud platform.

We present policy and technical innovations developed in PRIMED's working groups (WGs) to address data sharing challenges due to the large and heterogeneous group of investigators, institutions, and secondary use datasets involved. PRIMED has established multiple data sharing mechanisms to enable coordinated data access in AnVIL workspaces. Data is organized in multiple data sharing workspaces (i.e. cloud buckets) that align with data access permissions, enabling differential access to investigators based on approved data access requests. To facilitate this, PRIMED's Coordinating Center (CC) developed an AnVIL Consortium Manager application for programmatic tracking, management, and auditing of user accounts, workspaces, and data access permissions.
We also present PRIMED's methods for collaborative data harmonization and analysis in AnVIL. The WGs developed the PRIMED data model, which enables investigators to rapidly merge consistent data across AnVIL workspaces to create harmonized, analysis-ready datasets, and we describe the CC's data processing workflows that check uploaded data for conformity. We also describe how investigators conduct analyses in separate analysis workspaces, from which they use file pointers to the central data storage workspaces, avoiding the need to make multiple copies of large files. To support reuse and reproducibility of their work, PRIMED investigators are implementing PRS and other relevant methods as analysis workflows, which are shared with the community via the PRIMED Dockstore organization and can be deployed directly to AnVIL or other analysis platforms.
The technical and policy approaches PRIMED has developed to facilitate collaborative sharing and use of diverse, pre-existing datasets in the cloud will benefit future consortia and efforts to leverage the vast amount of extant biomedical data towards improving health outcomes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1060W: Enhanced Genotype Imputation in the Chinese Population Utilizing the GSRD-100K^{WCH} Haplotype Reference Panel from 22,374 Deep Whole Genome Sequenced Subjects ⭐

Authors:

F. Ai, F. Bu, H. Yuan, L. Ke, Z. Mingjun, Z. Jiayi, Z. Qian, c. jing, L. Yu, Y. Ye; West China Hosp., Chengdu, China

Abstract:

The Genome Sequencing of Rare Disease (GSRD) -100K^{WCH} Haplotype Reference Panel, aimed at improving genomic imputation across diverse Chinese populations, was developed using whole

genome sequencing (WGS) data from 22,374 subjects with a total of 321,365,014 variants. This effort is part of the broader GSRD-100K^{WCH} project, dedicated to understanding the genetics of rare diseases. With a median sequencing depth of 38.9x, the dataset encompasses 47 ethnic minorities, including minority groups from southwestern regions such as Bai, Dai, Shui, Wa, and Miao, which were not covered by previous panels, thereby capturing a wide spectrum of China's genetic diversity. The reference panel was benchmarked against other well-regarded reference panels to ensure robust comparison and validation of its imputation capabilities. The GSRD-100K^{WCH} Haplotype Reference Panel substantially enhances imputation accuracy, evidenced by a 42.12% increase in the number of imputable sites and an average accuracy rate of 99.3%, coupled with a sensitivity of 99.0%. Moreover, the non-reference discrepancy rate (NRD) among the 12 evaluated minority populations decreased by an average of 8.60%, and the correlation coefficient (R^2) between imputed and true genotypes for rare variants (MAF < 0.5‰) improved by an average of 8.58%. Further validation through a genome-wide association study (GWAS) on pemphigus phenotypes, utilizing both imputed and original datasets, identified new genetic loci and confirmed existing associations, demonstrating the panel's capability to enhance genetic resolution. Supported by an online imputation server, the GSRD-100KWCH Haplotype Reference Panel provides a vital resource for genetic research and clinical applications, enabling deeper insights into the genetic foundations of complex diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1061W: Evaluation of new built-in AI features from the database management system to enable natural language conversation with local data

Authors:

H. Qiu[1], F. Mentch[1], J. Glessner[1,2], H. Hakonarson[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Innovations in generative AI are rapidly transforming the field of data management. Recently major commercial databases have incorporated new artificial Intelligence (AI) capabilities. In the latest release of the database management system that we currently use, two new AI features stand out particularly interesting: Select AI and AI Vector search.
We prepared a de-identified test dataset which consists of patient demographics, diagnosis, clinical notes, genomic mutations and annotations. We used Select AI to automatically generate

SQL query based on user's natural language prompts. Sample queries include "How many patients have diagnosis of Type II Diabetes?", "What's the earliest age of the diagnosis?". Select AI relies on Large Language Models (LLMs) for SQL synthesis. It supplements LLMs with metadata about the local database schemas to generate accurate and contextual SQL queries. We utilized AI Vector search to convert unstructured data into vector embeddings and generate vector indexes and performed efficient semantic similarity searches. Vector embeddings data can also be used to augment LLMs to help Select AI to better handle queries like "find more patients with a similar test result or gene mutation."
The initial evaluation results are promising. We found the two new AI features present an opportunity to enhance our existing in-house information system. However, AI also has its own limitations of bias and misinformation. It also introduces additional operational complexity and costs. We are using these features in specific use cases, such as a chatbot for study cohort exploration and optimization. This will allow us to further assess the benefits and limitations of AI integration in our existing system.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1062W: FaceBase: Data Resources for Craniofacial and Dental Development, Molecular Genetics, and Genomics

Authors:

J. Hacia[1], A. Bugacov[2], J. Feng[3], T-V. Ho[3], V. Nguyen[3], L. Pearlman[2], R. Schuler[2], C. Williams[2], C. Kesselman[2], Y. Chai[3]; [1]Keck Sch. of Med. of USC, Pasadena, CA, [2]USC Information Sci. Inst., Marina del Rey, CA, [3]Herman Ostrow Sch. of Dentistry of USC, Los Angeles, CA

Abstract:

Objectives: FaceBase, supported by NIDCR (U01DE028729), is a trusted source of research and educational resources on craniofacial and dental development in humans and animal models. FaceBase supports dental, oral, craniofacial (DOC) research and community-building through a Cloud-based, TRUST-worthy repository of high-quality FAIR data resources. FaceBase welcomes new contributions across the full spectrum of translational research. **Methods:** FaceBase is a highly scalable, Cloud-based (STRIDES) data sharing and analysis hub for dental, oral, and craniofacial data resources. Cloud storage provides extremely robust, cost-effective, virtually unlimited capacity and improves accessibility while ensuring long-term sustainability. FaceBase embraces TRUST (Transparency, Responsibility, User-focus, Sustainability, Technology) principles throughout to ensure the integrity of FAIR (Findable, Accessible, Interoperable, Reusable) data

resources. FaceBase supports the full translational spectrum of DOC research by enhanced data curation procedures for data on clinical and public health while strengthening basic research data handling. FaceBase provides resources to help the community meet the new Data Management & Sharing (DMS) requirements. **Results:** FaceBase includes 1,100+ datasets from human, zebrafish, mouse, chimpanzee, and chick, and seeks data on Xenopus and other model organisms. Data span micro-CT imaging, microscopy, single-cell and bulk RNA-seq, ChIP-seq, enhancer activity detection, exome sequencing, morphometric analysis, GWAS, and others. Data are findable via faceted search interfaces, accessible on the Web, interoperable using standardized vocabulary (e.g., UBERON, NCBI Gene, and MGI), and reusable with clear descriptions and detailed protocol documentation. FaceBase hosts the EnamelBase resource on morphological and functional data on tooth development, developed by NIDCR UH3-supported amelogenesis consortium. FaceBase's imaging and facial scan collections are ideally suited for AI/ML. **Conclusions:** FaceBase is home to a growing collection of data from the community across the full translational spectrum as we collectively build a comprehensive, trustworthy data repository and educational resource on craniofacial development. FaceBase promotes multidisciplinary collaboration and research in craniofacial and dental development, molecular genetics and genomics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1063W: From query to visualization: programmatic access and interactive workspaces for Alzheimer's disease genetics and >100,000 omics datasets

Authors:

E. Greenfest-Allen, J. Cifello, P. Kuksa, S. Tate, L. Carter, O. Valladares, W-P. Lee, Y. Leung, L-S. Wang; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: The NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) is a repository for Alzheimer's disease (AD) genetic data. To facilitate data reuse, NIAGADS has developed an Open Access API enabling researchers to leverage hosted unrestricted genomic annotations and AD genetic associations curated from the literature, as well as the repository, in analysis pipelines. It also provides programmatic access to FILER, an indexed disease-agnostic database of >100k standardized and harmonized, searchable functional genomics tracks. Here we introduce a refactored API that includes a collection of tools that guide users from query

generation to data visualization, creating interactive workspaces for data exploration.
Method: The NIAGADS API meets OpenAPI standards to provide an intuitive interface for programmatic access. The API and its server-side processing are written in a high-performance Python web-framework that leverages type-hints to generate the API structure and documentation and to enhance the security of the API through typed-based validation of user requests. Calls to the API are simple, templated URLs that specify a data track or feature ID (incl., gene symbol, refSNP, genomic interval) and filter criteria using Boolean expressions. This is paired with client-side visualizations encapsulated in reusable JavaScript/React components that allow users to get API responses not only as structured text (JSON), but as interactive tables, charts, or as tracks loaded into NIAGADS customized IGV genome and LocusZoom browsers. Communication between the server-side API and client-side visualization tools is mediated by an in-memory database that provides temporary storage of query results; this system enables paging of large user queries so that users can very quickly retrieve data in manageable chunks or retrieve past query results for a limited time.
To help users navigate this system, we provide a user gateway that leverages a large language model (LLM) to decomposes natural language inquiries into tasks and subtasks that plans and optionally executes API calls to return the result or requested visualization endpoint. The LLM is trained using the OpenAPI specification defining the NIAGADS API, example tasks, and domain-specific controlled vocabularies.
Conclusion: The addition of generative AI driven search and interactive, visual endpoints to the NIAGADS Open Access API infrastructure makes NIAGADS services more accessible to a broader user community. Information about accessing this toolkit as it becomes available can be found at https://api.niagads.org.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1064W: AI cloud-based end-to-end technology for accurate, fast and affordable newborn screening of genetic disorders / rare diseases

Authors:

C. Kyriakidis[1], M. Garrido Navas[2], M. Chaushevska[1], A. Schack[1], Z. Velkoski[1], G. Madjarov[3], L. Krych[4]; [1]gMendel, Copenhagen, Denmark, [2]Univ. of Granada, Granada, Spain, [3]Univ. Ss Cyril & Methodius, Skopje, North Macedonia, The Republic of, [4]Univeristy of Copenhagen, Copenhagen, Denmark

Abstract:

Current Newborn Screening (NBS) relies on less accurate conventional techniques like tandem mass spectrometry, as advanced genomic technologies such as Next Generation Sequencing (NGS) are cost-prohibitive and complex. With the societal burden of rare diseases in the US reaching $900 billion, there's a glaring clinical need for improved, swifter, and more affordable NBS. Various technologies are employed to diagnose genetic disorders, depending on the specific genetic alteration. For chromosomal abnormalities such as Klinefelter syndrome (KS) or Turner syndrome (TS), karyotyping/FISH/microarrays are used although the resolution is limited for small chromosomal rearrangements and mosaicism. For point mutations such as cystic fibrosis (CF), standard NGS is performed after a clinical suggestion by sweat chloride test (SCT). Finally, diagnosis of spinal muscular atrophy (SMA) caused by bi-allelic mutations in the *SMN1* is diagnosed by RT-PCR for mutations in *SMN1* and for copy number of *SMN2*. The necessity to employ diverse methodologies depending on the specific disease presents a suboptimal approach in newborn screening. A unified, comprehensive screening methodology would greatly enhance the efficiency and accuracy of NBS protocols. A unified, comprehensive screening methodology would greatly enhance the efficiency and accuracy of NBS protocols. **Methodology:** We collected samples from patients affected with KS, TS, CF or SMA from Denmark, Egypt and Paraguay and prepared sequencing libraries for ONT sequencing using our standard operating procedure. Our approach allowed simultaneous assessment of 384 samples, equivalent to 96 subjects, each in 4 technical replicates. We multiplexed all samples into one R10.4 flow cell and sequenced them in a GridION device. The analysis based on artificial intelligence was performed on CE-IVD Phivea platform resulting in disease classification. **Results:** As anticipated from preliminary data utilizing DNA sourced from biobanks, the Phivea platform has demonstrated its capability to accurately identify various genetic disorders with high sensitivity and specificity. This achievement not only reduces costs but also accelerates the diagnostic process, showcasing its practicality for newborn screening. **Conclusions:** The Phivea platform has proven adept at detecting chromosomal aneuploidies, point mutations, and copy number alterations, making it an ideal screening tool for conditions such as KS, TS, CF, and SMA, among others. Presently, we are conducting experiments involving the screening of the 64 genetic disorders included in NBS across all European countries.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1065W: Genome structure and function in individuals with sex chromosome aneuploidies: A systematic review.

Authors:

M. Staszewski[1], M. Mastronardo[1], G. Butera[2], A. Dakic[1], A. Raznahan[1], M. Legüe[1]; [1]Section on Dev.al Neurogenomics, Human Genetics Branch, Natl. Inst. of Mental Hlth., NIH, Bethesda, MD, [2]Office of Res. Services, Div. of Library Services, NIH Library, Bethesda, MD

Abstract:

Background: Sex chromosome aneuploidies (SCAs) are conditions defined by an atypical number of sex chromosomes with karyotypes other than XX or XY (e.g., XO, XXY, XXX, XYY, XXYY). SCA increases the likelihood of psychopathology while also impacting other organ systems. These clinical manifestations stem from the proximal effects of altered sex chromosome dosage (SCD) on genome structure and function. Such SCD effects may also be relevant for some phenotypic sex differences. As interest and tools for the genomic study of SCDs are growing, we sought to survey existing literature to reveal trends and gaps in research practice to date, as well as consensus findings regarding SCD effects of epigenetic and transcriptomic properties of the genome. **Methods:** Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we selected the following search concepts: “sex chromosome aneuploidies” (population), “gene expression or transcriptional regulation” (exposure), and “impact on genome structure and function” (outcome). We first refined our PubMed search using the tool PubVenn. We then searched eight databases for original articles, dissertations, theses, or reports from database conception up to the search date. We used Covidence to screen the titles and abstracts of 2,149 articles. Of these, 130 full text articles were reviewed, and data extraction was completed for 58 included articles. **Results:** We report a recent surge in publications on the genomic effects of SCA. This trend survives control for secular increases in publication rate. Most research has focused on the 45,XO or 47,XXY karyotypes, resulting in an underrepresentation of other karyotypes. Studies combining multiple karyotypes remain rare. Gene expression was the most studied SCA effect. Source materials included blood, stem cells, gonads, and fibroblasts. Studies on 45,XO and 47,XXY gene expression account for most of the included articles. We found high heterogeneity in effect measures, methods, and tissues reported. Despite this heterogeneity, we recover several convergent signals of SCA impacts on the genome, including extreme dosage sensitivity of gametolog genes. **Conclusions:** This review provides an integrative overview of over 15 years of research on the impact of SCA on genomic organization. High-priority areas for future research include the following: combining information across different karyotypes and -omic features, linking genomic effects to cellular outcomes, understanding tissue-specific effects, and incorporating patient-derived *in vitro* models for clinically important but less accessible organs like the brain.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1066W: Genome-wide allele-specific expression analysis aids interpretation of disease associated common and rare variants ⭐

Authors:

T. Huan[1], A. Keshawarz[2], R. Joehanes[1], C-T. Liu[3], C. Liu[4], J. Ma[5], S-J. Hwang[1], C. Demirkale[6], P. Courchesne[1], P. Munson[6], J. Murabito[7], D. Levy[1]; [1]NHLBI/NIH, Framingham, MA, [2]Yale Univ., New Haven, CT, [3]Boston Univ. SPH, Boston, MA, [4]Boston Univ. Sch. of Med., Boston, MA, [5]Tufts Univ., Boston, MA, [6]NIH, Bethesda, MD, [7]Boston Univ. Sch. of Med., Framingham, MA

Abstract:

Background: Allele-specific expression (ASE) refers to the phenomenon where one parental allele at a genetic locus is preferentially expressed versus another allele. ASE analysis is a more powerful and targeted approach than expression quantitative trait locus (eQTL) analysis, as it can evaluate the expression of an allele at a regulatory or pathogenic site at both the population and individual levels. We hypothesized that identifying ASE will help reveal the functional significance of common and rare disease-associated SNPs identified in genome-wide association studies (GWAS).

Methods and Results: We performed genome-wide ASE analysis on whole blood samples from 2,591 participants in the Framingham Heart Study. Genotypes of the studied samples were measured by whole-genome sequencing, and expression levels were measured by RNA-seq. ASE analysis was performed at each heterozygous site in each individual using GATK. We identified 12,817 SNPs in 3,144 genes that showed significant ASE in ≥5 individuals at a Bonferroni-corrected $P<2.0\times10^{-7}$ using Binominal tests. Approximately 50% of ASE SNPs have not been identified as *cis*-eQTLs, particularly for low-frequency (0.01≤MAF<0.05, N=1298, 76%) and rare variants (MAF<0.01, N=501, 85%). ASE SNPs were significantly enriched for synonymous, missense, and stop-gain variants, as well as variants in the 3' untranslated region and splice regions. There are 1,799 ASE SNPs that coincide with reported clinical phenotype-related variants and 488 SNPs predicted by CADD (Combined Annotation Dependent Depletion) to be deleterious variants. For example, rs138650483 in *MS4A6A* is a rare variant (MAF=0.004) with CADD score=26.2 that has been reported in association with late-onset Alzheimer's disease. Rs138650483 was not associated with overall gene expression levels of *MS4A6A* in the studied population (P=0.82). However, among the individuals carrying the rare allele (C), the rare allele was expressed significantly higher than the reference allele (T) at $P<2.0\times10^{-7}$. We also observed that individuals carrying the rare allele had significantly lower expression of isoform ENST00000420732.6 ($P=7.5\times10^{-17}$) and higher expression of isoform ENST00000528851.5

(P=4.3x10^{-46}) of *MS4A6A*, suggesting that the rare allele may affect gene function and contribute to disease.
Conclusions: Our study provides a valuable ASE resource, offering additional insights beyond eQTL analysis by identifying more expression-associated genetic variants. It also helps re-evaluate the functional impact of disease-associated variants identified in patients or through GWAS.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1067W: Genome-wide de novo tandem repeat variation in a four-generation family extensively sequenced with multiple long- and short-read technologies.

Authors:

H. Dashnow[1], T. A. Sasani[2], T. J. Nicholas[3], T. Mokveld[4], E. Dolzhenko[5], M. Goldberg, E[3], B. Pedersen[3], M. Noyes[6], G. De Sena Brandine[7], H. Happ[3], D. Porubsky[6], K. Munson[8], K. Hoekzema[8], G. Garcia[6], M. Eberle[9], D. Neklason[2], L. Jorde[3], E. E. Eichler[8], A. R. Quinlan[3]; [1]Univ. of Colorado, Denver, CO, [2]Univ of Utah, Salt Lake City, UT, [3]Univ. of Utah, Salt Lake City, UT, [4]PacBio, Den Haag, Netherlands, [5]Pacific BioSci.s of California, Tampa, FL, [6]Univ. of Washington, Seattle, WA, [7]PacBio, Menlo Park, CA, [8]Univ of Washington, Seattle, WA, [9]PacBio, Oceanside, CA

Abstract:

DNA Tandem Repeats (TRs) include Short Tandem Repeats (1-6 bp motifs) and Variable Number Tandem Repeats (7+ bp motifs). These highly mutable and polymorphic loci comprise >8.5% of the human genome. More than 60 are implicated in human disease and many more influence gene expression and complex traits. Identifying genome-wide *de novo* TR mutations is critical for understanding their unique mutational dynamics and our ability to identify rare disease variants.
In the Platinum Pedigree project, we characterized genome-wide *de novo* TR variation in 28 individuals across four generations of a large CEPH/Utah family (pedigree 1463), from whom the fourth generation was recently collected. Using deep whole-genome sequencing from Pacific Biosciences HiFi data, we genotyped >7 million TRs using our novel computational methods TRGT and TRGT-denovo, which directly interrogate read evidence, reducing false positive *de novo* mutation (DNM) detection. We further assessed DNMs by transmission to the subsequent generation and orthogonal validation using ultra-long Oxford Nanopore Technologies (ONT) and Element Biosciences AVITI short reads. Long reads allowed us to implement a novel phasing

strategy and interrogate much larger and more complex loci than prior studies.
Across STRs and VNTRs we estimated a DNM rate of 1.6e-5 per locus per generation. In contrast, SNVs had a DNM rate of 1.17e-8. We found substantial variation across repeat motif and allele sizes, observing a “trough” in DNM rate for intermediate motif sizes. While homopolymers remain challenging, the inclusion of AVITI allowed us to validate many of these. TR DNMs were predominately paternal (74.7%). 32 TR loci were recurrently mutated, with two or more confident DNMs found within this pedigree. Of these, 16 were highly mutable, many showing expansion and contraction over multiple generations. At one locus we identified 12 mutation events across 22 transmissions, making it one of the most hypermutable loci identified to date.
This study is the first to comprehensively characterize TR mutations in a large family by integrating evidence from multiple long and short-read sequencing technologies. We identified DNMs at larger and more complex loci than prior studies. Participants consented to public release, including sequencing data, phased genome assemblies, and variant calls, making this resource an invaluable truth set for variant discovery, recombination, and transmission studies. This study highlights the complementary value of short and long-read sequencing technologies to gain insights into the structural dynamics of the human genome and enhance the analysis of TRs in studies of rare genetic diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1068W: Genomic Return of Results for The United States’ *All of Us* Research Program

Authors:

S. Topper[1], A. Zhou[1], C. Neben[1], **H. Hoban**[1], R. Gibbs[2], S. Gabriel[3], H. Rehm[3], K. Doheny[4], G. Jarvik[5], C-L. Wei[5], T. Dutka[6], R. Moore[6], A. Musick[6], A. Wise[6], All of Us Research Program; [1]Color Hlth., Burlingame, CA, [2]Baylor Coll. Med., Houston, TX, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Ctr. for Inherited Disease Res. (CIDR), JHU, Baltimore, MD, [5]Univ. of Washington, Seattle, WA, [6]NIH, Bethesda, MD

Abstract:

All of Us (the US Precision Medicine Initiative) is advancing precision medicine by developing a database of 1 million+ health-contextualized genome sequences, derived from a uniquely diverse population. Longitudinal participant engagement is a core aim, and return of genetic results to participants is an essential engagement strategy.

A return of results protocol was developed to provide meaningful preliminary health-related information to research participants, and engaging non-clinical genetic content, derived from research data. Multiple risk-mitigating components were incorporated, including: research confirmation for positive hereditary disease risk findings, no VUS reporting, genetic counseling support, and subsidized clinical confirmation. The protocol is IRB approved and operates under an FDA Investigational Device Exemption. Since 2020, the program has offered >310,000 eligible participants access to genetic ancestry, “traits”, and health-related results including hereditary disease risk and pharmacogenomics (PGx). Participants opt-in and view results through an online, digital platform. Genetic counseling is available to all participants.
To date >70% of eligible participants have engaged with the genomic return-of-results program by clicking on the email notification to consider accessing content. Of those requesting results, about 82% were individuals who have been underrepresented in biomedical research (UBR). Additionally, an equivalent portion of UBR availed themselves of the results program, although the composition is slightly different. 3% of result recipients have a positive finding in one of the ACMG Secondary Findings v2 Gene List and were offered free clinical confirmation testing. Over 3,000 participants have utilized an appointment with a genetic counselor where 93% stated they agreed their genetic counselor helped them understand their results. 91% of PGx result recipients have potentially impacted medications in one of seven PGx genes (CPIC Level A/FDA). Participant feedback about the results return process was overall positive via Likert-scaled questions and free text responses.
All of Us has established a new standard for engaging participants in research and has directly returned results to one of the largest and most diverse cohorts to date. Participants are highly engaged with their research results, which can help drive recruitment and retention and ultimately increases the value of the program’s research database. We have also demonstrated the feasibility of using research participation as an opportunity to provide personally meaningful and impactful health-related genetic information.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1069W: Harmonizing 60,000 whole genome sequence data to advance Alzheimer’s Disease genetics research - Joint effort by the Alzheimer’s Disease Sequencing Project and the Genome Center for Alzheimer's Disease ⭐

Authors:

P. Gangadharan[1], W-P. Lee[1], A. Kuzma[1], H. Nicaretta[1], L. Qu[1], L. Carter[1], Y. Ren[1], L. Cantwell[1], O. Valladares[1], Y. Zhao[1], T. Iqbal[1], M. Schmidt[2], P. Mena[2], Alzheimer's Disease Sequencing Project, C. Dalgard[3], B. Kunkle[2], W. Bush[4], E. Martin[2], A. Naj[1], J. Haines[4], M. Pericak-Vance[2], L-S. Wang[1], G. Schellenberg[1], F. Leung[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Miami, Miami, FL, [3]Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [4]Case Western Reserve Univ, Cleveland, OH

Abstract:

Background: The Genome Center for Alzheimer's Disease (GCAD) coordinates the integration and meta-analysis of all available Alzheimer's disease (AD) relevant whole genome sequencing (WGS) data to facilitate the goal of identifying AD risk or protective genetic variants and eventual therapeutic targets. The WGS datasets are generated through the collaboration of investigators from the Alzheimer's Disease Sequencing Project (ADSP) and GCAD. To minimize data heterogeneity introduced by different sequencing machines and protocols, GCAD processes all samples using standardized pipelines and performs quality control (QC) and quality assurance checks.
Methods: Raw sequencing data (FASTQs or BAMs) are aligned to GRCh38/hg38 by BWA, and variants (SNVs and indels) are called using GATK. Genomic VCFs (gVCFs) are merged and jointly called using GLNexus. SNVs and indels are annotated using the ADSP-developed annotation pipeline. To facilitate access and use of large joint-genotyped VCF files, we release compact VCF and GDS formatted files. Lastly, structural variants (SVs) are called per sample using Smoove and Manta, and joint genotyping is performed with GraphTyper2.
Results: The dataset (ADSP Release 5, (R5), anticipated 2024) includes ~60,000 genomes from >50 diverse cohorts representing four major ancestries: 47% Non-Hispanic White, 29% Hispanic or Latino, 16% Black or African American and 8% Asian. Data are deeply sequenced (average genome coverage: >30x). Individual sample CRAMs, gVCFs from GATK, and SV VCFs for a subset of the R5 samples (n = 36,361) were deposited into NIAGADS Data Sharing Service (DSS) (https://dss.niagads.org/) for public distribution in 2022. New samples in R5 will be similarly released after the joint call is complete. Additionally, joint-genotype VCFs on SNVs, indels, and SVs will be available, and these will undergo full QC and annotation process.
Conclusion: The ADSP and GCAD generate high quality SNV, indels and SV calls. Currently, the project is processing ~60,000 WGS samples sequenced primarily through the ADSP Follow-Up Study, which emphasizes ancestrally diverse populations. We anticipate this 2024 release will continue to benefit the research community studying AD genetics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1070W: High confidence small Open Reading Frames (smORFs) show evidence of translation outside annotated transcript boundaries

Authors:

R. Dawes[1], H. Kim[1], Y. Chen[1], J. Yu[2], K. Hjorleifsson[3], J. Howson[2], N. Whiffin[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom, [3]Digital Sci. and Innovation, Computational Biology – AI & Digital Res., Novo Nordisk A/S, Maløv, Denmark

Abstract:

Despite their significant contribution to human disease susceptibility, our mechanistic understanding of cis-regulatory elements in the human genome is far from complete. Novel small open reading frames (smORFs), identified in recent ribosome profiling (riboSeq) studies in different human tissues and cell lines likely act as negative regulatory elements in many protein-coding genes, however quality control and annotation of these regions has not been standardised, making it difficult to explore their role in human disease.
We aggregated 10 publicly available riboSeq datasets across 12 tissues. After mapping to Ensembl v108, and conducting several quality control filters, we identified 159,269 unique smORFs. We then clustered smORFs with matching start or end positions and high sequence similarity. We considered 13,164 (8.3%) smORFs found in ≥2 datasets as 'high-confidence' 30% of which had non-canonical start codons (i.e. non-ATG). The majority (61.8%) were found in the 5' untranslated regions (5'UTRs) of protein-coding genes, either located entirely within the 5'UTR ('uORFs') or overlapping the CDS region ('uoORFs'). Overall, 24% of protein-coding genes contained at least one high-confidence smORF, compared to only 2% of non-coding genes. High-confidence smORFs represent possible regulatory features which could contribute to disease if perturbed, especially in dosage-sensitive genes. 2,318 high-confidence smORFs could be precisely mapped to diagnostic genes for monogenic disease (PanelApp Green genes), while an additional 445 un-mapped smORFs overlapped these genes but could not be precisely mapped to any transcript within Ensembl v108. Un-mapped smORFs were often located just beyond currently annotated UTR boundaries. For example one smORF found in 3 riboSeq datasets partially overlaps the 5'UTR of EP300, linked to autosomal-dominant Rubinstein-Taybi syndrome 2. The start codon of the uORF is beyond the gene's currently annotated transcription start site, however transcripts detected in long-read sequencing of prenatal and postnatal cortex include the full uORF. UTR length is a major driver of transcript diversity in the human brain, and we identify the inclusion and exclusion of uORFs as cis-regulatory elements differentially impacting protein translation as a possible result of this diversity.
Our high-confidence set of smORFs highlights regions where the regulatory impact of genetic variation should be considered. Our comprehensive and standardised set of high-confidence

human smORFs is a valuable resource to enable improved understanding of the impact of regulatory variation on human disease.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1072W: Landscape of Colorado Center for Personalized Medicine research freeze 4

Authors:

N. Rafaels[1], S. Chavan[1], M. Fisher[1], J. Shortt[1], N. Pozdeyev[2], E. Todd[1], K. Crooks[2], C. Gignoux[1], C. on behalf of[3]; [1]Colorado Ctr. for Personalized Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]

Abstract:

The Colorado Center for Personalized Medicine (CCPM) has completed a research freeze with 94,826 individuals comprising 72,428 samples from whole exome sequencing (WES) and 22,398 samples genotyped on the Multi-Ethnic Genotyping Array (MEGA). 44,761,373 common sites were imputed across all platforms. Filtering of WES and MEGA data to American College of Medical Genetics (ACMG) Recommendations for Reporting of Incidental Findings revealed 2416 and 738 patients, respective to data type, and 1442 variants identified for confirmatory genotyping. GWAS summary statistics are available for 1722 phecodes derived from electronic health records (EHR), as well as imputed blood and HLA types. Briefly, 24,096 new samples were sequenced using hg38 Twist comprehensive exome kit with additional non-exonic regions sequenced at low coverage (GxS) to yield genomic backbone coverage sufficient for whole-genome imputation. After removing 757 samples for contamination, poor depth of coverage, gender discordance, unexpected duplicates, and patient withdrawal, we performed variant QC on 7,244,061 and 1,415,009 variants from WES and GxS, respectively, keeping 6,776,929 and 1,224,474 variants, respectively. We then merged the most recent batch of WES data with 71,587 samples, 49,189 samples with WES and 22,398 non-overlapping samples genotyped on MEGA (QC methods: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10806731), keeping sites genotyped in both datasets with concordant alleles, after which principal component analysis was performed. We then performed imputation against TOPMed r2 reference panel, restricting to variants with MAF>0.00005 or inclusion on Illumina Global Diversity Array to achieve sufficient overlap, leaving 1,853,339 autosomal variants. We then removed variants with R^2<0.55 (overlapping sequenced and imputed sites are highly concordant (>70%) down to R^2=0.55). We then merged the new RGC imputed batch with previously imputed data from

MEGA and RGC using IMMerge, keeping variants in common. WES QC'd variants excluded from imputation input replaced imputed variants and we adjusted R2=1 and recalculated allele frequency. ACMG variants from previous freezes were re-evaluated based on SF v3.2 and clinvar release 09302023 adding an additional 585 patients and 206 variants. Our current research freeze is of one of the largest Biobank GWAS datasets in the United States, and genotyping samples from additional whole exomes and GWAS array is underway. PheWAS results are made available on the Colorado Biobank Portal, a cloud-based solution for interactive exploration of GWAS results. In research freeze 5, we will re-impute all samples to TOPMed r3.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1074W: Leveraging semantic information in biomedical literature to annotate publicly-available samples and studies

Authors:

M. Ahmadian[1], H. Yuan[2], P. Hicks[1], L. Valtadoros[1], A. Krishnan[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Michigan State Univ., East Lansing, MI

Abstract:

Currently, more than 7 million genomic samples are publicly available in repositories like the Gene Expression Omnibus. Despite their potential to be reused by the biomedical community to significantly advance research and knowledge discovery, these data remain acutely underused. The primary impediment is the difficulty in efficiently finding all samples and studies relevant to a researcher's area of interest due to the metadata of these samples/studies are in the form of unstructured, unstandardized plain text. Recently, we developed an approach called txt2onto that can take as input unstructured metadata and use a large language model (LLM) and machine learning (ML) to annotate it with terms in controlled vocabularies of tissues, cell lines, and diseases. This approach has very good performance overall in annotating human samples and studies, including genomics, proteomics, and clinical trials. However, the dependence of this approach on the quality and number of labeled examples curated manually and available for training the underlying ML classifiers poses a major challenge. Either due to the tedium of manual curation or the lack of enough research, many tissue/cell-type and disease terms have little or no training data. Here, we address this challenge by developing a novel framework that uses text from more than 35 million PubMed abstracts to automatically generate labeled data containing several thousands of instances of unstructured biomedical text associated with terms in controlled vocabularies of diseases and cell lines. These instances are directly extracted from

snippets of text surrounding annotations from PubTator3, a text mining tool that provides annotations for various biomedical entities including species, diseases, and cell lines, in every PubMed abstract. These extracted text snippets, rather than sample metadata, are then used as input features to train our ML classifiers. Preliminary analyses show that the extracted text windows contain relevant terms that are predictive of the target ontology term, indicating that semantic information can be transferred from biomedical literature to biomedical metadata. By transforming the painstaking process of data labeling into an effortless task, our framework enables accurately annotating samples with attributes irrespective of how well-studied that attribute is along with the new benefit of doing so for samples from nearly any species, not just human. Overall, our work dramatically improves the findability of invaluable public data and facilitates its effective data reuse.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1075W: Long-chain fatty acid oxidation disorders: A comprehensive novel database of LC-FAOD gene variants, genotypes and phenotypes

Authors:

H. Richbourg[1], V. Rangel Miller[1], O. K. Japalaghi[1], M. AlSayed[2], P. Baker, II[3], S. Daugherty[1], T. Ekstein[4], S. C. Grünert[5], M. Kiel[6], A. Khan[7], L. Korngut[8], D. Marsden[1], S. A. Monteleon[6], I. Schwartz[9], J. Vockley[10], N. Miller[1]; [1]Ultragenyx Pharmaceutical Inc., Novato, CA, [2]Coll. of Med., Al-Faisal Univ., Riyadh, Saudi Arabia, [3]Univ Colorado, Anschutz Med. Campus, Aurora, CO, [4]Invitae corporation, San Francisco, CA, [5]Ctr. for pediatric and adolescent medicine, Univ. Med. Ctr., Freiburg, Germany, [6]GENOMENON, Ann Arbor, MI, [7]M.A.G.I.C. (Metabolics and Genetics in Canada) Clinic Ltd., Calgary, AB, Canada, [8]Dept. of Clinical NeuroSci.s, Univ. of Calgary, Calgary, AB, Canada, [9]Hosp. de Clinicas, Porto Alegre, Brazil, [10]UPMC Children s Hosp. of Pittsburgh and Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

Abstract:

Long-chain fatty acid oxidation disorders (LC-FAOD) are rare, life-threatening, treatable conditions detected through acylcarnitine profile analysis, identified through newborn screening (NBS) or clinical presentation. Confirming diagnosis with genetic testing is complicated by the rarity of the disorder, genetic and phenotypic heterogeneity, and high frequency of variants of uncertain significance (VUS). A new locus-specific database for variants in the six genes associated with LC-FAOD was established to collect and disseminate information

about disease associated variants in *ACADVL, CPT1A, CPT2, HADHA, HADHB* and *SLC25A20*. The database integrates data from a systematic literature review (using Mastermind) and a sponsored gene panel program with NBS results, biochemical and phenotype data, to support accurate and timely diagnosis of LC-FAOD. As of March 2024, the LC-FAOD database contains 5595 variants from 3397 individuals. 1870 (55%) had a molecular diagnosis (MDx) categorized as positive (+) (≥2 pathogenic or likely pathogenic) and 244 (7%) as potential positive (p+) (1 pathogenic or likely pathogenic with ≥1 VUS) broken down by gene as follows: *ACADVL* (828, 39%), *HADHA* (575, 27%), *CPT2* (445, 21%), *SLC25A20* (81, 4%), *CPT1A* (100,5%), *HADHB* (85, 4%). 1200 individuals had only 1 variant and 27 had variants in ≥2 LC-FAOD genes. 958 unique genotypes were reported among 2114 patients with a +/p+ MDx. Phenotypes were reported for 939 +/p+ patients, averaging 3.8 clinical signs per patient. Hypoglycemia and liver abnormalities were most common in patients < 1 year and myalgia, myoglobinuria and rhabdomyolysis were most common in patients ≥13 years. The LC-FAOD gene database is now available at www.rarediseasegenes.com. This interactive and searchable website includes variant and genotype tables annotated with associated frequency, classification, and impact, along with biochemical findings, phenotypes, geographies, and publications reporting the variant. Variant locations are visualizable relative to gene structure. Summary analyses, including variant, biochemical, and phenotypes, are displayed in data charts that can be filtered and customized by users. **Conclusions:** Gene locus specific databases are important in rare diseases where variant-associated clinical information is scarce, and VUS are frequent. Genotype-phenotype data can aid diagnosis and VUS resolution. This comprehensive searchable LC-FAOD gene database is available online (www.rarediseasegenes.com) and provides an opportunity to add new variants, which will increase the diagnostic yield from genetic testing and be an important tool for LC-FAOD research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1076W: Matched Annotation from NCBI and EMBL-EBI (MANE): Current Status, Stability and Future Plans for expansion of the MANE dataset

Authors:

V. Joardar[1], J. Loveland[2], A. Astashyn[1], R. Bennett[2], A. Berry[2], E. Cox[1], C. Davidson[2], O. Ermolaeva[1], C. Farrell[1], R. Fatima[2], T. Goldfarb[1], J. M. Gonzalez[2], D. Haddad[1], M. Hardy[2], T. Hunt[2], J. Jackson[1], M. Kay[2], V. Kodali[1], K. McGarvey[1], J. Mudge[2], M. Murphy[1], S. Rangwala[1], F. Thibaud-

Nissen[1], A. Vatsan[1], C. Wallin[1], D. Webb[1], A. Frankish[2], S. Pujar[1], T. Murphy[1]; [1]NCBI, Natl. Library of Med., NIH, Bethesda, MD, [2]EMBL-EBI, Hinxton, United Kingdom

Abstract:

The collaborative Matched Annotation from the NCBI and EMBL-EBI (MANE) project defines a comprehensive set of matched, representative transcripts and proteins for human genes on the GRCh38 reference assembly. The MANE Select set represents an exactly matched pair of RefSeq and Ensembl/GENCODE transcripts for each protein-coding locus, chosen based on biologically relevant criteria. This set is useful as a universal standard for clinical reporting and for other applications that require a single transcript per gene. Additionally, a second transcript, the MANE Plus Clinical, is provided when MANE Select alone is insufficient to report clinical variants. Access to the MANE dataset is available on multiple RefSeq and Ensembl resources, and MANE data is available for download via FTP.

The current version of MANE (v1.3) encompasses >99% of human protein-coding genes and contains 19,288 MANE Select transcripts, including all genes in the ACMG SF v3.2 set. Sixty two MANE Select transcripts are annotated on alternate loci as they cannot be represented on the primary assembly. This release also includes 64 MANE Plus Clinical transcripts. To realize the full benefit of the MANE set, we strongly recommend the use of GRCh38. However, mappings of RefSeq accessions corresponding to MANE transcripts on GRCh38 are also available on the T2T-CHM13 and GRCh37 to support users working with these genome assemblies. These transcripts may have sequence differences to their GRCh38 equivalents.

MANE Select is intended as a stable set. Updates or changes to the set are rare and made only in egregious cases. An example is the MANE Select for *GP6* (glycoprotein VI platelet), which was changed to a different transcript accession based on AlphaFold protein structure prediction score.

We are working on expanding the MANE Select set to include non-coding genes with an established role in disease. These will be included in the next release. We are open to recommendations for candidates for the non-coding, and MANE Plus Clinical sets, and welcome your requests and feedback. While MANE Select and MANE Plus Clinical are valuable for clinical variant reporting, analysis of all transcripts of a gene representing the diversity of protein isoforms is crucial for variant discovery. In the long term, we aim to define a core set of transcripts that represent the functional alternative splicing diversity within a gene.

This work was supported in part by the National Center for Biotechnology Information of the National Library of Medicine, National Institutes of Health. Additional support is from NIH U24HG007234, Wellcome Trust 226083/Z/22/Z and Ensembl Core Funds.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1077W: Measuring the accuracy of electronic health record-based phenotyping in the All of Us Research Program using pathogenic *BRCA1* and *BRCA2* carrier frequencies in putative ovarian cancer cases

Authors:

J. Baierl, M. Jones, P-H. Lai, P-C. Peng, P. Pharoah; Cedars-Sinai Med. Ctr., Los Angeles, CA

Abstract:

Biobank programs like UK Biobank and the All of Us Research Program play a central role in the study of human genetics and disease prediction, by providing researchers with low-cost access to hundreds of thousands of whole genomes. However, these data rely on electronic health records (EHRs) to identify disease phenotypes of interest in participants. The accuracy of EHR-based phenotyping strategies is understudied. We have evaluated different strategies for identifying epithelial ovarian cancer (EOC) cases in the All of Us cohort by comparing the observed carrier frequencies of loss-of-function variants in the high-penetrance *BRCA1* and *BRCA2* genes in cases under different case definitions. We find substantially lower carrier frequencies of both *BRCA1* and *BRCA2* pathogenic variants among participants with only a single qualifying EHR diagnosis code ($Freq_{BRCA1}=0.027$, $Freq_{BRCA2}=0.017$; n=513) compared to participants meeting a more stringent EHR-based phenotype criteria developed through the Electronic Medical Records and Genomics (eMERGE) Network ($Freq_{BRCA1}=0.045$, $Freq_{BRCA2}=0.049$; n=368). Both groups have elevated carrier frequencies relative to cancer-free controls ($Freq_{BRCA1}=0.0019$, $Freq_{BRCA2}=0.0038$; n=115,834). Taking the *BRCA1* carrier frequency under the more stringent eMERGE criteria as an estimate of the true frequency in cases gives a point estimate for the precision of the single-EHR case criterion of 77%. We show that this significantly reduces statistical power for downstream genetic association analysis. While the eMERGE criteria improves phenotype specificity by reducing false positive case calls, this comes at the cost of lowered sensitivity. We estimate that the eEMERGE criteria misclassifies 33% of affected individuals as controls. We discuss the implications of this, as well as potential mitigation strategies by merging survey data on personal health history with these EHR phenotype criteria.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1078W: MedGen at NCBI: Empowering Biomedical Discovery and Clinical Care by Integrating Multiple Ontologies

Authors:

A. Malheiro, M. Kane, T. Venkatapathi, B. Kattman; NCBI, NLM, NIH, Bethesda, MD

Abstract:

To support the full spectrum of biomedical discovery and clinical care, MedGen integrates multiple clinically relevant and sometimes orthogonal ontologies and vocabularies under one resource. The source ontologies in MedGen include the Human Phenotype Ontology (HPO), Mondo, and OrphaNet; vocabularies include a subset of the Unified Medical Language System (UMLS), Online Mendelian Inheritance in Man (OMIM), PharmGKB, GeneReviews, and NCATS' Genetics and Rare Disease (GARD) dataset. Additionally, to support submissions to NCBI projects novel terms are created and assigned an identifier. The result is a harmonized, robust resource of standardized disease and phenotype concepts and supplemental data. Data are integrated based on the frequency of data release from the individual sources to provide the most current information to users. As new data are processed, equivalent concepts are identified between sources and mapped to a single MedGen record that can be searched by concept name(s) or identifier(s). Significant efforts are involved in bringing in the ontologies and mapping concepts between sources with varying degrees of overlap. For example, the subset of UMLS concepts in MedGen including SNOMED CT, MeSH, and NCI thesaurus have significant overlap with HPO and OMIM, while Mondo and PharmGKB have fewer overlaps. Automated mapping of terminologies uses identifier mappings from the source (when available) or customized string mapping algorithms. When the automated pipelines find discrepancies in mapping concepts across vocabularies, the terms are assigned to curators for review. These cases are mapped manually when possible or the sources are contacted to resolve ambiguities. Annually, MedGen curators average 130 requests for creation, obsoletion or corrections to the Monarch Initiative GitHub boards and 100 review requests with UMLS. MedGen's concepts are relationally linked based on gene-disease associations and disease hierarchies, enabling identification and retrieval of related concepts. Definitions for the concepts, alternate names, and source identifiers are imported from primary sources and supplemental resources including MedlinePlus to support searches on the website by clinicians looking for genetic phenotype information. The goal is to provide a standardized set of phenotype data to facilitate data sharing and interoperability in all settings from research to point of care. This harmonized dataset is available via web interface, FTP and API, supporting multiple use cases. This research

was supported by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1079W: Mucopolysaccharidosis Type VII (MPS VII): A novel, online *GUSB* gene variant database

Authors:

S. Daugherty[1], T. Hamazaki[2], M. Hegde[3], R. Giugliani[4], T. MacKenzie[5], L. Dilworth[1], M. Kiel[6], O. K. Japalaghi[1], B. Lianoglou[5], D. Marsden[1], J. Merritt, II[1], T. Nalin[1], V. Pansare[1], L. Pollard[7], V. Rangel Miller[1], H. Richbourg[1], T. Sparks[5], N. Miller[1]; [1]Ultragenyx Pharmaceutical Inc., Novato, CA, [2]Osaka Metropolitan Univ., Osaka, Japan, [3]Revvity, Waltham, MA, [4]HCPA/UFRGS, Porto Alegre, Brazil, [5]UCSF Sch. of Med., San Francisco, CA, [6]GENOMENON, Ann Arbor, MI, [7]Greenwood Genetic Ctr., Greenwood, SC

Abstract:

Background: Mucopolysaccharidosis type VII (MPS VII) is an ultrarare, autosomal recessive, lysosomal storage disease caused by beta-glucuronidase (GUSB) enzyme deficiency. Patients present variably with fetal effusions, skeletal dysplasia, dysmorphology, short stature, and cardio-pulmonary signs. MPS VII is diagnosed clinically in association with GUSB enzyme deficiency and/or in association with two disease-associated variants in the *GUSB* gene. However, diagnosis can be confounded by the rarity of the disorder, variable presentation, and high frequency of variants of uncertain significance (VUS). A new locus-specific database (LSDB) was established to collect and disseminate information about disease-associated variants in *GUSB*. **Methods:** We tabulated data from multiple clinical laboratories, ClinVar, and a consolidated literature search, curated in Mastermind, to develop a comprehensive landscape of variants in the *GUSB* gene. Using Large Language Models (LLM), we have begun mining public literature for patient demographic and phenotype data associated with *GUSB* variants and genotypes. **Results:** As of 17 May 2023, the database reports 318 unique *GUSB* variants distributed as 221 (70%) missense, 21 (7%) nonsense, 21 (7%) splicing, 16 (5%) frameshift, 9 (3%) UTR, 7 (2%) intronic, 6 (2%) CNV, 3 (<1%) deletion, 1 (<1%) insertion, 2 (<1%) start-loss, and 11 (3%) synonymous. Clinical classifications available for 314 variants were 65 (20%) Pathogenic, 39 (12%) Likely Pathogenic, 209 (67%) VUS, and 1 with conflicting interpretations.
The *GUSB* LSDB is available at www.rarediseasegenes.com. This interactive, searchable website includes a variant table annotated with associated frequency, classification, impact, and

publications. Data are provided in aggregate to protect patient privacy in this ultrarare disease. **Conclusions:** This database reports 67% VUS among 318 *GUSB* variants, and we recently found that ~half of MPS VII patients had ≥1 VUS. With the increasing use of genome-based newborn sequencing, timely reporting and classification of MPS VII disease-associated variants is necessary to support accurate diagnosis and enable clinical study of in utero fetal therapy for this disorder. Rapid reporting of novel and previously observed *GUSB* variants to public databases, including www.rarediseasegenes.com, ClinVar and LOVD, aids a better understanding of MPS VII and supports timely diagnosis. A planned update of the *GUSB* database in Q4 2024 will integrate data from LLM analysis including patient genotypes, age, sex, geography, clinical and biochemical manifestations, disease severity, and fetal effusion (incl. NIHF) history.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1080W: Multi-organ imaging-derived polygenic indexes for brain and body health

Authors:

X. Yang[1], P. F. Sullivan[2], B. Li[3], Z. Fan[4], D. Ding[4], J. Shu[5], Y. Guo[1], P. Paschou[6], J. Bao[4], L. Shen[4], M. Ritchie[7], G. Nave[8], M. L. Platt[8], T. Li[2], H. Zhu[2], B. Zhao[4]; [1]Purdue Univ., West Lafayette, IN, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univ. of California, Los Angeles, Los Angeles, CA, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Purdue Univ., Plano, TX, [6]PURDUE Univ., West Lafayette, IN, [7]Univ. OF PENNSYLVANIA, Philadelphia, PA, [8]Univ. of Pennsylvania, Philadephia, PA

Abstract:

The UK Biobank (UKB) imaging project is a crucial resource for biomedical research, but is limited to 100,000 participants due to cost and accessibility barriers. Here we used genetic data to predict heritable imaging-derived phenotypes (IDPs) for a larger cohort. We developed and evaluated 4,375 IDP genetic scores (IGS) derived from UKB brain and body images. When applied to UKB participants who were not imaged, IGS revealed links to numerous phenotypes and stratified participants at increased risk for both brain and somaticdiseases. For example, IGS identified individuals at higher risk for Alzheimer's disease and multiple sclerosis, offering additional insights beyond traditional polygenic risk scores of these diseases. When applied to independent external cohorts, IGS also stratified those at high disease risk in the All of Us Research Program and the Alzheimer's Disease Neuroimaging Initiative study. Our results demonstrate that, while the UKB imaging cohort is largely healthy and may not be the most

enriched for disease risk management, it holds immense potential for stratifying the risk of various brain and body diseases in broader external genetic cohorts.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1082W: NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS): 2024 Update

Authors:

J. Manuel[1], A. Kuzma[1], O. Valladares[1], H. Nicaretta[1], P. Gangadharan[1], Z. Katanic[1], M. Kirsch[1], Y. Ren[1], L. Bass[1], N. Saravanan[1], J. Brettschneider[1], A. Wilk[1], Y. Zhao[1], L. Qu[1], M. Moon[1], A. Lerro[1], P. Keskinen[1], J. Cifello[1], W. Horng[1], S. Tate[1], R. Flawless[1], H. White[1], L. Carter[1], P-C. Ho[1], Y. Jin[1], S-Y. Chou[2], A. Naj[1], E. Greenfest-Allen[1], P. Kuksa[1], F. Leung[1], W-P. Lee[1], G. Schellenberg[1], L-S. Wang[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Lehigh Univ., Bethlehem, PA

Abstract:

NIAGADS is a national genomics data repository that facilitates access of genotypic and sequencing data to qualified investigators for the study of the genetics of Alzheimer's disease (AD) and related neurological diseases. Collaborations with large consortia and centers such as the Alzheimer's Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Alzheimer's Disease Sequencing Project (ADSP), and the Genome Center for Alzheimer's Disease (GCAD) allow NIAGADS to lead the effort in managing large AD datasets that can be easily accessed and fully utilized by the research community.NIAGADS is supported by National Institute on Aging (NIA) under a cooperative agreement. All data derived from NIA funded AD genetics studies are expected to be deposited in NIAGADS or another NIA approved site. NIAGADS manages a Data Sharing Service (DSS) that facilitates the deposition and sharing of genomic data and association results with approved users in the neurodegenerative disease research community. In addition, researchers are able to freely use the NIAGADS Alzheimer's Genomics Database to search annotation resources that link published AD studies to AD-relevant sequence features and genome-wide annotations.As of May 2024, NIAGADS houses 124 datasets comprised of over 209,000 samples including array data, sequencing, gene expression, annotations, deep phenotypes, summary statistics, among others. Qualified investigators can retrieve ADSP sequencing data with ease and flexibility through the NIAGADS DSS. To date, the ADSP and other contributing studies have completed whole exome sequencing (WES) of 20,499 samples and whole genome sequencing (WGS) of 36,352 samples. Raw WES and WGS files, quality

controlled VCF files, and phenotype data files are available via qualified access. The next round of sequencing currently underway will generate around 30,000 additional genomes to be released in mid-2024.NIAGADS is a rich resource for AD researchers, with the goal of facilitating advances in Alzheimer's genetics research. By housing datasets from many projects and institutions, NIAGADS enables AD researchers to meet their research goals more efficiently.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1083W: Optimization of Insert Sizes for Illumina WGS with a Combination of Covaris AFA based Shearing and Bead-Based Size Selection

Authors:

M. Ambavaram, V. Process, M. Werner, U. Thomann; Covaris, Woburn, MA

Abstract:

The drivers for the whole genome and exome sequencing include growing demand for personalized medicine, advancements in sequencing technologies, and increasing applications in molecular diagnostics. To accelerate biomedical and health research, the All of Us Research program is aiming to generate whole-genome sequence (WGS) and genotyping data on one-million participants (1). In precision oncology, WGS is increasingly applied to identify disease-causing genetic variation (SNPs or structural variants), disease susceptibility, cancer evolution and drug response, among a plethora of other applications (2). Standard WGS libraries currently tend to have a fragment size of ~350-500 bp for optimal sequencing results. A challenge when using Illumina's sequencing technology is that the insert sizes of the libraries are impacted by the process of cluster generation in which the libraries are distributed on the surface of the flow cell and then amplified. During this process, shorter fragments cluster and amplify preferentially resulting in an apparent disparity between library insert size determined by electrophoresis and sequenced insert size (3). To address this challenge, we conducted sequential experiments using Coriell DNA (NA12878) samples that were sheared using R230 Focused-ultrasonicator to three different target size distributions (175 bp, 350 bp & 550 bp) followed by library prep, an optimized final clean-up using Covaris-developed magnetic bead-cleaning reagents and sequencing on a MiSeq. Our results demonstrate that the novel protocol results in optimal insert sizes that maximizes coverage for WGS analysis. REFERENCES: (1) Genomic data in the All Of Us research program. 2024, Nature (627), 340-346. (2) Clinical utility of whole-genome sequencing in precision oncology. 2022, Seminars in Cancer Biology (84)32-39. (3) Optimizing

library insert size representation on NovaSeq X series instruments. 2024. Illumina Knowledge Article #8912.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1084W: Pathogenic variants of pulmonary hypertension associated genes in COPDGene

Authors:

V. Foris[1,2], A. Boueiz[1,3], K. Kim[1], L. Martini[1], C. Tern[1], R. Estepar[4], G. Washko[3], E. Silverman[1,3], M. Cho[1,3], COPDGene investigators, TOPMed investigators; [1]Channing Div. of Network Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [2]Med. Univ. of Graz, Graz, Austria, [3]Div. of Pulmonary and Critical Care Med., Dept. of Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [4]Surgical Planning Lab., Lab. of Mathematics in Imaging, Dept. of Radiology, Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA

Abstract:

Introduction: Pulmonary hypertension (PH) frequently occurs in individuals with chronic obstructive pulmonary disease (COPD), impacting their prognosis. Understanding the genetic background of COPD-PH could enhance disease classification and treatment decisions. We aimed to evaluate known pathogenic variants of genes associated with pulmonary arterial hypertension (PAH) in a large cohort of smokers. **Methods:** We analyzed whole genome sequencing and Phase 1 clinical data from the Genetic Epidemiology of COPD (COPDGene) cohort, consisting of non-Hispanic white and African-American smokers with and without COPD, as part of the Trans-Omics in Precision Medicine (TOPMed) program. We used WGSA v0.95, integrating outputs from ANNOVAR, VEP, SnpEff, and ClinVar, to annotate 12 genes with definitive evidence (BMPR2, ACVRL1, ATP13A3, CAV1, EIF2AK4, ENG, GDF2, KCNK3, KDR, SMAD9, SOX17, TBX4), 3 genes with moderate evidence (ABCC8, GGCX, TET2), 6 genes with limited evidence (AQP1, BMP10, FBLN2, KLF2, KLK1, PDGFD), and 5 genes with disputed evidence (BMPR1A, BMPR1B, NOTCH3, SMAD1, SMAD4) for association with PAH according to ClinGen. We assessed the prevalence of different classes of pathogenic variants and the clinical and chest CT imaging characteristics of variant carriers. **Results:** We analyzed 10,550 whole genomes and 133,060 intronic and exonic variants in previously identified PAH genes. Only 2,256 variants (1.7%) were reported in ClinVar, with 1,501 classified as benign/likely benign, 724 as variants of uncertain significance (VUS) or with conflicting interpretations, and 31 as pathogenic/likely pathogenic. We identified pathogenic/likely pathogenic heterozygous variants

in 6 genes with definitive evidence (14 variants in 19 subjects), including 5 pathogenic variants in EIF2AK4 carried by 6 subjects. Carriers of pathogenic variants, including those in genes with moderate or disputed evidence (36 subjects, representing 0.3% of the COPDGene cohort), showed no differences in demographic factors, lung function, or pulmonary vascular imaging phenotypes compared to carriers of VUS (5,279 subjects) or benign variants (4,845 subjects) in both univariable or models adjusted for age, sex, race, and genetic ancestry. **Conclusion:** Clinically annotated pathogenic variants in known PAH genes were identified in the COPDGene cohort. Cross-sectional analysis revealed no substantial clinical differences between carriers and non-carriers of these variants. Future research will incorporate non-ClinVar variants and leverage multi-omics and longitudinal data for a more comprehensive understanding of these genetic influences.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1085W: PhenoGemini: Enhancing Molecular Diagnoses of Mendelian Disorders through Identifying Twin Patients with Large Language Models ⭐

Authors:

Z. Chen[1,2], C. Jihao[1], Y. Yang[3], T. Hospedales[4], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Div. of Spine Surgery, Dept. of Orthopedics, Nanfang Hosp., Southern Med. Univ., Guangzhou, China., Guangzhou, China, [3]Queen Mary Univ. of London, London, United Kingdom, [4]The Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract:

Mendelian disorders (MDs) affect millions of births each year, underscoring the pressing need for molecular diagnoses. For now, clinicians diagnose patients predominantly by matching the patient's genotype and phenotype to cases documented in the literature. However, this approach faces significant challenges due to the extensive number of phenotype-genotype associations and the rapid growth of knowledge about MDs. The main challenges include: 1) Relatively limited phenotype-genotype associations in existing databases; 2) The difficulty in accurately quantifying the probability of phenotype occurrences; 3) The reliance on structured inputs, such as HPO codes, rather than natural language, complicates interactions, especially when phenotypes are not listed in HPO. To address these challenges, this study aims to leverage the unique features of large language models (LLMs) to establish an automated phenotype-genotype database for MDs, enabling clinicians to prioritize candidate genes using natural

language input. First, we fine-tuned an LLM to perform semantic searches across medical databases, including PubMed, which contains over 34,957,128 papers. The search strategy retrieved 1,355,400 papers related to Mendelian disorders and extracted 2,451,642 phenotype-genotype associations from their full texts. These results were validated against the Human Gene Mutation Database (HGMD) and demonstrated substantial enhancements, showing improvements of 97.2% in recall and 47.3% in precision compared to a traditional keyword-based search strategy. Second, we established a patient-level database, PhenoGemini, that facilitates the identification of causative genes. PhenoGemini enables the matching of twin patients among 240,624 individuals recorded in the database. Using the ability to rank causative genes in top-10 lists as an evaluation metric, baseline analysis demonstrated that PhenoGemini outperformed previous phenotype-driven gene prioritization tools by a relative increase of 72.8-118.5% in independent, real-world cohorts (N = 1769; PhenoGemini: 52.9%; PhenoApt: 30.6%, PMID: 35063063; Phrank: 24.2%, PMID: 29997393). Third, we introduced a large language model-based retrieval pipeline, PhenoGemini GPT, employing the Retrieval Augmentation Generation (RAG) technique. This pipeline enables clinicians to search for twin patient cases and engage in real-time, natural language dialogues. It further enhances diagnostic accuracy by 10.8%, particularly in hard-to-solve cases (PhenoGemini GPT: 58.6%; PhenoGemini: 52.9%). For further information, please visit PhenoApt.org or PhenoGemini.org.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1086W: Phenome-wide association study of *MTHFR* variant in UK Biobank and *All of Us*

Authors:

A. Awan[1], Y. Tufa[1], D. Shittu[2], C. Wilczewski[1], T. Tran[1], T. Ferrara[3], L. Biesecker[4], J. Denny[5], C. Turner[6], H. Mo[7]; [1]NHGRI, Bethesda, MD, [2]Natl. Human Genome Res. Inst., Silver Spring, MD, [3]NIH, Bethesda, MD, [4]NIH NHGRI, Bethesda, MD, [5]NIH, Bethesda, MD, [6]NIH/NIHGRI, Bethesda, MD, [7]NHGRI, Rockville, MD

Abstract:

Methylenetetrahydrofolate reductase (*MTHFR*) is an important regulator in folate and homocysteine metabolism. NM_005957.4:c.665C>T p.(Ala222Val), historically known as the C677T variant, has been linked to decreased enzymatic activity resulting in elevated homocysteine levels. Previous studies, often using case reports or small case series, have linked p.(Ala222Val) to conditions such as cardiovascular disease risk, thrombosis, pregnancy

complications, neural tube defects, psychiatric diseases and cancer risk. However, follow-up meta-analyses have not supported these associations. *MTHFR* is one of the most frequently ordered genetic tests in clinical settings, despite multiple expert groups such as ACMG and NSGC not recommending variant testing.
We performed phenome-wide association studies (PheWAS) in both *All of Us* and UK Biobank cohorts to assess clinically relevant associations with this variant. In the *All of Us* cohort with 245,386 sequenced individuals, we identified disturbances of amino-acid transport (OR=1.53 [1.3 -1.77]) and disturbances of sulphur-bearing amino-acid metabolism (OR=1.64 [1.40-1.92]) above the Bonferroni cut-off. Previous *MTHFR* associated conditions such as thrombosis (OR=1.31 [0.90-1.89]), major depressive disorder (OR=1.00 [0.98-1.02]), and breast cancer (OR=1.00 [0.96-1.04]) were not statistically significant in the *All of Us* dataset. In the UK Biobank cohort with 469,829 individuals, we were not able to identify any phenome-wide significant phenotypes. Previous *MTHFR* associated conditions such as thrombosis (OR=0.68 [0.44-1.04]), major depressive disorder (OR=1.00 [0.98-1.02]), and breast cancer (OR=1.01 [0.99-1.04]) were not statistically significant in the UK Biobank dataset.
This is the largest known dataset analysis of the p.(Ala222Val) variant. Both of the phenome-wide significant conditions may be related to abnormal biochemistry or genetic tests. Despite the large size of the two biobanks, we did not find any adverse outcomes significantly associated with this variant. Our PheWAS results support the current recommendations advising against *MTHFR* testing.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1087W: Prevalence of errors in lab-made plasmids across the globe

Authors:

C. Fernandes, X. Bai, J. F. Hong, G. S. Yu, D. Y. Hu, A. Y. Chen, C. A. Rich, S. J. Shi, S. Y. Xu, K. J. Müssar, D. W. Meng, J. L. Chen, B. T. Lahn; VectorBuilder Inc., Chicago, IL

Abstract:

Plasmids are an indispensable reagent in life sciences research and therapeutics development. Currently, most labs custom-build plasmids in a “do it yourself” (DIY) fashion at the bench. As yet, no systematic data on the quality of DIY plasmids exists. Here, we report a broad survey of plasmids made by hundreds of labs worldwide from both academia and industry. We found that nearly half of them contained design and/or sequence errors that could potentially compromise

function. For adeno-associated virus (AAV) vectors, which are widely used in gene therapy, we found that about 40% of transfer plasmids used to package AAV carried mutations in the inverted terminal repeat (ITR) regions relative to the wildtype due to their inherent sequence instability, and such instability is strongly influenced by flanking GC content. We also list genes that are difficult to clone into plasmid or package into virus due to their toxicity to host cells. Our findings raise serious concerns over the trustworthiness of a ubiquitous research reagent, which parallels the widespread mycoplasma contamination and misidentified mammalian cell lines reported previously, and highlights the need for community-wide standards to uphold the quality of gene delivery vectors in research and medicine.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1088W: Rapid Point-of-Care Genotyping for Sickle Cell Anemia Using CRISPR Technology

Authors:

A. Siji[1], S. Sharma[2], V. Gupta[1], V. Chandru[1], R. Arora[1]; [1]CRISPRBITS PVT LTD, Bengaluru, India, [2]TeOra, Bengaluru, India

Abstract:

Sickle cell anemia (SCA) is a genetic blood disorder characterized by the presence of abnormally shaped red blood cells (RBCs) that impair oxygen transport, caused by a single nucleotide mutation in the beta hemoglobin gene (*HBB*). Early diagnosis is critical for reducing mortality and improving patient management, particularly in regions with high prevalence, such as certain tribal communities in India. India contributes to 15% of the global annual births of children with SCA. In countries where early SCA screening is routine, over 94% of affected children survive, in stark contrast to sub-Saharan Africa, where 50-90% of children with SCA die before reaching five years of age. Current nucleic acid based diagnostic tests, which rely on sequencing, require the transport of samples to central laboratories, resulting in a turnaround time of several days from sample collection to results. Despite their accuracy, these tests are impractical for point of care settings due to logistical challenges, high costs, the necessity for expensive equipment, and prolonged wait times.We have optimized a point of care CRISPR-based diagnostic assay to detect SCA and carrier states at the genotype level. The assay combines Polymerase chain reaction (PCR) with CRISPR Cas12 recognition of specific nucleotide changes, facilitating the detection of normal (AA), diseased (SS), and carrier (AS) genotypes through fluorescence signal amplification. Specifically, a guide RNA is customized to target the

DNA sequence associated with either HbA or HbS. If the target DNA is present, a Cas12:guide RNA: target DNA complex forms, leading to trans-cleavage of labelled ssDNA and an increase in fluorescence signal. The assay includes two detection mixes one for HbA and one for HbS enabling discrimination between homozygous and heterozygous samples. A total of 52 samples were tested so far with 100% sensitivity and specificity despite presence of additional polymorphisms in the region which can confound guide binding. Our results demonstrate that the assay is able to distinguish between AA, SS, and AS phenotypes with high sensitivity and specificity, displaying distinct fluorescence patterns for each genotype.The assay is rapid, requiring minimal equipment and technical expertise and can be deployed in primary health care settings. This innovative approach promises to bridge the gap between laboratory-based diagnostics and field-based implementation, providing an effective solution for early and accurate detection of sickle cell anemia in resource-limited environments.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1089W: Taiwan View: A multiomic visualization tool for Taiwan Biobank data.

Authors:

C-W. Lin, M-W. Su, C-Y. Pan, Y-L. Wang, S-H. Wu, J-Y. Hsiao, H-M. Shih; Academia Sinica, Taipei, Taiwan

Abstract:

The Taiwan Biobank (TWB) is a large longitudinal population-based cohort offering a wide range of multiomics and phenotypic data to advance our understanding of human health. TWB has completed its first-phase goal of recruiting 200,000 participants and launched Taiwan View (https://taiwanview.twbiobank.org.tw), a web platform which provides access to summary information on genome, methylome, human leukocyte antigen (HLA), and metabolome data. For genomic data, we have 1,492 high-coverage whole genome sequences and approximately 150,000 samples genotyped using the Affymetrix Axiom TWB or TWB2 array available. Genotyped data were further imputed with a merged reference panel consisting of East Asian population data from the 1000 Genomes Project phase 3 and a population-specific Taiwanese panel. Phenome-wide association scans were conducted for imputed genomic data across 242 traits (131 binary, 111 continuous) using SAIGE, which accounts for case-control imbalance and relatedness. DNA methylation was profiled in 2,469 participants' whole blood using the Illumina EPIC array which covers over 850,000 CpG sites. NGS-based HLA types at up to 4-field resolution

for class I (HLA-A, -B, and -C) and class II (HLA-DR, -DQ, and -DP) antigens of 1,097 participants were obtained with NXType kit on the Ion S5 system. Concentrations of 150 metabolites in blood plasma from 2,262 participants were measured on the Bruker IVDr nuclear magnetic resonance (NMR) platform. Additionally, urine plasticizer and melamine concentrations of 1,799 participants were measured by liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS). Taiwan View summarizes all the information into five modules: genomic variants, DNA methylation levels, HLA types, metabolite levels, and a large-scale association visualization tool based on the PheWeb platform. This is the largest multiomics dataset of the Taiwanese population currently available and should significantly benefit biomedical research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1090W: The *All of Us* Researcher Academy Internship Program: An Experiential Learning Opportunity Designed to Support Research Skills Development of Diverse Early-Stage Researchers

Authors:

S. Hood, H. Sow, J. Carter, S. Brian, T-R. Hawkins, C. Richardson, I. Adewumi, S. Cates, B. Walker, J. Uhrig, M. Lewis; RTI Intl., Research Triangle Park, NC

Abstract:

Background: The *All of Us* Researcher Academy Internship Program is a 3-month experiential learning opportunity that provides mentored training to graduate and undergraduate students on accessing and analyzing data from the *All of Us* Research Program to understand and address complex health issues. Interns are paired with mentors who use the *All of Us* Researcher Workbench (RW), a cloud-based data analysis platform. Interns also participate in networking activities, weekly internship supervisor meetings, an RTI office visit, and a virtual presentation to share their project results. The inaugural internship cohort (Cohort 1) took place in summer 2023. The internship included six interns (five undergraduates and one graduate) who were full-time students enrolled at an Historically Black College and University (HBCU). Topics studied by Cohort 1 interns include sickle cell disease, breast cancer, diabetes, sleep disorders, asthma, and mental health. The second internship cohort (Cohort 2) will take place in summer 2024 and will include seven interns (six PhD students and one MPH student) from HBCUs.
Methods: Cohort 1 interns were invited to complete an online post-internship evaluation survey. The survey consisted of Likert-type scales, ranging from 1 (lowest) to 4 (highest), and included open-ended response fields for qualitative feedback. The survey inquired about the

following domains: Research Skills Development, Overall Internship Experience, Researcher Workbench Experience, Mentorship Experience, and Internship Resources.
Results: All interns rated their overall internship experience as extremely satisfactory (mean=4.0). Interns' experience using the RW was positive (mean= 3.7). Most interns agreed that they developed their research skills (mean=3.7). All interns rated their mentorship experience as extremely satisfactory (mean=4.0). The most highly rated internship resources included mentors (mean=4.0), networking opportunities (mean=3.0), internship supervisor meetings (mean=3.0), and the virtual presentation opportunity (mean=3.0). Qualitative feedback was positive, and suggested that interns felt supported, and the internship experience increased their sense of pride and confidence as researchers.
Conclusion: The Cohort 1 internship survey results demonstrate the importance and impact of inclusivity in applied research opportunities. The *All of Us* Researcher Academy Internship Program serves as an exemplary model for future experiential learning opportunities to support early-stage researchers' utilization of diverse datasets, such as *All of Us*, for studying a wide range of topics, including genetics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1091W: The Alzheimer's Disease Sequencing Project Follow Up Study ADSP-FUS: *APOE* genotype status and demographic characteristics across African and African American datasets.

Authors:

P. Mena[1], A. Zaman[2], K. M. faber[3], L. Adams[2], J. Inciute[2], P. Whitehead[2], T. Foroud[4], D. Reyes-Dumeyer[5], A. Kuzma[6], H. Issen[7], A. Naj[6], E. Martin[8], C. Dalgard[9], G. Schellenberg[10], L-S. Wang[11], R. Mayeux[12], B. Vardarajan[13], J. Vance[2], M. Cuccaro[14], B. Kunkle[2], M. Pericak-Vance[15]; [1]Univ. of Miami Hussman Inst., Miami, FL, [2]Univ. of Miami, Miami, FL, [3]Indiana Univ. Sch. of Med., Indianapolis, IN, [4]Indiana Univ Sch Med, Indianapolis, IN, [5]Columbia Univ., New York City, NY, [6]Univ. of Pennsylvania, Philadelphia, PA, [7]Univ. of Pennsylvania, Glenside, PA, [8]Univ. of Miami Miller Sch. of Med., Miami, FL, [9]USUHS - TAGC, Bethesda, MD, [10]Univ Pennsylvania Sch Med, Philadelphia, PA, [11]UPENN, Philadelphia, PA, [12]Columbia Univ, New York, NY, [13]Columbia Univ., New York, NY, [14]John P. Hussman Inst. for Human Genomics, Miami, FL, [15]Univ. of Miami Miller Sch. of Med, Miami, FL

Abstract:

Background: The ADSP is a National Institute on Aging (NIA) initiative focused on identifying genetic risk and protective variants for Alzheimer Disease (AD). Initial phases were predominantly non-Hispanic Whites of European Ancestry (NHW-EA). The ADSP expanded the population diversity in the Follow Up Study (ADSP-FUS), focusing on whole genome sequencing (WGS) of non-European populations including non-Hispanic Black with African Ancestry (NHB-AA), Hispanic/Latino (HL), and Asian populations. Support for these efforts include newly funded initiatives such as The DAWN Project, focused on recruitment of African, African-American (AA) and Hispanic American populations, and the Asian Cohort for Alzheimer's Disease (ACAD).
Methods: ADSP cohorts consist of studies of AD, dementia, and age-related conditions. Clinical classifications are assigned based on standard criteria and derived from clinical measures and history, as well as additional neuropathologic data. In addition to production of WGS, APOE genotyping is available for all ADSP samples.
Results:The ADSP currently consists of 40 cohorts comprised of ~36,300 individuals, with plans to sequence >110,000 individuals from diverse ancestries. Genotyping, sequencing, and clinical adjudication has been performed on 5,296 Blacks including 964 Africans (66% female (F), 8.5% AD cases, 17.5% ADRD cases, 74% controls) and 4,332 AA (77% F, 32.7% AD cases, 14.2% ADRD cases, 53.1% controls). African AD and ADRD cases mean age of onsets are 76.1 and 74.5 respectively, while controls mean age is 81.4. APOE genotype proportions are similar across African and AA ancestry. For example, 39.1% of African participants are APOE ε3/ε4 carriers vs 45.1% of AA, and 3.9% of African participants are ε4/ε4 carriers vs. 7.1 % of AA participants.
Conclusions:The growth of the ADSP-FUS 2.0, and its expansion of diverse ancestral genetic data, is central to the ADSP. The data produced by the FUS, all of which is available to qualified investigators via NIAGADS, will be integrated with ADSP programs focused on phenotype harmonization, association analyses, functional genomics, and machine learning. In concert with these programs, the ADSP-FUS 2.0 will accelerate the identification of AD genetic factors across all populations with the target of developing new treatments that are globally effective.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1092W: The BIG initiative's contributions to inclusive genomics

Authors:

F. Marsico[1], S. Buonaiuto[1], A. Mohammed[2], L. Chinthala[2], R. Rooney[3], T. Finkel[4], R. Williams[1], R. L. Davis[2], C. Brown[1,3], V. Colonna[5,3]; [1]Dept of Genetics, Genomics and Informatics, UTHSC, Memphis, TN, [2]Ctr. for BioMed. Informatics, UTHSC, Memphis, TN, [3]Dept of Pediatrics, Genetics

Div., UTHSC, Memphis, TN, [4]Dept of Pediatrics, Div. of Rheumatology, UTHSC, Memphis, TN, [5]Dept of Genetics, Genomics and Informatics, UTHSC, MEMPHIS, TN

Abstract:

Until recently, most genetic data available for human research has predominantly been of European origin. The Biorepository and Integrative Genomics (BIG) initiative in Tennessee is a large state-wide biorepository linked to electronic health records (EHR) data that represents a genetically diverse and ethnically rich study cohort, making it a valuable resource for inclusive genomics. The BIG initiative has enrolled over 40k participants (97% of age <18 years) with EHR, collected 16k biosamples, and generated sequence data from 13k genomes. Participants have been diagnosed with over 8k clinical phenotypes (ICD9-10), broadly categorized into 1,337 PheCode groups. Among sequenced participants, the most prevalent phenotype category is respiratory diseases, which is consistent with the age of the participants.We discovered and phased 6.8 million high-confidence variable sites through exome sequencing and genotype by sequencing (average marker distance 15 and 3.5k bp in coding and intergenic regions, respectively). The majority of variants are rare and non-coding variants account for 69% of the total. Among coding variants, 2% are predicted to have a high impact on the gene or protein function.
Global and local ancestry deconvolution revealed that approximately 50% of participants have non-European origins, including 20% African and 30% with admixed origins, with multiple admixture patterns. These figures mirror self-reported ancestry (41.8% as Black or African American, 2% as Multiple). Thus, projecting to all participants to the BIG initiative, this suggests that approximately 19k are likely to be of non-European or admixed origin. Analyzing the ancestral segment length in admixed individuals through an Approximate Bayesian Computation approach, we estimated that most of the admixture events occurred between 12 and 13 generations ago. Finally, we found comparable genome sharing identical by descent up to first cousins in pairs of participants of the same and different self-reported races. This suggests that, while self-reported race is a valuable indicator in studying environmentally-based diseases, it might be a confounding factor in the analysis of genetically based diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1093W: The Ensembl Variant Effect Predictor: enhanced annotation for human variant interpretation

Authors:

O. Austine-Orimoloye, S. N. Hossain, D. Lemos, N. Saraiva-Agostinho, L. Surapaneni, J. Allen, S. E. Hunt, M. A. Freeberg; EMBL-EBI, Wellcome Genome Campus, Hinxton, United Kingdom

Abstract:

With the increasing use of genome and exome sequencing in rare disease research and diagnosis, new resources have been produced to enable improved variant interpretation. Despite these advancements, variant interpretation remains challenging, necessitating tools to aggregate and enable effective application of these data. The Ensembl Variant Effect Predictor (VEP) is a robust toolset for annotation, interpretation and filtering of genetic variants, enabling access to the latest resources for accelerating variant interpretation.
Ensembl VEP predicts variant molecular consequence and provides easy access to a wide range of tools and reference data including ClinVar assertions for pathogenicity and gnomAD population frequencies. Recently, multiplexed assays of variant effect (MAVEs) have been integrated to report molecular and cellular phenotypes of potential variants within genomic regions. MAVE scores and related publications for assays in MaveDB can be accessed through Ensembl VEP interfaces.
To further aid in identifying variants linked to human phenotypes, Ensembl VEP reports when a variant is found in the Geno2MP database, a catalogue of rare variants observed in individuals with Mendelian diseases. Additionally, phenotype information associated with orthologous genes from model organisms is available to enhance variant interpretation. Ensembl VEP can also report variants for which a paralogous protein location has a ClinVar pathogenic variant assertion. Given that prediction of likely deleterious alleles is crucial when no other information about the locus is available, Ensembl VEP has been extended to integrate new scores including AlphaMissense from Google DeepMind and VARITY that score the likely pathogenicity of missense variants. Enformer scores also provide predictions of variant impact on gene expression. SpliceVault uses GTEx data to predict splicing effects such as cryptic donor or acceptor sites or exon skipping.
Ensembl VEP offers a wide array of analysis options, interface choices and filtering tools, thereby facilitating comprehensive variant annotation and expediting variant interpretation. As this is an exciting time in the field of variant interpretation, Ensembl VEP provides easy access to a wide range of consolidated data resources, simplifying the process and significantly aiding in the interpretation of genetic variants.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1094W: The impact of a research resource: A 5-year review of peer-reviewed publications utilizing samples from the NIGMS Human Genetic Cell Repository

Authors:

M. Mitchell, J. Schulz, N. Turan; Coriell Inst. for Med. Res., Camden, NJ

Abstract:

The NIGMS Human Genetic Cell Repository (HGCR) was established by the NIGMS at the Coriell Institute for Medical Research in 1972 and it stores, processes, and distributes cell lines and DNA samples to researchers around the world. The HGCR collection currently comprises more than 11,900 fibroblast and lymphoblastoid cell lines, over 6,000 DNA samples, and 150 induced pluripotent stem cell lines. The primary goal of the HGCR is to stimulate and facilitate biomedical research, teaching, and diagnosis and prevention of human genetic diseases by establishing, maintaining, and distributing a repository of high quality, uncontaminated, well-characterized, and clinically documented cell lines and DNA from individuals with genetic disorders and chromosomal abnormalities, and from individuals from various geographical locations for studies on human diversity. To better understand the specific uses of these specimens, we surveyed the peer-reviewed literature that cited the use of samples from the HGCR in studies published from 2019 until the present. Publications were identified for review by keyword search from the Google Scholar index and filtered for inclusion based on an explicit citation of an HGCR ID. Publication metadata that was analyzed was gathered from the article text, article keywords, and the content of the abstract. We identified over 1,500 publications, referencing HGCR samples more than 5,400 times. These publications originated from research groups in over 60 countries, representing both the academic and commercial sectors, and included research applications that ranged from technological improvements to basic science investigations. Over time, these analyses will provide metrics on the value and impact of this resource and highlight evolving trends in research that utilize NIGMS HGCR samples. Understanding the research impact of its samples provides an opportunity to improve the HGCR and the diversity of samples in the collection with the aim of promoting impactful research and stimulating the advancement of scientific knowledge.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1095W: The impact on clinical success from the 23andMe cohort

Authors:

X. Wang[1], S. Karagounis[1], S. Shringarpure[1], Q. Su[1], S. Pitts[2], A. Auton[1]; [1]23andMe, Sunnyvale, CA, [2]23andMe, South San Francisco, CA

Abstract:

90% of therapeutic programs that enter clinical trials ultimately fail. Human genetic variation provides a set of “natural experiments” that can inform successful strategies for therapeutic discovery. Previous work has estimated that drug targets with human genetics supported mechanisms have a 2-3x increased likelihood of succeeding in the clinic compared to those without. 23andMe, Inc. is a direct-to-consumer genetics company that has created a human genetics dataset approximately an order of magnitude larger in sample size than current publically available cohorts. As of 2024, 23andMe has approximately 15 million individuals with genotype and phenotype data, of which ~80% consent to participation in research. In this work, we explore how both the scale of the genetic data and improved methods to link genetic associations to putative causal genes impact the prediction of clinical success. Comparing the total number of target-indication pairs that have reached at least phase I that are also supported by genetic evidence, the number of target-indication pairs with support from 23andMe is 60% larger than that with support from all GWAS datasets in the public domain. Including 23andMe genetic evidence approximately doubles the number of target-indication pairs in the clinic that are supported by human genetic evidence. Furthermore, we show that genetic associations derived from entirely self-reported phenotypes are 2-3x enriched for clinical success, just as for clinically derived phenotypes. In contrast to conclusions from the recent publication of Minikel et al., we found that minor allele frequencies and effect sizes from GWAS influence the relative success estimates for program approvals, and that drug programs supported by rare and large effect associations have greater (3-4x) likelihood to be approved compared to common variant associations with small effects. Finally, improved gene mapping to identify the likely causal genes underlying genetic associations can result in up to 4-5x enrichment for trial success. With the increased power and scale of the 23andMe genetic dataset, we identify an expansive set of opportunities that may be pursued in the clinic, emphasizing the importance of cohort size and gene mapping confidence in deriving clinical value.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1096W: Unraveling the genetic predisposition to non-syndromic Biliary atresia through genomic investigations

Authors:

Q. Lin[1], C. S. Tang[1,2], V. C. Lui[1,2], P. C. Sham[3,2], P. K. Tam[4]; [1]Dept. of Surgery, Li Ka Shing Faculty of Med., The Univ. of Hong Kong, Hong Kong, China, [2]Dr Li Dak-Sum Res. Ctr., The Univ. of Hong Kong-Karolinska Inst.d Collaboration in Regenerative Med., Hong Kong, China, [3]Dept. of Psychiatry, Li Ka Shing Faculty of Med., The Univ. of Hong Kong, Hong Kong, China, [4]Faculty of Med., Macau Univ. of Sci. and Technology, Macao, China

Abstract:

Biliary atresia (BA) is a rare congenital cholangiopathy and is the most common reason for liver transplantation in the pediatric population. Elucidating the genetic predisposition of BA may help decode the underlying molecular mechanisms to improve disease outcomes. To uncover novel disease-susceptibility genetic factors, rare damaging mutations were identified in a dataset of whole exome/genome sequencing of 188 BA patients and 324 unaffected parents. Through bioinformatic analyses, four gene sets were found to be associated with the development of BA: (i) cilia, (ii) extracellular matrix (ECM), (iii) NOTCH, and (iv) transforming growth factor-beta (TGFβ) signaling pathways. *THBS1* emerged as the central hub of candidate genes in the protein-protein interaction (PPI) network analysis. Additionally, the integration of transcriptome data identified several candidate genes related to liver diseases, including two known BA-associated genes, *EFEMP1* and *KRT19*, and a novel candidate, *MAT1A*. Gene-based burden tests on the rare damaging variants highlighted the involvement of the catabolic process in the development of BA, which was relevant to the pathomechanism of MAT1A. Cell type enrichment and gene expression deconvolution analyses using single-cell RNA sequencing data further highlighted the significance of hepatic stellate cells (HSCs) and their interplay with ECM in BA. Overall, this study utilized various genomic methodologies to uncover novel disease-susceptibility genes, biological processes, and pathways involved in the development or progression of BA.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1097W: Updates from the frontline of international genomic data sharing

Authors:

A. Page[1], H. Rehm[2], P. Goodhand[3]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Broad Inst., Cambridge, MA, [3]Global Alliance for Genomics And Hlth., Toronto, ON, Canada

Abstract:

The Global Alliance for Genomics and Health (GA4GH) continues to lead the way in fostering international collaboration to accelerate progress in genomics research and its application to health. Over the past year, GA4GH has made significant strides in several key areas, advancing the global genomic data ecosystem. This presentation will provide an overview of the most impactful updates and achievements from the past year. Key Highlights include: **(i) Enhanced Data Sharing Frameworks**: GA4GH has introduced new and improved policy frameworks for data sharing, emphasizing privacy, security, and interoperability. These updates facilitate more seamless data exchange across international borders, enhancing the global research community's ability to collaborate effectively. **(ii) Updates to the GA4GH Roadmap**: The organization has rolled out updates to its suite of policy tools and standards, including GA4GH Passports and AAI, refget 2.0, and Beacon 2.0. These additions support more efficient genomic data access, sharing, and analysis, benefiting both research and clinical applications. **(III) Collaborative Projects and Case Studies**: This year saw the initiation and completion of several high-impact collaborative projects. The presentation will highlight key case studies demonstrating the successful application of GA4GH standards in real-world research settings (e.g., implementations within NIH, Genomics England, Australian Genomics, Pan-Canadian Genomics Library, Qatar Genome and more), showcasing the tangible benefits of global collaboration. The past year has been marked by significant progress and innovation within GA4GH, with a particular emphasis on implementation. These updates not only enhance the capability of researchers worldwide but also ensure that the benefits of genomics research are shared equitably. This presentation will provide a comprehensive overview of these developments, demonstrating GA4GH's ongoing commitment to advancing global health through genomics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1098W: Validation of the variant classification rules for ornithine transcarbamylase, N-acetylglutamate synthase, and carbamylphosphate synthetase 1

Authors:

L. Caldovic[1,2], N. Ah Mew[1], D. Argueta[3], R. Ayyadurai[4], K. Azizi-Namini[2], J. Bedoyan[5], G. Berry[6], L. Burrage[7], A. Cortes Fernandez[8], M. Crenshaw[8], A. Feigenbaum[9], E. Groopman[1], A. Gropman[10], J. Häberle[11], R. Haider[3], C. Harding[12], L. Konczal[13], U. Lichter-Konecki[14], G. Lipshutz[15], S. McCandless[16], L. Medeiros[17], S. Melby[2], H. Morizono[1,2], M. Naybor[1], Y. Nguimbous[18], K. Oishi[19], M. Osundiji[20], V. Rubio[21], S. Sako[19], K. Simpson[1], E. Spector[22], M. St. Jacques[2], Y. Sudo[23], J. Trainor[2], T. Tsunogai[19], A. Upadhye[24], J. Volpi[3], M. Warrier[25], M. Weaver[26], D. Wong[15], D. Zastrow[27], A. Thomas-Wilson[28]; [1]Children's Natl. Hosp., Washington, DC, [2]The George Washington Univ., Washington, DC, [3]Baylor Coll. of Med., Houston, TX, [4]Emory Univ., Atlanta, GA, [5]Univ Hosp. Cleveland Med Ctr., Cleveland, OH, [6]Boston Children s Hosp, Boston, MA, [7]Baylor Coll. Med., Houston, TX, [8]Univ. of Colorado, Aurora, CO, [9]Univ. of California San Diego, San Diego, CA, [10]Children s Natl. Hosp., Washington, DC, [11]Univ. Children's Hosp. Zurich, Zurich, Switzerland, [12]Oregon Hlth.& Sci. Univ., Portland, OR, [13]Univ. Hosp. of Cleveland Med. Ctr., Cleveland, OH, [14]Children s Hosp. of Pittsburgh, Pittsburgh, PA, [15]Univ. of California Los Angeles, Los Angeles, CA, [16]Univ. of Colorado Sch. of Med., Aurora, CO, [17]Hosp. de Clínicas de Porto Alegre, Porto Alegre, Brazil, [18]Howard Univ., Washington, DC, [19]The Jikei Univ. Sch. of Med., Tokyo, Japan, [20]Mayo Clinic, Acottsdale, AZ, [21]Inst. for Biomedicine of Valencia-CSIC and CIBERER-ISCIII Group, Valencia, Spain, [22]Univ. of CO Sch. of Med., Highlands Ranch, CO, [23]Fujita Hlth.Univ., Toyoake, Japan, [24]Rensselaer Polytechnic Inst., Troy, NY, [25]LabCorp, Durham, NC, [26]ACMG, Bethesda, MD, [27]Invitae, San Bruno, CA, [28]New York Genome Ctr., Saratoga Springs, UT

Abstract:

The ClinGen Urea Cycle Disorders Variant Curation Expert Panel (UCD VCEP) was established in 2021 to evaluate evidence and classify genomic variants identified in genes encoding 6 enzymes and 2 transporters of the urea cycle. Since more than 5000 variants reported in these genes have been deposited in ClinVar, the UCD VCEP prioritized the development of gene-specific variant classification rules based on disease severity and prevalence. Ornithine transcarbamylase (*OTC*), N-acetylglutamate synthase (*NAGS*) and carbamylphosphate synthetase 1 (*CPS1*) deficiencies typically present with the earliest and most severe symptoms. *OTC* deficiency, an X-linked condition, generally accounts for about half of all individuals with UCDs. Autosomal recessive UCDs may be more frequent than *OTC* deficiency in populations with high consanguinity rates. Variant interpretation rules have been specified for *OTC*, *NAGS,* and *CPS1* deficiencies. Criteria and strength of evidence related to variant type, allele frequency, functional assay(s), and phenotype were adjusted for each disease. Allele frequency thresholds were defined based on the prevalence of each disease and frequency of the most common pathogenic variants in population databases (PM2, BA1, BS1 criteria). *NAGS* and *CPS1*-specific splicing tables provide supporting evidence that exon skipping or in-frame deletion of or duplicating one or more exons results in the loss of protein function

via frameshift and nonsense-mediated decay or loss of critical residues (PVS1 criterion). A functional assay based on the growth rate of yeast cells expressing human OTC variants allowed the determination of the odds of their pathogenicity (PS3/BS3_Moderate and _Supporting). The VCEP chose REVEL scores for the classification of missense variants and modified the strength of PP3 and BP4 criteria according to the latest ClinGen recommendations. A point scheme based on biochemical testing, enzyme assays in patient-derived cells, biochemical and genetic testing to exclude alternate causes, and for *NAGS* deficiency, response to monotherapy with carglumic acid, was established to determine the strength of evidence for variant pathogenicity provided by the patients' phenotype (PP4 criterion). The strength of PP4 criteria can be increased to PP4_Moderate when multiple lines of phenotypic evidence are provided for a proband. These variant classification rules will be validated by the curation of
select *OTC*, *NAGS,* and *CPS1* variants. The variants have been selected to assess gene-specific thresholds for various criteria as well as different combinations of classification criteria.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Wednesday Session

Board 1099W: VarChat: the genAI-based genomic assistant available as an open-access platform

Authors:

S. Berardelli[1,2], F. De Paoli[2], E. Rizzo[2], I. Limongelli[2], **S. Zucca**[2]; [1]Univ. of Pavia, Pavia, Italy, [2]enGenome srl, Pavia, Italy

Abstract:

Background: In the current genomics era, the scientific community is experiencing rapid growth in the volume of publications. While this abundance of data offers invaluable insights, it also places a pressing responsibility on genetic professionals and researchers to stay informed about the latest findings. One major challenge in genomic variant interpretation is the identification and assimilation of the most relevant scientific papers, crucial for accelerating the transition from clinical assessment to actionable reporting. Computer-aided literature searches and summarization tools are key in addressing this challenge, helping to synthesize complex genomic findings into clear, concise summaries, transforming extensive genomic datasets into insights that are relevant to clinical practice.
Methods: We developed VarChat, an innovative tool that leverages generative AI (https://doi.org/10.1093/bioinformatics/btae183), designed to find and summarize the fragmented scientific literature associated with genes and genomic variants. VarChat is an open

access platform, accessible at varchat.engenome.com. It uses a customized version of the BM25 ranking algorithm to determine the relevance of publications, which primarily relies on the classic term frequency. More weight is given to most recent papers that cite the variant in the abstract or in the full text, while papers that report the variant only in the supplementaries are penalized.
Results: VarChat enables users to search for genomic variants through transcript- or gene-based HGVS coordinates for both coding and protein sequences, genomic coordinates, or dbSNP IDs. Searches can also be conducted using official gene symbols or their synonyms, to get a comprehensive summary about gene function and its association with diseases. It uses both information retrieved from the scientific literature and its own internal knowledge to provide insightful summaries. Every piece of information is referenced, so that users are in full control of the information. Additionally, when a variant is listed in ClinVar, VarChat provides details about the related submission and includes a direct link to the database entry. The tool also supports translation of answers in different languages. VarChat's performance was evaluated on a dataset of 200 between germline and somatic variants. We performed both an automated analysis and a manual curation by a geneticist expert to assess the accuracy and the reliability of each answer.
Conclusion: The VarChat platform is open and accessible at varchat.engenome.com to support the genomic community.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2002W: A Case Series of Pancreatitis in Patients with Poorly Controlled Classic Homocystinuria

Authors:

R. Aljaberi[1], L. Romo[2], H. Levy[2], C. Ficicioglu[3], T. Mellor[3], M. Steffensen[3], B. Hainline[4], M. Gambello[1], H. Li[1]; [1]Dept. of Human Genetics, Emory Univ., Sch. of Med., Atlanta, GA, [2]Div. of Genetics & Genomics, Dept. of Med., Boston Children's Hosp. and Dept. of Pediatrics, Harvard Med. Sch., Boston, MA, [3]Div. of Human Genetics, Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med., Indianapolis, IN

Abstract:

Background: Classic homocystinuria (HCU) is a rare autosomal recessive disorder caused by cystathionine beta-synthase deficiency, leading to elevated levels of homocysteine (Hcy) and methionine (Met). Clinical manifestations typically involve the eyes, skeleton, central nervous system, and vascular system. Thromboembolic events, including stroke and deep vein

thrombosis, are well-recognized complications. Although acute pancreatitis (AP) has been reported in only a few cases in the literature, we describe four individuals with classic HCU and poor compliance who developed AP with various outcomes. **Case description:** We present four cases of classic HCU diagnosed through newborn screening and confirmed molecularly, all of whom were pyridoxine non-responders with significantly elevated total homocysteine (tHcy) levels at 100s-300s umol/L. Case 1, a 15-year-old male, presented with acute abdominal pain and necrotizing pancreatitis complicated by insulin-dependent diabetes. Case 2, a 24-year-old male who became non-compliant in his teenage years had a history of pulmonary embolism, spinal compression fractures, and leukoencephalopathy. He developed recurrent pancreatitis episodes four times, complicated by insulin-dependent diabetes, and ultimately passed away due to respiratory failure. Case 3, a 14-year-old female without known HCU complications, had multiple pancreatitis episodes since the age of 8, leading to chronic pancreatitis and diabetes. Case 4, a 27-year-old male with Marfanoid habitus, ADHD, and adult-onset psychosis, had recurrent pancreatitis three times starting at age 15, complicated by pancreatic insufficiency and diabetes. Only this patient's abdominal CT showed a splenic thrombus with an atrophied pancreas. **Conclusion:** Our newly reported four cases, along with six published cases of pancreatitis in HCU, show that pancreatitis is linked to very high levels of homocysteine, indicating that poorly controlled HCU increases the risk of acute pancreatitis. The etiology may be attributed to a combination of microthrombosis, inflammation, and dysregulation of the glutathione pathway leading to glutathione depletion. Early recognition of pancreatitis in patients with a known diagnosis of HCU is critical and can affect medical management, such as avoiding high-protein TPN. We also recommend considering HCU as a differential diagnosis in idiopathic pancreatitis cases and including total homocysteine in standard laboratory investigations to identify missed cases.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2003W: A founder variant in tubulin binding cofactor B is associated with autism spectrum and hereditary spastic paraparesis

Authors:

K. Weiss[1], S. Bratman-Morag[1], C. Itzkovich[2], A. Kurolap[3], M. Shohat[4], A. Durr[5], J-M. de Sainte Agathe[6], A. Koifman[7], A. Alkelai[8], A. Shuldiner[9], A. Mory[10], T. Harel[11], T. Paperna[12], H. Baris Feldman[3], D. Kornitzer[2], A. Salzberg[2]; [1]Rambam Hlth.Care Campus, Haifa, Israel, [2]Technion - Israel Inst. of Technology, Haifa, Israel, [3]Tel Aviv Sourasky Med. Ctr., Tel Aviv, Israel, [4]Maccabi Hlth.Services, Petach Tikva, Israel, [5]ICM UMRS1127 SORBONNE UNIVERSITE, Paris, France, [6]Sorbonne Université, Paris, France, [7]Samson Assuta Ashdod Univ. Med. Ctr., Ashdod,

Israel, [8]Regeneron Genetics Ctr., Tarrytown, NY, [9]Regeneron Genetics Ctr., Miami, FL, [10]Tel Aviv Sourasky Med. center, Tel Aviv, Israel, [11]Hadassah-Hebrew Univ. Med. Ctr., Jerusalem, Israel, [12]Rambam Med Ctr, Haifa, Israel

Abstract:

Hereditary spastic paraparesis (HSP) is a group of Mendelian disorders related to length-dependent axonal degeneration. Microtubule (MT) dysfunction is one of the molecular mechanisms for HSP. MT are dynamic structures controlling axonal outgrowth and retraction, their structural diversity is attributed to variability in its subunit, the αβ-tubulin heterodimer. *TBCB* encodes a tubulin-binding co-factor, not previously associated with disease, that regulates α-tubulin dissociation from the MT thus negatively regulating axonal growth. We describe a founder pathogenic variant in *TBCB* as a novel cause of complex HSP. Through GeneMatching tools and automated data sharing, we identified 9 patients from 7 unrelated Ashkenazi Jewish families harboring the same c.589T>A; p.(Tyr197Asn) variant with a minor allele frequency of 0.65% in this population. The patients show global developmental delay, autism spectrum, and late childhood onset spastic paraparesis. Western blot analysis revealed reduced levels of TBCB in fibroblasts of affected individuals, with normal localization on immunofluorescence. A corresponding mutation in the Saccharomyces cerevisiae orthologue *ALF1*, resulted in benomyl sensitivity resembling the expected knock-out phenotype. Furthermore, *dTBCB* knockdown in neuronal cells results in a shorter lifespan and impaired locomotor activity in Drosophila. CRISPR-Cas9 generated homologous mutation in Drosophila produced semi-lethality and reduced climbing ability. In summary, we identified a novel form of HSP secondary to a missense *TBCB* variant with a surprisingly high carrier rate in the Ashkenazi Jewish population warranting its addition to preconception carrier screening. Our results support that TBCB has an important role in CNS development and potentially in axonal function in humans.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2004W: A homozygous frameshift mutation in *TRDN* causes a Cardiac Arrhythmia Syndrome in a Malian family

Authors:

M. Diakité[1,2], O. Sanni[3], S. Bamba[4], M. Goïta[4], O. Traoré[4], A. Maiga[2], B. Dembélé[1], W. Ji[5], S. Diarra[4], M. Khokha[5], S. Lakhani[5], S. Coulibaly[1,4], G. Landoure[4,6], O. Samassekou[2]; [1]Service de Cardiologie, Ctr. Hosp.ier Univ.ire "Point G", Bamako, Mali, [2]USTT-B, Bamako, Mali, [3]Faculté de Médecine et d'odontostomatologie, Bamako, Mali, [4]Faculté de Médecine et

d'odontostomatologie, USTT-B, Bamako, Mali, [5]Yale Univ., New Haven, CT, [6]Service de Neurologie, Ctr. Hosp.ier Univ.ire "Point G", Bamako, Mali

Abstract:

Introduction: Triadin is a protein expressed in cardiac and skeletal muscle that has an essential role in the structure and functional regulation of calcium release units and excitation-contraction coupling. Genetic variations in the triadin gene are reported to cause certain malignant arrhythmic syndromes including Catecholaminergic polymorphic ventricular tachycardia (CPVT), particularly in younger children. CPVT is clinically characterized by episodic syncope, triggered by physical activities and emotional stress. Although cases of CPVT are described in other populations, data in Africa are scarce. Here report two siblings from a consanguineous family with clinical and electrical characteristics of CPVT, and in which DNA sequencing revealed a pathogenic homozygous frameshift variant in the TRDN gene.
Material and methods: Two affected siblings and their relatives were enrolled in our protocol after giving full consent. Family and medical history was taken to draw pedigree. A thorough physical examination was performed in all available individuals in the family by cardiologists and medical geneticists. Additionally, Electrocardiography (ECG), heart Doppler Ultrasound and stress test were done. WES was performed in the probands, their parents and unaffected siblings, and prediction tools used to assess deleteriousness of putative variants.
Result: A 17-year-old boy, his 11-year-old sister, and their relatives were referred to our clinic for repetitive syncope. They were born to healthy consanguineous parents. The first episode occurred at the ages of 7 and 11, respectively. Physical examination was unremarkable. ECG showed electrical patterns of bradycardia in both, while the stress test was positive. Echocardiography was uneventful. WES identified a homozygous frameshift in variant in the proband (Gln205GlufsTer3), leading to a premature stop codon. This variant was found in a heterozygous state in both parents and is classified as pathogenic by Clinvar and the ACMG criteria.
Conclusion: Our study reports a fatal cardiac arrhythmic syndrome linked to a homozygous variant in TRDN in a Malian family. To the best of our knowledge, this is the first report of a genetically confirmed case in Africa, expanding the genetic spectrum of this condition in Africa.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2005W: A mechanism underlying anophthalmia by MAB21L1^{R51W} de novo mutation

Authors:

Y. Sun, L. Peng, Y. Yang; Tongji Univ., Shanghai, China

Abstract:

Anophthalmia or microphthalmia(A/M) is a rare genetic birth defect with an absent or very small malformed orbital globe. Its prevalence is 1-3 in 10,000 live births. Here we identified a causative novel variant in *MAB21L1*(OMIM: *601280, c.151C>T, p.R51W) in an anophthalmia family via whole-exome sequencing. The MAB21L1^{R51W} mutant generated using CRISPR/Cas9-mediated gene editing in zebrafish showed a host of ocular defects, including microphthalmia, small and disorganized lenses, and cornea dysgenesis. Meanwhile, MAB21L1 null mutant produced a similar striking eyeless phenotype in zebrafish, suggesting a loss-of-function of MAB21L1^{R51W} mutation. Moreover, injection of MAB21L1 mRNA in MAB21L1^{R51W} mutant zebrafish significantly reduced the proportion of A/M zebrafish, which further verified the crucial role of MAB21L1 in eye development.
In mechanism, we find that MAB21L1 promotes the expression of PAX6, a factor being essential for eye development in both human lens epithelial cells and corneal epithelial cells. Specifically, docking and coIP results indicate that MAB21L1 can bind with transcription factor MEIS2 to form a complex that can enter the nucleus to promote *PAX6* transcription. Bioinformatics analysis indicates that the MAB21L1^{R51W} mutation resulted in a disruption of the protein salt bridge, leading to a decrease of the protein stability. We further confirm that the MAB21L1^{R51W} mutant protein is more susceptible to degradation compared with wild MAB21L1 by using cycloheximide assay. Finally, we suppose that the unstable MAB21L1^{R51W} protein may be degraded via the ubiquitin-proteasome system (UPS) rather than the lysosomal, which is fail to bind with MEIS2 to initiate the expression of PAX6, finally resulting in eye development retardation.
In conclusion, we identify a de novo mutation of MAB21L1^{R51W} that causes anophthalmia by a mechanism, in which the mutant protein fails to associate with MEIS2 to maintain PAX6 expression, leading to the development disorder of the eyeball.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2006W: A Novel *De Novo KDM3B* Mutation in a Male Patient with Evolving Diets-Jongmans Syndrome Features

Authors:

S. Jo[1,2], C. Lee[1,2], L. White[3], R. Hopkin[2,3], G. Hijazi[3], E. Sperry[3], A. Shillington[2,3], A. Slavotinek[2,3], Y. Wu[2,3]; [1]Univ. of Cincinnati, Cincinnati, OH, [2]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [3]Cincinnati Children's Hosp., Cincinnati, OH

Abstract:

The *KDM3B* gene encodes a JmjC domain-containing protein, shown to function as a histone lysine demethylase, specifically for histone H3K9, and involved with transcriptional regulation. Heterozygous pathogenic variants in *KDM3B* are associated with Diets-Jongmans syndrome (DIJOS), an autosomal dominant genetic disorder first described in 2019. Developmental delay or mild to moderate intellectual disability was observed in all affected individuals old enough for evaluation. Other commonly reported clinical features include distinct facial features, behavioral problems such as ADHD and autism, short stature, and neonatal feeding difficulties. Epilepsy, hearing loss, childhood hypotonia, sleeping difficulties, or sleeping problems have been reported in some DIJOS patients. Heterozygous missense, nonsense, frameshift, and splice site variants have been reported in affected individuals. While most cases reported in the literature involve a *de novo* variant, some patients have inherited the variant from a similarly affected parent. Here, we report the identification of a *de novo* frameshift *KDM3B* mutation in a nine-month-old male with small stature, facial asymmetry, microcephaly, reduced facial movement on one side, and mild gross motor delays. The patient was first evaluated for slow weight gain, short stature, microcephaly, and mild left-sided facial asymmetry at seven months. Prior genetic testing, including microarray, was non-diagnostic. A clinical exome placed at nine months identified a *de novo* likely pathogenic variant in *KDM3B* (c.2446del;p.Ser816lfs). This variant has not been reported in the literature or public databases. At fourteen months, the patient exhibited age-appropriate development and patterns of interacting and socializing with others. At fifteen months, the patient showed additional clinical findings of mild bilateral epicanthal folds, short philtrum, mildly upslanting palpebral fissure, broad nasal bridge, and mildly simplified ears, while his physical and behavioral development remained age-appropriate. To our knowledge, this is the youngest patient with a *de novo KDM3B* variant identified before one year old. He does not have signs of significant neurodevelopmental delays consistently noted in other DIJOS patients. Additionally, this patient presents with unilateral facial asymmetry, which had not been reported in DIJOS patients. This patient's unique and evolving clinical features will significantly contribute to understanding the phenotypic diversity and long-term prognosis of Diet-Jongmans Syndrome.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2007W: A novel in-frame deletion of HNF1B in a family presenting with renal cysts and maturity onset diabetes of the young

Authors:

J. Han[1], J. Park[2]; [1]Daejeon St.Mary's Hosp., Daejeon, Korea, Republic of, [2]Jeonbuk Natl. Univ. Med. Sch. and Hosp., Jeonju, Korea, Republic of

Abstract:

Background: Maturity-onset diabetes of the young comprises a cluster of inherited disorders within non-autoimmune diabetes mellitus (DM), typically emerging during adolescence or young adulthood. We report a novel in-frame deletion of HNF1B in a family with renal cysts and MODY, furthering our understanding of HNF1B-related phenotypes. Methods: We conducted sequential genetic testing to investigate the glucose intolerance, renal cyst, hepatic cyst, and agenesis of the dorsal pancreas observed in the proband. A comprehensive clinical exome sequencing approach using a Celemics G-Mendeliome Clinical Exome Sequencing Panel was employed. Results: Considering the clinical manifestations observed in the proband, the gene panel sequencing identified a heterozygous HNF1B variant, c.36_38del/p.Leu13del, as the best candidate for causing MODY in the proband. The clinical presentation of the patient was consistent with the MODY caused by HNF1B variant, indicative of glucose intolerance, renal cyst, hepatic cyst, and agenesis of the dorsal pancreas. Sanger sequencing confirmed the same HNF1B variant with the phenotype and established the paternally inherited autosomal dominant status of the heterozygous variant in the patient, as well as in his father and sister. Conclusion: The presence of early-onset diabetes, renal cysts, a family history, and nephropathy appearing before or after the diagnosis of DM suggests a diagnosis of HNF1B-MODY5. Early diagnosis is crucial for preventing complications of DM and allows for family screening and pre-conceptional genetic counseling.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2008W: A novel *PPARG* germline variant segregating with familial lymphedema: challenges in disease gene discovery at the single-family level

Authors:

Z. Dai[1,2], B. Haque[1], S. Walker[3,4], A. Deshwar[1,2,5,6], G. Costain[1,2,6]; [1]Program in Genetics and Genome Biology, SickKids Res. Inst., Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]Genomics England, London, United Kingdom, [4]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [5]Program in Dev.al and Stem Cell Biology, SickKids Res. Inst., Toronto, ON, Canada, [6]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, and Dept. of Paediatrics, Univ. of Toronto, Toronto, ON, Canada

Abstract:

The *PPARG* gene encodes for the peroxisome proliferator-activated receptor-gamma of the PPAR family of nuclear receptors. *PPARG* has a critical role in adipogenesis, and dysregulation of *PPARG* has been associated with obesity, diabetes, and cancer. Germline loss-of-function variants in *PPARG* cause familial partial lipodystrophy, but the consequences of constitutive activation of *PPARG* are unknown. We report a novel *PPARG* germline missense variant (p.Thr457Met) in a 14-year-old individual with a personal and family history of lymphedema. The variant was identified by trio genome sequencing and segregated with lymphedema within the family. A previous study of somatic *PPARG* mutations in bladder tumours confirmed experimentally that the *PPARG* variant identified in the patient's family results in gain-of-function gene activity. We are now searching for germline activating variants in *PPARG* in large-scale public and private datasets with genome-wide sequencing data from millions of individuals and studying the impact of constitutive activation of *PPARG* in a non-human animal model system. The results to date underscore the challenges of establishing new disease-gene associations to a clinical standard based on single-family observations.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2009W: A novel*CUL7*variant in a 3-M syndrome Turkish case

Authors:

A. Ersoy Karayumak, O. Hanoglu, C. Onaran, E. Ceylan, A. Guney, B. Geckinli; Marmara Univ. Faculty of Med., Istanbul, Turkey

Abstract:

Background
3-M syndrome (3MS) is an autosomal recessively inherited skeletal dysplasia involving multiple organ systems and is characterized by short stature with distinctive facies. *CUL7, OBSL1* and *CCDC8* genes are accounted for the 3 types of 3MS. So far, 100 affected individuals have been reported in the literature. *CUL7* gene mutations account for more than %75 of the cases and all of them are point mutations.
Methods
DNA was isolated from the patient's peripheral blood and Clinical Exome Sequencing (CES) was performed via next-generation sequencing (Illumina Nextseq 500).
Results
The patient was a 2-months-old male who was consulted to medical genetics clinic for dysmorphism, hypotonia and left ventricle hypertrophy. His parents were from the same small

village. He was born at term with 2330 gr (-2,36 SDS) of weight and 39 cm (-5 SDS) of height and 36 cm (0.79 SDS) of head circumference. He was admitted to the NICU because of respiratory distress. At 1,5-years-old height was 65 cm (-5 SDS), weight was 7 kg (-4 SDS) and head circumference was 47 cm (-0.91). His dysmorphic features were frontal bossing, dolichocephaly, triangular face, pointed and prominent chin, long philtrum, upslanting palpebral fissures, fleshy upturned nose and anteverted nares. He had bilateral transverse palmar crease and an enlarged abdomen. Hearing tests and vision examination were normal. CES analysis revealed a novel homozygous *CUL7* (NM_014780) c.1938dup p.(Thr647Aspfs*33) likely pathogenic variant. The variant was not reported in ClinVar database and in silico tools agree on the disease causing property of the variant. Segregation analysis from the parents are ongoing.

Conclusion

Here, we report a novel variant in *CUL7* gene detected in a Turkish patient. Left ventricular hypertrophy is not reported in literature. More research is needed for phenotype genotype relationship and presence of cardiac pathology in 3MS.

Grants

None

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2010W: A null variant in the voltage-gated calcium channel beta-1 subunit (*CACNB1*) causes autosomal recessive muscular dystrophy

Authors:

N. Assia Batzir[1], D. Yagel[2], R. Jaron[1], R. Greenberg[2], S. Aharoni[3,4], L. Basel Salmon[3,5,6], N. Orenstein[1,3], S. Ben-Shachar[2,3]; [1]Pediatric Genetics Unit, Schneider Children's Med. Ctr. of Israel, Petah-Tikva, Israel, [2]Clalit Genomic Ctr., Clalit Hlth.Care, Ramat-Gan, Israel, [3]Faculty of Med. and Hlth.Sci., Tel-Aviv Univ., Tel-Aviv, Israel, [4]Inst. of Pediatric Neurology, Schneider Children's Med. Ctr. of Israel, Petah-Tikva, Israel, [5]Raphael Recanati Genetics Inst., Rabin Med. Ctr., Petah-Tikva, Israel, [6]Felsenstein Med. Res. Ctr., Petah-Tikva, Israel

Abstract:

Congenital muscular dystrophies (CMD) are a group of genetic disorders characterized by early-onset muscle weakness alongside histological signs of dystrophy. While many genes have been clinically associated with CMDs, in about 40% a genetic etiology cannot be detected.
Here, we report two affected individuals of a large consanguineous Bedouin family, presenting with proximal muscle weakness, myopathic facies and severe muscle wasting with extremely low body weight in both. Individual 1 has evidence of cardiomyopathy with features of

noncompaction of the left ventricle observed on MRI. Both patients had elevated creatine phosphokinase (up to 1900 U/L and 4900 U/L in individuals 1 and 2 respectively) and myopathic EMG. Exome sequencing revealed a homozygous out-of-frame deletion (NM_000723.5:c.124_133del; p.Asp42ArgfsTer37) in *CACNB1* in the affected individuals, but not in healthy family members. No pathogenic variants were detected in the known CMD and other myopathy-related genes. No other variants segregated with the phenotype. The *CACNB1* gene encodes the beta-1 subunit in voltage-gated L-type calcium channels, contributing to the regulation of muscle contraction and overall muscle function. Variants affecting the alpha-1S subunit of the receptor, and its interactors (RYR genes) are known to cause congenital myopathy in human. Previous studies with knockout mouse models for the *Cacnb1* gene indicate that its biallelic malfunction causes defects in excitation-contraction (EC) coupling. Similarly, zebrafish nonsense variants in the gene affected embryonal motility. The overall data implicate that *CACNB1* is a novel gene associated with CMD. Further work is needed to evaluate its prevalence among patients with CMD of unknown etiology, as well as its exact role in CMD pathology.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2012W: A rare case of Knobloch Syndrome in a consanguineous Malian family

Authors:

M. K. Goita[1], S. Bamba[1,2], L. Cisse[3], A. Maiga[1], O. Traore[1], S. Diarra[1,2], O. Samassekou[1], M. Khokha[2], S. Lakhani[2], G. Landoure[1,4]; [1]Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, [2]Pediatric Genomics Discovery Program, Dept. of Pediatrics, Yale Univ., New Haven, CT, [3]Service de Médecine, Hôpital Nianankoro FOMBA de Ségou, Segou, Mali, [4]Service de Neurologie, Ctr. Hosp.ier Univ.ire Point G, Bamako, Mali

Abstract:

Introduction: Knobloch syndrome (KS) is a rare autosomal recessive disorder, classically characterized by high myopia, retinal detachment, and occipital encephalocele, but it is now known to have an increasingly variable phenotype including epilepsy. KS is due to genetic defects in the *COL18A1* gene, which encodes the alpha chain of type XVIII collagen. Although Variants in this gene have been reported worldwide, but to our knowledge none in Africa. Here, we report the first molecularly confirmed case of KS with visual impairment and generalized tonic-clonic seizures, in Mali. Methods: This study was approved by the Ethics committee of the Faculty of Medicine and Dentistry of Bamako, Mali, and written consent/assent was obtained

from participants. After a thorough neurological examination, EEG was performed in the patient. Peripheral blood was collected to extract DNA for genetic analysis, including Exome sequencing and Sanger sequencing. Prediction tools were used to assess deleteriousness. Results: A 13-year-old girl, with her unaffected sister, from a consanguineous Malian family, was seen for generalized tonic-clonic seizures and visual impairment. The disease started at the age of 5 with generalized tonic seizures, gradually evolving to tonic-clonic seizures and visual loss around the age of 10. Physical examination found a normal intelligence, bilateral persistent anisocoria, total visual loss, miscalculation, and left hemiparesis. Neurological examination revealed brisk reflexes in the lower limbs and Babinski's sign in the left foot. No abnormalities were noted on EEG. Exome sequencing identified a homozygous frameshift deletion in the *COL18A1* gene (c.3523_3524del p.Leu1175ValfsTer72). This variant segregates with the disease phenotype in the family. Furthermore, this variant is classified as pathogenic by many functional assessment tools, including ClinVar (SCV003934783.1) and the ACMG criteria PM2, PP5, and PVS1. Conclusion: This is the first report of a molecularly diagnosed case of KS in Mali. Furthermore, this is the first molecularly diagnosed case of KS in Africa, to the best of our knowledge, contributing further to supporting the implication of *COL18A1* in KS and highlighting the importance of analyzing genetic variants from diverse populations. In addition, this report adds to the clinical and genetic knowledge of this rare condition.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2013W: A scan for genomic disorders of infertility in 1500 cases of spermatogenic failure

Authors:

A. Akbari, J. Ahuja, A. Riera-Escamilla, D. Conrad; Oregon Hlth.& Sci. Univ., Portland, OR

Abstract:

Male infertility affects up to 15 % of the male population and estimations indicate that the prevalence is on the rise. Massively parallel sequencing technologies revolutionized the discovery of genetic causes of infertility in the last decade with an emphasis on monogenic and Mendelian forms. However, 30-40 % of cases remain idiopathic necessitating more research into genetic causes. Despite the advances made, the role of CNVs in male infertility pathogenesis in cohort settings remains fairly unexplored. Our genetics of male infertility initiative (GEMINI) cohort consists of 1502 male patients and 639 controls recruited from several countries putting us in a unique position for robust comparative CNV analysis. Segmental duplications (segdups) are known to cause CNVs through non-allelic homologous recombination. The aim of this study

is to identify CNVs unique to patients with potential segdup-mediated recurrence as infertility CNV hotspots.CNV calling was performed on exome alignment files based on read-pair method using Manta and Smoove and coverage method using XHMM and ExomeDepth. VCF outputs were annotated by ANNOVAR for gene annotations and SVAFOTATE for minor allele frequencies. Data aggregation, subsequent filtering steps and statistical analyses were done in R. Results were validated by IGV and PCR.A total of 717 CNVs were called in 569 infertility genes. Generalized linear model indicated patients as significantly more enriched for deletions larger than 100 kb against controls ($p = 0.00809$) and for segdup-flanked variants ($p = 0.02$). Deletions in *SYCE1*, *CDC45C*, *VCX*, *PUDP*, *BMBP8B* and *FAM47C* were found in potential CNV hotspots along with those in *DPY19L2* and *AZFc* genes. Moreover, rare deletions in dominant genes of *DMRT1*, *DNMT1*, *HNRNPL*, *TDRD5* and *SDF4*, and in recessive and X-linked genes of *ADAD2*, *STX2*, *STRA8*, *RBBP7* and *DMD* were found only in patients.In conclusion, our findings nominate new hotspots of segdup-mediated, disease-causing CNVs, reaffirming the idea of genomic disorders of infertility. Moving forward, the adoption of whole genome sequencing technology coupled with long-read solutions will enhance the accuracy and detection rate of large genomic events contributing to male infertility. CNV analysis, male infertility, genomics, whole exome sequencing, cohort study

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2014W: Accuracy of rare diseases inference in generative AI (ChatGPT3.5): comparison with in-house AI diagnostic supporting system

Authors:

T. Kaname[1], K. Yanagi[1], A. Igarashi[1], M. Yamamori[1], K. Naritomi[2], Y. Matsubara[1]; [1]Dept Gen Med, Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan, [2]Univ Ryukyus Sch Med, Okinawa, Japan

Abstract:

Background and Aims: The generative artificial intelligence (AI) has had an impact on various fields in recent years, including healthcare. However, the effect on highly specialised fields of medicine, such as rare diseases remains unclear. In this study, we aimed to investigate the accuracy of diagnosis estimation from clinical findings (symptoms) information by generative AI (ChatGPT3.5) for rare diseases with high specialisation.
Methods: As a prompt, clinical findings were entered into ChatGPT3.5 in the form of a question, and responses were requested three times in succession, and were considered according to whether or not the name of the disease was included among those presented. The first role

(doctor) was also given in order to increase accuracy. The clinical findings to be entered are the same 31 cases used for the initial validation in in-house AI diagnosis supporting system for rare disease, whose supporting system initially had an accuracy of 71% and now has an accuracy of 82%.
Results: The correct answer was provided in 14 of 31 cases (45.2%) among the candidate diseases given in the ChatGPT3.5 responses. There was an impression that accuracy tended to be higher for rare diseases with high recognition in the general population (e.g. Marfan syndrome) and lower for less recognised diseases. When comparing the percentage of correct answers for questions in Japanese with those in English, the percentage of correct answers was slightly (one case) higher for questions in English.
Conclusions: The ChatGPT3.5 may be useful for some relatively well-known rare diseases, depending on the disease in question, as it may answer the correct disease at a high rate. However, it should be noted that when symptoms (clinical findings) are entered into the ChatGPT3.5 and a possible disease question is asked, often a plausible answer is given, whereas the answer may be for a completely different disease. Although inputting clinical findings requires expertise, generative AI is generally available, hence it is important to assume that patients (or their family members) may enquire about the output results with their doctor.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2015W: An innovative approach for identifying low-penetrance structural birth defect genes

Authors:

D. Scott[1,2], T. Scott[3], A. Hardcastle[4], R. Belanger Deloge[5], M. Sy[6], T. Gordon[6], E. Huth[7], X. Zhao[1,8], S. Lalani[6,2], J. Rosenfeld[1], B. Kunisetty[6], B. Martin-Giacalone[9], A. Agopian[10], P. Luna[1], C. Shaw[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Univ. of North Carolina Med. Sch., Chapel Hill, NC, [4]McGovern Med. Sch. at UTHlth., Houston, TX, [5]Boston Children’s Hosp., Boston, MA, [6]Baylor Coll. Med., Houston, TX, [7]Univ. of California San Francisco, San Francisco, CA, [8]Baylor Genetics, Houston, TX, [9]Washington Univ. Sch. of Med., Saint Louis, MO, [10]UTHlth.Sch. of Publ. Hlth., Houston, TX

Abstract:

A molecular cause is not identified in most children with structural birth defects who undergo genetic testing. This is due, in part, to an incomplete understanding of the genes that contribute to their development. Genes that are associated with low penetrance in humans may go unrecognized even though they play a critical role in development. Identifying these low-

penetrance genes is a prerequisite to accurate genetic counseling for families affected by structural birth defects and enhances the ability of physicians to generate individualized medical plans for affected individuals. Deleterious coding variants in low-penetrance genes are also likely to represent very rare variants with modest to intermediate effect sizes, which are difficult to identify in traditional genetic studies. Such variants may represent important genetic modifiers of penetrance and may significantly contribute to the pathogenesis of structural birth defect cases that are associated with polygenic or multifactorial inheritance. To identify low-penetrance genes, we have developed an alternative gene discovery approach in which candidate genes are prioritized based on their similarity to genes already known to cause a phenotype. This prioritization is accomplished using an innovative machine learning algorithm that leverages data from large-scale genomic knowledge sources to generate phenotype-specific rank annotation scores for all RefSeq genes. Additional evidence in support of a gene's role in the development of the phenotype is then identified from clinical, research, and public databases, expression profiles, animal models, and isolated case reports. Using this approach, we have implicated 17 genes for congenital diaphragmatic hernia, 8 for anorectal malformations, 5 for esophageal atresia/tracheoesophageal fistula, 3 for anomalous pulmonary venous return, and 9 for microphthalmia/anophthalmia/coloboma. These results provide compelling evidence that this approach can be used to identify low-penetrance genes for a wide variety of structural birth defects.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2016W: An update on microduplications in 17q21.31 region potential cause of *KANSL1* gene disruption

Authors:

A. Alaqeel[1], H. AL-Balool[2], L. Matsa[3], A. Alahmad[4], A. Aladwani[4]; [1]New Ahmadi (KOC) Hosp., Ahmadi City, Kuwait, [2]Kuwait Med. Genetics Ctr. (KMGC), Kuwait City, Kuwait, [3]iGenomix Labs, Dubai Branch, Dubai City, United Arab Emirates, [4]Kuwait Med. Genetic Ctr. (KMGC), Kuwait City, Kuwait

Abstract:

Background: Koolen-De Vries syndrome (KDVS) is a genetic syndrome that is caused by either a 17q21.31 microdeletion, or intragenic variants lead to loss of function of the *KANSL1* gene. KDVS has wide range of phenotype including variable severity of intellectual disability, hypotonia, failure to thrive, characteristic facial features, various ophthalmic findings, and other neurologic and cardiovascular defects.Method: Copy number analysis using Affymetrix Cytoscan

HD Microarray assay to analyse patients with unclear clinical diagnosis. In addition, multiplex ligation-dependent probe amplification (MLPA) analysis was carried out to further confirm the aberrations.Results: Following our previous work (Alaqeel et al; 2021 ASHG conference), another 20 (total of 25) unrelated cases were identified with microduplication at 17q21.31 (approximately chr17:44,169,848-44,788,546). Although this type of aberration has not been reported as causative for a known microduplication syndrome; it is partially overlaps with the *KANSL1* gene and may potentially compromise its function. In addition, MLPA data have also confirm the duplication in at least 5 cases out of 25. In these 25 cases, several phenotype(s) have been observed, including, developmental delay, hypotonia, hypertension, congenital heart defect, single kidney. Furthermore, 2 out of the 25 cases also harbor additional pathogenic aberration, the first one with 22q11.21 deletion (DiGeorge syndrome), and the second with 7q11.23 deletion (Williams Syndrome). Although such cases did exhibit features related to their syndromes, other features were observed that overlap with 17q21.31 microdeletion syndrome. Conclusion: More recurrent aberration has been identified in this work located within 17q21.31. Duplication that partially involved the *KANSL1* gene has been confirmed by two methods, confirming the genuity of such aberration. The data also suggest that the patients do exhibit similar characteristics as KDVS. Although there is some evidence supporting the pathogenicity of a microduplication disrupting the *KANSL1* gene function; however, more evidence is needed to confirm its affect. Therefore, further testing is required to investigate this aberration.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2017W: Beta-thalassemia in India: A decade of molecular diagnostics in a tertiary care centre

Authors:

K. Mandal, S. Shende; Sanjay Gandhi Post-Graduate Inst. of Med. Sci., Lucknow, India

Abstract:

Background: Beta-thalassemia is a common recessive single gene disorder with a carrier frequency of 1.5% globally, and 3-4% throughout India (ranging from <1% to 17%). In low and middle-resource countries like India, carrier screening and prenatal testing is the corner stone of management. Study design: Descriptive study describing variants in HBB gene in a total of 2647 samples which included 1164 probands, 1115suspected carriers and 368 prenatal samples over a period of 10 years (2014 to 2023) done in a Medical Genetics facility in North India. Methods: Molecular testing of affected children and suspected carriers (along with HPLC for Hb variant and RBC indices) were done by Sanger sequencing and GAP-PCR for 619 bp deletion. Targeted

testing in prenatal samples was done after ruling out maternal contamination. Reports were generated or revised as per HGVS nomenclature using NM_00518: ID for HBB gene. The study was approved by Institutional ethics committee. Results: Yearly distribution showed that there was gradual increase in proportion of prenatal samples with time; albeit there was a dip in number of all samples during the covid 19 pandemic (2020 and 2021). Among the probands 36% were found to be homozygous (22% for c.92+5G>C and 14% for other variants). Among individuals who were found to be compound heterozygous, 59% had c.92+5G>C as one of the two alleles. We had a new observation that all individuals who had c.92G>C allele also had c.-92C>G variant. It prompted us to check both the variants together. Among all individuals 1% had very rare variations. We did not find any pathogenic or likely pathogenic variants in individuals who were screened for HBB mutation with HbA2 less than 3.5% in HPLC for Hb variant. Maternal contamination was detected in 1% of the samples requiring repeat testing from fresh samples or cultured fetal cells. The error rate in the prenatal testing was 0.5%. Discussion: Families opted for molecular testing in the affected child or suspected carriers mainly for pre-natal testing. The other major indication was availability of government support for treatment. Though there had been other methods like ARMS-PCR for molecular testing, we straight away opted for Sanger sequencing along with GAP-PCR. Majority of our families presented during pregnancy for prenatal testing. We made in-house protocols for screening and mutation testing to facilitate early diagnosis and to minimize errors. We will be discussing rare mutations, interesting case scenarios, intricacies of genetic counselling and errors and caveats in our study.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2018W: Biallelic variants in *LLGL1*, a hippo-yap pathway regulator, are a candidate cause for a rare neurodevelopmental syndrome

Authors:

H. Yousaf[1], K. Gilmore[2], N. Vora[2], E. E. Davis[1]; [1]Stanley Manne Children's Res. Inst., Ann & Robert H Lurie Children's Hosp. of Chicago, Chicago, IL, [2]Dept. of Obstetrics and Gynecology, Div. of Maternal-Fetal Med., The Univ. of North Carolina Sch. of Med., Chapel Hill, NC

Abstract:

Hippo-Yap signaling is a transcriptional regulating pathway that plays a pivotal role in cellular proliferation, differentiation, and survival and thus regulates organ development and homoeostasis. This pathway functions through core kinases and activator proteins including MST1/2, SAV1, LATS1/2, YAP, and TAZ. Along with other polarity factors, lethal giant larvae (LLGL) proteins regulate Hippo-Yap pathway and function upstream of the core kinase complex

to negatively regulate Yap/Taz transcriptional output. In addition, previous studies have reported functional relevancy of LLGL proteins to apicobasal cell polarity, cell shape, cell division and cell cycle exit. In vertebrates, there are two paralogs, *LLGL1* and *LLGL2*, and neither of them have been implicated in Mendelian disease. Here, we report an affected individual with compound heterozygous rare, highly conserved, predicted deleterious variants in LLGL1 (Paternally inherited: NM_004140.4:c.101C>G p.(Pro34Arg); Maternally inherited: c.196G>A p.(Val66Met)). The paternal variant is absent in gnomAD while the maternal variant is ultrarare with VAF of 0.00000889 (v4.1.0). Our case had prenatal onset neuroanatomical (microcephaly, corpus callosum agenesis, cerebellar hypoplasia, polymicrogyria), ocular (aphakia, bilateral cataracts), and craniofacial patterning (hypertelorism and retrognathia) anomalies. Published in vivo studies have shown that *Llgl1 -/-* mice exhibit abnormal neural morphology, hydrocephalus, and neonatal lethality. Furthermore, *Llgl1* deficient neural progenitor cells failed to exit the cell cycle and underwent apoptosis. Published zebrafish models of llgl1 depletion displayed larger and dysmorphic cardiomyocytes, pericardial effusion, impaired blood flow, and aberrant valvulogenesis. To explore further the role of *LLGL1* in physiological processes related to neurodevelopment, we generated and are currently characterizing zebrafish models of *llgl1* ablation to understand the underlying pathophysiology in anterior structures during development. Together, our work implicates *LLGL1* in a hitherto unreported prenatal manifesting human genetic condition and will strengthen further the role of LLGL proteins in neurogenesis.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2019W: *CAMK2D* and isolated dilated cardiomyopathy: genotype-phenotype expansion and functional analyses support an association with disease

Authors:

B. Mitchell[1], S. McLean[1], J. Erickson[2]; [1]Baylor Coll. of Med., San Antonio, TX, [2]Univ. of Otago, Dunedin, New Zealand

Abstract:

Background: A family with dilated cardiomyopathy in two generations underwent extensive genetic testing, culminating in trio exome sequencing that identified a maternally inherited missense variant in *CAMK2D* [c.824, G>A (p.R275H)]. CAMK2D is highly conserved and encodes a type II multifunctional calcium/calmodulin-dependent protein kinase that may regulate heart function based on a murine model. Two publications (PMID 37372357 and 38272033) describe

various *CAMK2D* missense variants in 9 individuals, all with neurodevelopmental disabilities and some with dilated cardiomyopathy. Consequently, we investigated whether R275H is pathogenic and sought to better correlate genotype and phenotype.
Methods: This family underwent thorough clinical evaluations, and we performed functional studies of the R275H variant. All five siblings and their parents were evaluated with comprehensive histories, physical examinations, and serial echocardiography over 7 years. We evaluated the impact of the R275H variant on *CAMK2D* function using Camui, a fluorescence-based CaMK2D activity sensor that can be genetically modified to account for mutations.
Results: Variant testing confirmed a robust segregation of the R275H variant with dilated cardiomyopathy within this kinship, with mother and 4 of 5 children affected and harboring R275H. Their neurologic and developmental features were normal. The Camui sensor demonstrated the R275H variant to be significantly more active at baseline with substantial activation at a much lower calcium concentration, supporting that the mutant kinase is more active at baseline and at peak with increased sensitivity to calcium.
Conclusion: This study is the first to describe a family with a *CAMK2D* variant presenting with isolated dilated cardiomyopathy. Our observations, reinforced by Camui-sensor data, strongly support the pathogenicity of the R275H variant and expand the phenotypic spectrum of *CAMK2D* variants. This will materially improve counseling about the natural history of the condition. Additional investigations will be necessary to determine how the R275H variant contributes to cardiomyopathy and delineate the phenotypic spectrum of *CAMK2D*-related cardiomyopathy.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2021W: Complete chromosome 21 centromere sequences from a Down syndrome family reveal size asymmetry and differences in kinetochore attachment ⭐

Authors:

F. Mastrorosa[1], A. N. Rozanski[1], W. T. Harvey[1], J. Knuth[1], G. Garcia[1], K. M. Munson[1], K. Hoekzema[1], W. He[2], S. Sherman[2], T. C. Rosser[2], E. G. Allen[2], G. A. Logsdon[1,3], E. E. Eichler[1,4]; [1]Univ. of Washington, Dept. of Genome Sci., Seattle, WA, [2]Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA, [3]Dept. of Genetics, Epigenetics Inst., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [4]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA

Abstract:

Down syndrome, the most common form of human intellectual disability, is caused by precocious segregation and nondisjunction of chromosome 21. Differences in centromere structure have been hypothesized to play a potential role in this process in addition to the well-established risk of advancing maternal age. Using long-read sequencing, we completely sequenced and assembled the centromeres from a parent-child trio where Trisomy 21 (T21) arose in the child as a result of a maternal meiosis I error (MMIE). The proband carries three distinct chromosome 21 centromere haplotypes that vary by 11-fold in length--both the largest (H1) and smallest (H2) originating from the mother. The longest H1 allele harbors a less clearly defined centromere dip region (CDR), as defined by CpG methylation, and a significantly reduced signal by CENP-A chromatin immunoprecipitation sequencing when compared to H2 or paternal H3 centromeres. These epigenetic signatures suggest less competent kinetochore attachment for the maternally transmitted H1. Analysis of H1 in the mother indicates that the reduced CENP-A ChIP-seq signal, but not the CDR profile, pre-existed the meiotic nondisjunction event. A comparison of the three proband centromeres to a population sampling of 35 completely sequenced chromosome 21 centromeres shows that H2 is the smallest centromere sequenced to date and all three haplotypes (H1-H3) share a common origin of ~15 thousand years ago. We are currently assembling six additional T21 families using long-read sequencing data. We observe two additional families with a short chromosome 21 centromere for which we are assessing kinetochore competency. These results suggest that recent asymmetry in size and epigenetic differences of chromosome 21 centromeres may contribute to nondisjunction risk but not in all T21 cases.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2023W: *De novo* variants in *CDKL1* and *CDKL2* are associated with childhood-onset epilepsy, developmental delay, and poor motor coordination

Authors:

A. H. Bereshneh[1,2], J. C. Andrews[1,2], G. Bademci[3], N. A. Borja[3], S. Bivona[3], Undiagnosed Diseases Network, Baylor College of Medicine Center for Precision Medicine Models, S. Yamamoto[1,2], M. F. Wangler[1,2], S. McKee[4], M. Tekin[3], W. K. Chung[5], H. Bellen[1,2], O. Kanca[1,2]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, United States., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, United States., Houston, TX, [3]Dr. John T. Macdonald Fndn. Dept. of Human Genetics and John P. Hussman Inst. for Human Genetics, Univ. of Miami Miller Sch. of Med., BioMed. Res. Building (BRB), Miami, FL, United States., Miami, FL, [4]Northern Ireland Regional Genetics Service, Belfast City Hosp.,

Belfast, Northern Ireland, United Kingdom., Belfast, United Kingdom, [5]Dept. of Pediatrics, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, United States., Boston, MA

Abstract:

The *CDKL* (cyclin-dependent kinase-like) genes are a group of serine-threonine kinases consisting of five members in humans: *CDKL1-5*. *CDKL5* is the only member of this gene family that has been associated with a Mendelian disease, developmental and epileptic encephalopathy type 2 (DEE2), presenting with childhood-onset epilepsy, developmental delay, and poor motor coordination. In this study, we have identified four *de novo* variants in *CDKL2,* including three missense variants (*p.Ser119Gly, p.Met215Leu*, and *p.Asn468Ser*) and an in-frame insertion (*p.Glu332_Glu333insVal*), in five individuals (three unrelated probands and pair of monozygotic twin). These individuals presented with overlapping disease manifestations, including global developmental delay, intellectual disability, seizures, dyspraxia, growth abnormality, and speech deficits. In addition, we investigated paralogues of *CDKL2* in *de novo* databases and we found two *de novo* missense variants in *CDKL1* (*p.Val115Ala* and *p.Arg169Cys*) in two unrelated probands from Deciphering Developmental Disorders (DDD) project and GDX cohort with developmental disorder. We generated genetic reagents to model the *CDKL1* and *CDKL2* variants in *Drosophila melanogaster* and probe their phenotypic consequences *in vivo*. The fly gene, *CG7236* (hereafter called *cdkl*), is the only orthologue of human *CDKL1* and *CDKL2*. We generated a new loss of function allele of *cdkl* using CRISPR-mediated homologous recombination ($cdkl^{CRIMIC\text{-}T2AGAL4}$) that expresses the *GAL4* transcription factor in the expression domain of *cdkl*. We determined that *cdkl* is very sparsely expressed in the central nervous system throughout the lifespan. Interestingly, *cdkl* is specifically expressed in most or all *Drosophila* sensory neurons. Loss of *cdkl* is semi-lethal and escaper flies have climbing defects, heat-induced seizures, and reduced longevity that can be fully rescued by expression of either of the human *CDKL1* or *CDKL2* reference cDNAs, suggesting the functions of these genes are evolutionarily conserved. In contrast, the *CDKL1* and *CDKL2* variants do not fully rescue these phenotypes. Interestingly, ubiquitous or cell-type specific overexpression of the reference cDNAs do not cause an obvious phenotype, whereas overexpression of the human variants leads to climbing defects, heat-induced seizures, bang sensitivity, and reduced longevity, similar to the *cdkl* loss of function allele in flies. These data suggest that the *CDKL1* and *CDKL2* variants act as dominant negative alleles and are responsible for the clinical symptoms observed in the individuals. Further investigation is ongoing to probe the molecular basis of the observed phenotypes.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2024W: Deficiency of lncRNA *CHASERR* upregulates chromatin remodeler CHD2 and causes global transcriptomic changes in a novel neurodevelopmental disorder ⭐

Authors:

E. Yoon[1], V. Ganesh[2], A. O'Donnell-Luria[3], G. Carvill[1]; [1]Northwestern Univ., Chicago, IL, [2]Broad Inst. / Brigham and Women s Hosp., Cambridge, MA, [3]Boston Childrens Hosp. and the Broad Inst., Boston, MA

Abstract:

CHD2 is a chromatin remodeler critical for chromatin organization and developmentally regulated gene expression. Pathogenic loss of function *CHD2* variants cause treatment-resistant epilepsy. Upstream of the *CHD2* gene is a conserved lncRNA *CHASERR.* We identified *de novo CHASERR* deletions in three unrelated individuals with severe neurodevelopmental delay, facial dysmorphisms, cerebral dysmyelination, and growth retardation. These individuals have a more severe clinical presentation than those with *CHD2* haploinsufficiency, who unlike the *CHASERR* deficiency individuals, are mostly able to communicate verbally, ambulate independently, and perform fine motor tasks.

To investigate the downstream consequences of *CHASERR* disruption on *CHD2* regulation, we used patient-derived *CHASERR*$^{+/-}$ fibroblasts, induced pluripotent stem cells (iPSCs), and genome-edited diploid HAP1 *CHASERR* deletion lines as our model. Like the patients' deletions, the genome-edited HAP1 deletion lines removed the 5' UTR region and the first three exons of *CHASERR*.

CHASERR deletions increased *CHD2* expression at the RNA and protein level in both the HAP1 and patient iPSC lines. Moreover, gene ontology analysis of differentially expressed genes upregulated in HAP1 *CHASERR* deletion lines compared to wildtype (n=1199) generated pathways such as nucleosome assembly, nucleosome organization, regulation of cyclin-dependent protein kinase activity. Our findings align with current understanding of CHD2's function in chromatin remodeling and emerging evidence of CHD2's role in cell proliferation. Additionally, differentially expressed gene analysis between the *CHASERR* deletion (*CHD2* overexpression) and *CHD2* knockout (*CHD2* haploinsufficiency) lines showed upregulation of genes involved in cell fate and differentiation, including *MMP2,* a metalloproteinase involved in myelinogenesis and axonal regeneration. Dysregulated upregulation of such genes may contribute to premature neuronal differentiation and abnormal myelinogenesis and explain why the *CHASERR* individuals have a more severe disorder with microcephaly and cerebral dysmyelination, compared to *CHD2* haploinsufficiency individuals. Future studies will involve epigenomic approaches (CUT&RUN) to elucidate which changes in

global gene expression are directly due to increased CHD2 protein abundance via *CHASERR* deletion.
We conclude that *CHD2* is a dosage sensitive gene with a tight therapeutic window. Our study postulates that other mendelian disorders with suggested dosage sensitivity may be regulated by upstream lncRNAs and recommend careful analysis of such lncRNAs in clinical genomic analysis.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2025W: Defining the molecular mechanism of *NARS1*-mediated dominant neurological disease

Authors:

S. Peeples, A. Antonellis; Univ. of Michigan, Ann Arbor, MI

Abstract:

Aminoacyl-tRNA synthetases (ARSs) are essential enzymes that ligate amino acids to cognate tRNAs in the cytoplasm and mitochondria. Although ARSs are essential and ubiquitously expressed, loss-of-function missense mutations in seven dimeric ARS enzymes have been associated with dominant inherited neuropathy. Heterozygosity for missense and protein-truncating variants in asparaginyl-tRNA synthetase (*NARS1*; the gene encoding the enzyme that ligates asparagine to tRNA molecules in the cytoplasm) have been associated with distinct clinical phenotypes where patients present with a late-onset neuropathy or with an early-onset neurodevelopmental phenotype, which can include neuropathy. Thus, ARS genes cause a spectrum of dominant neurological diseases. Interestingly, all ARS enzymes implicated in dominant neurological disease (including NARS1) function as dimers and we previously showed that neuropathy-associated alanyl-tRNA synthetase (*AARS1*) alleles act via a dominant-negative mechanism. Here, we test pathogenic *NARS1* variants for a dominant-negative effect to determine if this mechanism is common to ARS-associated dominant neurological disease. We performed yeast complementation assays to test *NARS1* variants in isolation, which revealed that all alleles exhibit a loss-of-function effect. To test for dominant-negative properties of ARS-mediated neurological alleles, we developed a humanized yeast assay to co-express mutant human *NARS1* with wildtype human *NARS1*. These studies revealed that four out of five mutant alleles repress the ability of the wild-type allele to rescue yeast cell growth, consistent with a dominant-negative effect. Here, we will present our unpublished functional data on disease-associated *NARS1* variants and will discuss how differential dominant-negative effects may

impact patient phenotypes. In sum, this study expands our understanding of the molecular mechanisms of *NARS1*-related dominant human disease.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2026W: Development of patient-friendly diagnostic method for pediatric mitochondrial cases in India

Authors:

D. Saha[1], S. Kothari[2], S. D. Kulkarni[2], D. K. Das[1]; [1]ICMR-NIRRCH, Mumbai, India, Mumbai, India, [2]Bai Jerbai Wadia Hosp. for Children, Parel, Mumbai, India

Abstract:

The diagnosis of mitochondrial disorders (MD) is extremely challenging which requires a combinatorial approach of biochemical respiratory chain enzyme assays, histopathology and genetic analysis. A broader and high-end testing is often achieved from skeletal muscle biopsy, however, due to its highly invasive nature, it is practically very difficult to obtain muscle biopsies from infants and children. Therefore, in order to reduce the risks and expense associated with muscle biopsy, an alternative non-invasive approach using buccal swabs have been used in our study. Forty-three pediatric Indian patients suspected with MD along with age matched 37 healthy controls were recruited for the study. Using spectrophotometric analysis, a significant decrease in enzyme activities of mitochondrial Complex I, III and IV ($p= <0.05$) have been observed in buccal samples. A comparative analysis between muscle biopsy and buccal swab samples ($n=4$) was also analyzed and a correlation was observed in the buccal swab samples similar to the deficiencies present in muscle biopsies. Further, in order to analyze the diagnostic potential of buccal swab as a non-invasive predictive tool, receiver operating characteristic (ROC) curve analysis had shown to significantly predict complex I, III and IV deficiency in buccal samples. The sensitivity of the buccal swab approach was observed to be approximately 78% which makes it a considerable non-invasive preliminary diagnostic tool for pediatric mitochondrial cases, prior to planning for a skeletal muscle biopsy.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2027W: Diagnostic yield of whole genome sequencing for patients with neurodevelopmental delays

Authors:

R. Rigobello, A. Neogi, J. Chibuk, L. Vossaert, L. Meng, F. Xia, C. Eng; Baylor Genetics, Houston, TX

Abstract:

Objectives: Whole genome sequencing (WGS) is increasingly being adopted as a first-tier diagnostic tool for patients with neurodevelopmental delays (NDD), in accordance with guidelines from the American College of Medical Genetics and Genomics (ACMG). However, there is limited information on the utility of WGS for patients with NDD. In this study, we reviewed the diagnostic yield of WGS for patients presenting with various NDDs.
Methods: A retrospective analysis was performed to examine the clinical and genetic findings from over 1100 consecutive patients tested by WGS in our clinical laboratory. Clinical notes were queried for indications associated with NDD, and the frequency of phenotype-related positive results detected by WGS (pathogenic (P), likely pathogenic (LP), and high-suspicion variants of uncertain significance (VUS)) was calculated. Results: 22% of patients had an NDD-related indication. Of these, half had global developmental delays (GDD, 51%). Additional NDD symptoms included speech and language delays (35%), motor delays (26%), intellectual disabilities (ID, 20.5%), and developmental regression (11%). Additional neurological features included other brain imaging abnormalities (45%), autism (20.5%), and seizures (37%). Most patients (97%) had at least one non-neurological symptom. The overall diagnostic yield of WGS for patients with NDD was 47% (40% with P/LP only). Copy number variations accounted for 10% of diagnoses, and included two patients with Angelman syndrome, and one patient each with Turner syndrome, Prader-Willi syndrome, and Duchenne/Becker muscular dystrophy. Two patients had ATXN2 expansions associated with spinocerebellar ataxia 2 and one had uniparental disomy associated with Temple syndrome. For patients with a clinical indication of GDD, the diagnostic yield was 54% (45% with P/LP only). For 19 patients with both GDD and ID, the diagnostic yield of WGS was 79% (74% with P/LP only).
Conclusions: WGS identified a genetic diagnosis in nearly half of patients with NDD. Several variant types were detected, which can often be missed by other standalone genomic technologies. The diagnostic yield was higher in patients with both global developmental delays and intellectual disability, with nearly 80% of this cohort testing positive.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2028W: DNA methylation episignature analysis is useful for reclassification of a maternal inherited missense variant with uncertain significance in the *KDM5C* gene

Authors:

K. Takano[1,2], T. Kawai[3], R. Takeda[4], T. Yamaguchi[1,2,5], K. Nakabayashi[3], S. Saitoh[6], T. Kosho[2,1,5]; [1]Ctr. for Med. Genetics, Shinshu Univ. Hosp., Matsumoto, Japan, [2]Dept. of Med. Genetics, Shinshu Univ. Sch. of Med., Matsumoto, Japan, [3]Dept. of Maternal-Fetal Biology, Natl. Res. Inst. for Child Hlth.and Dev., Tokyo, Japan, [4]Dept. of Med. Genetics, Nagano Children's Hosp., Azumino, Japan, [5]Div. of Clinical Sequencing, Shinshu Univ. Sch. of Med., Matsumoto, Japan, [6]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan

Abstract:

[Background] Claes-Jensen syndrome (CJS: MIM #300534) is an X-linked intellectual disability disorder, which is caused by pathogenic variants in the *KDM5C* gene. *KDM5C* encodes a histone demethylase involved in histone modifications and chromatin remodeling. CJS is characterized by developmental delay/intellectual disability, short stature, and characteristic facial features. Interpretation of *KDM5C* variants can be difficult due to its X-linked inheritance and non-specific clinical manifestations. CJS has been known to have unique genomic DNA methylation patterns called "episignatures" and episignature analysis have been reported to be useful for interpretation of variants of unknown significance (VUS) in the *KDM5C* gene. **[Case report]** The patient, a 7-year-old boy, was born at term as the first child without perinatal complications. At the age of 5 years, he developed epilepsy, but his seizures were controlled with valproic acid. At the age of 6 years, he visited our hospital and presented with severe global developmental delay, short stature (-2.9SD), misalignment of teeth, and brachydactyly. A custom NGS panel-based analysis for intellectual disability detected a hemizygous missense variant in the *KDM5C* gene classified as VUS. This variant was inherited from his asymptomatic mother except for short stature. Episignature analysis of the patient's genomic DNA reveals CJS-specific pattern. **[Conclusion]** DNA methylation episignature analysis would be useful for reclassifying a VUS in the *KDM5C* gene, the specificity of which could lead to appropriate genetic counseling based on an accurate diagnosis.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2029W: *Drosophila* models of *ABCD1*-related X-linked Adrenoleukodystrophy show possible implications in peroxisomal biogenesis

Authors:

S. Jangam[1,2], Y. Manor[3], C. Hyunglok[4], P. Bhagwat[1,2], J. Andrews[1,2], H. Chester[1,2], S. Srivastav[1,2], J. Botas[1,2], A. Moser[5], M. Wangler[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [3]Metabolic Disease Unit, Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr.,, Ramat Gan, Israel, [4]Stanley H. Appel Dept. of Neurology, Houston Methodist Academic Inst., Weill Cornell Med. Coll., Houston, TX, [5]Hugo W Moser Res. Inst., Kennedy Krieger Inst., Baltimore, MD

Abstract:

X-linked adrenoleukodystrophy (X-ALD) is a progressive neurodegenerative disorder caused by a loss-of-function (LOF) mutation in the *ATP binding cassette subfamily D member 1* (*ABCD1*) gene, leading to the accumulation of very long-chain fatty acids (VLCFAs). This disorder exhibits striking heterogeneity; some male patients develop a neuroinflammatory demyelination disorder, while other patients, including most affected women, experience chronic progressive myelopathy. All patients are born asymptomatic, and it is currently impossible to predict symptoms based on genetics, complicating clinical decision-making. Previous fruit fly models highlighted an upstream deficiency in VLCFA synthetases, leading to reduced lifespan, retinal degeneration, and VLCFA accumulation. Here, we present a study of LOF of *ATP binding cassette subfamily D member 1* (*Abcd1*) as a *Drosophila* model of X-ALD. First, we used UAS-$Abcd1^{RNAi}$, *which*, when overexpressed ubiquitously and spacio-temporally, showed a severe reduction in survival, abnormal peroxisomal biogenesis, abnormal eye formation, and retinal degeneration. We also utilized a known LOF allele $Abcd1^{\Delta 4}$ and observed locomotion defects compared to the control. When we performed biochemical analysis of VLCFAs observed an accumulation of the signature hexatonic acid (C26:0). Next, we generated the fly model for the human *ABCD1* ($ABCD1^{Ref}$) along with the known X-ALD-causing variant $ABCD1^{R518Q}$ under the UAS promoter. When overexpressed ubiquitously and spacio-temporally, the human $ABCD1^{Ref}$ causes abnormal peroxisomal biogenesis and late larval lethality. However, the $ABCD1^{R518Q}$ variant also shows peculiar peroxisomal biogenesis that notably differs from the reference and causes partial larval lethality. $ABCD1^{R518Q}$ variant expressing flies fail to behave as reference animals, thus proving a LOF mechanism for the variant. This research in the fly model partially recapitulates the biochemical manifestations of X-ALD observed in humans. These results identify a potential link between *Abcd1* function and peroxisomal biogenesis in the fly's simplified mechanism of VLCFA import in peroxisomes.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2030W: *Drosophila* models of HNRNPH2-related neurodevelopmental disorder variants demonstrate developmental defects

Authors:

M. Mew, J. Andrews, S. Jangam, B. Sharma, B. Vuocolo, The Texome Project, M. Wangler; Baylor Coll. of Med., Houston, TX

Abstract:

Pathogenic mutations in the gene encoding heterogeneous nuclear ribonucleoprotein H2 (HNRNPH2) cause HNRNPH2-related neurodevelopmental disorder (HNRNPH2-NDD). This is an ultra-rare X-linked dominant disease characterized by symptoms that begin within the first year of life and include developmental delay, intellectual disability, behavioral and psychiatric disorders, and growth and musculoskeletal abnormalities. There are currently no treatments for HNRNPH2-NDD, and the mechanisms of pathogenesis for this disease remain unknown and largely unexplored. HNRNPH2 is an RNA-binding protein with known functions in alternative splicing and 3'-end RNA processing. The *Drosophila* ortholog of *HNRNPH2* is *glorund*, and this gene is critical for development and expressed through the central nervous system of the fly. We are generating *Drosophila* models of three *HNRNPH2* variants, including the known pathogenic missense mutations R206Q and R206W. We also report the identification of a human proband with a presumed de novo missense variant, G399R, who presented with intellectual disability, seizures, craniosynostosis, osteopenia, cerebral palsy, adrenal insufficiency, hypothyroidism, gastrointestinal symptoms, neurogenic bladder, amenorrhea, hypocalcemia, neutropenia, esotropia, and blindness. Tissue-specific RNAi knockdown (KD) of *glorund* using eye-, wing-, glia-, or thorax-specific drivers or a weak ubiquitous driver did not affect viability or morphology. However, KD with a moderate or strong ubiquitous driver reduced viability. We can therefore conclude that there is a minimum level at which *glorund* must be expressed for survival. In addition, *glorund* KD using an early neuron-specific driver was lethal, while flies with *glorund* KD in its endogenous expression pattern were viable. These results could be due to differences in the level of gene knockdown between these two lines. Alternatively, these results raise the possibility that Glorund has non-cell autonomous roles in nervous system development. Overexpression (OE) studies show that ubiquitous OE of reference- or variant-*HNRNPH2* is lethal. Furthermore, eye-specific OE of variant forms of *HNRNPH2* using an early eye-specific driver produced severe eye defects compared to

reference-*HNRNPH2* OE. This indicates that OE of *HNRNPH2* during eye development is deleterious. Overall, our results demonstrate that our *Drosophila* model is capable of identifying functional changes in variant forms of *HNRNPH2*. We will also use our models to elucidate the mechanisms of action by which variants in HNRNPH2 promote HNRNPH2-NDD pathogenesis.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2031W: Elucidating the role of *de novo*structural variations in a Mendelian disease cohort consisting of 169 families

Authors:

M. Jamsandekar[1], H. Du[1], Human Genome Sequencing Center, S. Jhangiani[2], J. Posey[1], R. Gibbs[2], J. Lupski[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX

Abstract:

Structural variants (SVs) include both copy number variants (CNVs) and copy number neutral variants such as balanced translocations, inversions, or mobile element insertion. They change the genome architecture and can have profound consequences for human disease. *De novo* SVs, new mutations only present in a proband from a trio genome analysis, but not in parental genomes, have profound impact on novel gene and regulatory element discovery, as well as genotype and phenotype association identification. Despite technological advances, computational identification of true positive *de novo* SVs, with wet-bench experimental validation, remains a challenge. In this study, we developed a *de novo* SV detection pipeline and implemented it on 169 parent-child trios in which the proband remained without an etiological diagnosis. We used short read genome sequencing (sr-GS) data files in which the subjects' personal genomes were sequenced at 30x coverage. These families originated from the Baylor College of Medicine Mendelian disease cohort of the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) consortium. SVs were called with the DRAGEN pipeline at the Human Genome Sequencing Center (HGSC). *De novo* CNVs larger than 10 kb were further validated with the internally developed visualization tool - VizCNV and then underwent novel genotype-phenotype analysis. On average, 10 *de novo* CNVs per family were called by DRAGEN, of which nearly 90% were either false positive as they failed e-validation with VizCNV or were embedded in a repetitive region and were thereby difficult to resolve without additional molecular methods. If the CNV region harbors one or more genes, the number of gene copies will further be determined with wet-bench experiments using ddPCR. Moreover, CNVs overlapping loci with known dosage sensitivity (e.g., genomic disorder regions, gene dosage, exon dosage) from DECIPHER and ClinGen will be assessed. *De novo* mobile element insertions,

e.g., *Alu*, which are known to disrupt the genomic integrity and contribute to genomic diseases will also be studied. This pipeline and visualization tool may identify potential pathogenic *de novo* SVs in a single family that contribute to a proband's clinical findings, gain insight into SV mutagenesis and mutation rates, as well as to understand the contribution of SV to conditions and clinical phenotypes manifesting in a disease cohort.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2032W: Establishing a comprehensive registry and natural history study for Pol II-associated developmental disorders

Authors:

K. Ward[1], A. Hansen[1], C. Hansen[2], J. Kuldell[3], C. Kaplan[3]; [1]Geneial, Houston, TX, [2]Geneial, Houston, UT, [3]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

RNA polymerase II (Pol II) is responsible for all protein coding gene synthesis in humans. Pol II-associated disease is a multisystemic developmental disorder with a wide range of presentations driven by germline mutations in the POLR2A gene, encoding the largest Pol II subunit and the majority of the Pol II active site. As Pol II is formed by multiple subunits, it is likely that mutations in other RNA polymerase II subunits, especially those that contribute to the Pol II active site, e.g. POLR2B, could emerge as driving related disorders. Previously identified rare variants in POLR2A have been associated with clinical manifestations that include developmental delay, intellectual disability, hypotonia, seizures, and dysmorphic features, among others. It is reasonable to hypothesize that these phenotypes relate to abnormal gene expression resulting from Pol II defects during development. Understanding the full genotype-phenotype spectrum and natural history of Pol II transcriptionopathy is essential for advancing patient care and translational research.

To address this need, we have designed a patient-centric registry on the Geneial privacy-preserving registry platform. Our model integrates electronic patient-reported outcomes (ePROs) with molecular diagnostic and phenotypic data to collect high-quality baseline and longitudinal data. The baseline survey extends the ClinGen-supported GenomeConnect schema-designed to capture a comprehensive phenotypic summary-by integrating standardized instruments for social determinants of health and quality of life, and a prospective, longitudinal component with recurrent surveying of relevant dynamic elements.

Surveys are administered through the Geneial Advocate app (web/iOS/Android). Participants or their legal guardians provide informed consent electronically before completing detailed

surveys about their health, development, and genetic testing results.
Only individuals with confirmed genetic testing results indicating mutations in POLR2A or related Pol II subunit genes will be recruited. The registry aims to enroll a diverse cohort, including children, to achieve better understanding of the relationship between genetic mutations and clinical outcomes.
By establishing this comprehensive registry, we aim to accelerate the discovery of correlations between genotypes and phenotypes, improve patient care, and advance translational research towards therapeutic development. The registry will serve as a valuable resource for researchers and clinicians, fostering collaborations and advancing our understanding of Pol II deficiency, ultimately improving outcomes for patients.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2033W: Estimating the prevalence of rare Mendelian diseases in a founder population, using whole genome sequences from the CARTaGENE Biobank in Quebec

Authors:

A. Green[1,2], J. Pelletier[1,2], D. Ashton-Beaucage[1,2,3], G. Gagnon[1,2], A-M. Laberge[4], M. Tétreault[5,6], D. Taliun[1,2], C. Bhérer[1,2]; [1]Dept. of Human Genetics, Faculty of Med. and Hlth.Sci., McGill Univ., Montréal, QC, Canada, [2]Canada Excellence Res. Chair Program in Genomic Med. and VP Dahdaleh Inst. of Genomic Med., McGill Univ., Montréal, QC, Canada, [3]Dept. of Med. Genetics, McGill Univ. Hlth.Ctr., Montréal, QC, Canada, [4]Dept. of Pediatrics, Université de Montréal, Montréal, QC, Canada, [5]Dept. of NeuroSci.s, Université de Montréal, Montréal, QC, Canada, [6]Univ. of Montreal Hosp. Res. Ctr. (CRCHUM), Montréal, QC, Canada

Abstract:

For most rare diseases, there is incomplete data on disease prevalence, limiting the possibility of targeting screening toward high-risk populations and individuals. It is well known that founder populations, such as the French Canadians (FCs) in Quebec, present with a high prevalence of some rare diseases due to the founder effect. Typically, prevalence is estimated using patient cohorts and newborn screening, but population biobanks have emerged as an alternative method. **Here, we use genomic data from the population-based CARTaGENE (CaG) biobank in Quebec to estimate the birth prevalence of rare diseases in the FC founder population**. We characterized 80,407,531 variants using high-depth whole genome sequences from 2,173 CaG participants, where 1,756 (~81%) are FCs. In total, 2,173 variants were previously interpreted as pathogenic or likely pathogenic (PLP) in the CLINVAR database, with

no conflicting evidence. Of those, 148 were significantly enriched (one-sided Fisher exact test, p < 2.3×10^{-5}) in FCs when compared to non-Finnish Europeans from gnomAD. Enriched variants included those underlying Mendelian diseases described to have a higher prevalence in the FC population, such as Leigh Syndrome FC type (MIM#220111; odds ratio=18.9, $p=7.92 \times 10^{-7}$), but also novel candidate founder variants, such as Rajab interstitial lung disease with brain calcifications 1 (MIM#613658, odds ratio=50.8, $p=4.66 \times 10^{-13}$). 209 genes for approximately 125 autosomal recessive (AR) conditions had PLP variants with a cumulative carrier rate above or equal to 1/100. To illustrate birth prevalence estimation from population-based cohorts, we curated a set of 27 diseases (20 AR, 7 autosomal dominant) previously identified as occurring more frequently in FCs. We estimated prevalence from allele frequencies assuming Hardy-Weinberg equilibrium and compared our new estimates to previously published estimates (where available). We found that many previous regional estimates are higher than our population-wide estimates, in line with a regional founder effect. For example, ACCPN (MIM#218000) has a CaG estimate of 5.06/100,000 births (95% CI: 2.23-10.7/100,000 births) and a previous estimate in the Saguenay-Lac-Saint-Jean (SLSJ) region of 47.2/100,000 births. Similarly, in Zellweger syndrome (MIM#614862), where the CaG estimate is 0.29/100,000 births (95% CI: 0.05-1.24/100,000 births), and the previous SLSJ estimate is 8.20/100,000 births. In future work, we will estimate the prevalence and variable expressivity for all 125 above AR diseases, which can be used to improve early diagnostic screening programs and provide timely intervention.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2034W: Examination of a novel biomarker for pyridoxine-dependent epilepsy, 6-oxo-pipecolate, and its potential implications for screening and treatment

Authors:

J. Souder, K. Clarke, A. Larson, J. Thomas, C. Coughlin, W. Timothy, O. D'Annibale; Univ. of Colorado, Aurora, CO

Abstract:

Pyridoxine-dependent epilepsy (PDE) is a disorder of lysine metabolism that typically presents in infancy with medication-refractory seizures responsive to pyridoxine (vitamin B6) therapy. Patients diagnosed and treated earlier in the disease course have been shown to have improved neurodevelopmental outcomes, but early diagnosis remains a challenge. The metabolite, 6-oxo-pipecolate (6-oxo-PIP), has been suggested as a reliable biomarker of PDE, but studies regarding

the specificity of this metabolite for PDE and whether it can be used as a marker of treatment efficacy are lacking. Here, we present a case of pyridoxine-dependent epilepsy diagnosed in an infant via ultra-rapid whole genome sequencing with subsequent initiation of triple therapy for PDE in the early infantile period. 6-oxo-PIP levels decreased by over 50% following initiation of treatment in this patient (1.29 uM before treatment and reduced to 0.61 uM after treatment), suggesting adequate treatment may reduce 6-oxo-PIP levels in patients with PDE. Additionally, we provide comparison with 6-oxo-PIP levels in a cohort of patients with non-PDE-related seizures of various etiologies from the newborn period to early adulthood. Ultimately, this provides a unique report of early PDE diagnosis by ultra-rapid whole genome sequencing and offers insight into the utility of 6-oxo-PIP as a clinically utilized biomarker for PDE.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2035W: Exploring craniofacial variation in a background mouse population to better understand the genetic architecture of Mendelian diseases, using achondroplasia as a case study

Authors:

K. Tran[1], H. Hoskens[1], J. Aponte[2], C. Unger[1], C. Rolian[1], P. Claes[3], B. Hallgrímsson[1]; [1]Univ. of Calgary, Calgary, AB, Canada, [2]Deep Surface AI, Calgary, AB, Canada, [3]KU Leuven, Leuven, Belgium

Abstract:

Approximately 446 million individuals are affected by rare diseases worldwide, 80% of which are genetic in origin. Genetic diseases are traditionally categorized under Mendelian or complex inheritance. However, the distinction between Mendelian and complex diseases is increasingly unclear. A recent study demonstrated that Mendelian and complex diseases that have similar phenotypes tend to share molecular and/or genetic mechanisms [PMID: 30290150]. This suggests that Mendelian disease *phenotypes* are neither rare nor unique as the *disease*. In fact, non-disease associated genes in the background population should then also be able to recapitulate disease phenotypes, as in complex inheritance. In this case study, we investigate achondroplasia as a classic Mendelian disease. Achondroplasia is the most common form of dwarfism in humans and patients present with characteristic craniofacial features. Interestingly, the Longshanks mouse, a selection model, exhibits the opposite craniofacial features of achondroplasia. Using geometric morphometrics (GM), a multivariate shape analysis, the shape variation generated by the Longshanks model captured the craniofacial features of achondroplasia. We thus modelled the achondroplasia phenotype using Longshanks animals.

We then scored Diversity Outbred mice, a heterogenous background population, according to their craniofacial similarity to the Longshanks/achondroplasia phenotype using GM. These scores were tested in a genome-wide association study to reveal two significant loci whose annotated genes (*RBFOX3* and *ADAMTS20*) have not been previously implicated in achondroplasia. To validate our results, we will test the joint effect of these genes on craniofacial shape in humans, using an in-house developed method of multivariate genotype-phenotype mapping. In this study, we validate the use of related or opposite phenotypes as a proxy to study disease phenotypes. This approach is specifically important for rare diseases in which lack of patients present an ongoing challenge for genetic research. Further, genes that contribute to disease phenotypes, such as those identified in this study, may modulate variability in disease presentation, known as disease penetrance. Lastly, we demonstrate that traditional Mendelian diseases, like achondroplasia, may have a more similar genetic architecture to complex diseases than widely accepted.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2036W: Exploring molecular mechanisms of GABA dysregulation in *SLC6A1*-related neurodevelopmental disorders

Authors:

P. Hall, K. L. Jay, J. C. Andrews, S. V. Jangam, R. German, V. Gomez, H. Pan, National Brain Gene Registry, Undiagnosed Diseases Network, S. Yamamoto, M. F. Wangler; Baylor Coll. of Med., Houston, TX

Abstract:

Variants in Solute Carrier Family 6 Member 1 (*SLC6A1*) cause a rare neurodevelopmental disorder that affects nearly 200 individuals. Patients experience a range of debilitating symptoms, including myoclonic-atonic epilepsy, autism spectrum disorder, intellectual disability, and movement disorders. *SLC6A1*-related disorders severely compromise quality of life of affected individuals, with epilepsy being the most severe co-morbidity. Often, patients attempt several medications and may spend years optimizing treatments before finding the most effective therapeutic. This trial-and-error process stems from a fundamental lack of understanding of disease pathogenesis and the mechanism that causes the seizures to arise. Currently, it is proposed that *SLC6A1* mutations lead to haploinsufficiency, where one healthy copy of *SLC6A1* is not enough to provide proper protein function. *SLC6A1* encodes the GABA transporter 1 (GAT1) protein that is responsible for GABA reuptake at the synapse. GABA is the main inhibitory neurotransmitter in humans, thus impaired GABA reuptake results in reduced

inhibitory activity, which may contribute to seizures. Variants in *SLC6A1* result in impaired or complete loss-of-function (LOF) of GAT1 activity. In this study, we modeled patient-specific variants in *Drosophila* to better define pathological mechanisms of disease, specifically *SLC6A1*^{A288V}, *SLC6A1*^{S295L}, *SLC6A1*^{G297R}, *SLC6A1*^{F339L}, and *SLC6A1*^{A334S}. We are evaluating seizure phenotypes with bang sensitivity assays, neuronal signaling with electroretinogram (ERG) recordings, and performing sleep and circadian rhythm analysis in the fly. We have observed bang sensitivity, ERG, and circadian rhythm abnormalities using our variant specific *SLC6A1* lines. We are using these readouts of GABA dysregulation to evaluate *SLC6A1* targeted therapeutics. Identifying therapeutics that best suppress seizure phenotypes by genotype will improve outcomes for those living with *SLC6A1*-related disorders and other forms of epilepsy.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2037W: Functional Validation of *HMCN1* as a Candidate Gene for Childhood-Onset Essential Hypertension

Authors:

K. Davey[1], E. Wonkam Tingang[1], Y. Han[1], M. Gupta-Malhotra[2], J. Posey[3], S. Hashmi[4], D. Milewicz[5], S. Shete[6], A. Haldipur[1], E. Banfield[1], N. Hanchard[1]; [1]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Baylor Coll. of Med., San Antonio, TX, [3]Baylor Coll. of Med., Houston, TX, [4]McGovern Med. Sch. at UTHlth., Houston, TX, [5]Univ Texas McGovern Med Sch, Houston, TX, [6]Univ Texas MD Anderson Cancer Ctr, Houston, TX

Abstract:

Childhood-onset essential hypertension (COEH) is an uncommon form of hypertension in which children under the age of 18 have sustained blood pressures greater or equal to the 95th percentile for their age, sex, and height. Childhood blood pressure is known to be a predictor of future blood pressure status; thus, children with COEH are at risk for developing hypertension-related complications as adults, including heart attack, stroke, and renal disease. The heritability of COEH is quoted as between 70-80%, however, the genes and genetic models underlying COEH susceptibility remain unknown and understudied. The relative rarity, high heritability, and overrepresentation in African ancestries of COEH suggest a role for rare genetic variation in its pathogenesis. We identified rare, putatively damaging variants in *HCMN1* in two families segregating in an autosomal recessive manner among probands with COEH. To evaluate its functional context, we employed cellular and mouse models of *HMCN1*, as this gene is an essential component of the extracellular matrix architecture in vascular smooth muscle (OMIM:

608548). Using vascular smooth muscle cells from human aorta, we performed an *in vitro* siRNA knockdown of *HMCN1* that resulted in increased cellular stiffness on atomic force microscopy, consistent with a hypertension phenotype. We then leveraged resources from the knockout mouse phenotyping project to breed *Hmcn1*$^{+/-}$ and *Hmcn1*$^{-/-}$ mice on a wild-type C57BL/6 background. We measured the blood pressure of juvenile mice (5-9 weeks) via non-invasive tail cuffs daily for two weeks, and found that mean arterial pressure, systolic blood pressure, and diastolic blood pressure in *Hmcn1*$^{-/-}$ mice (N=3) were significantly higher than in wild-type (N=1) or *Hmcn1*$^{+/-}$ (N=4) mice (all $p<0.005$). There are currently no published data on the direct effects of *HCMN1* on blood pressure or heart function, making our study the first of its kind. Additional studies to confirm our findings and understand the pathophysiological effects of *HCMN1* loss-of-function on blood pressure are underway. These studies will provide diagnostic insights into the genetics underlying COEH with the ultimate goal of developing pediatric therapeutic interventions in the future.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2038W: Gene variants identified in pediatric patients with Hypermobile Ehlers-Danlos Syndrome (hEDS)

Authors:

E. Elias, J. Bell; Children's Hosp. Colorado, Aurora, CO

Abstract:

Query: Pediatric patients with suspected hEDS are difficult to assess, even when using the new pediatric criteria published in 2023. The Multidisciplinary EDS Center of Excellence at Children's Hospital Colorado is one of the few centers in the US where children with suspected hEDS are evaluated and managed. In children, hEDS-like symptoms can be caused by multiple underlying disorders, each with different prognoses and interventions. Important clues found in the medical history, family history, and on exam in pediatric patients may point to genetic disorders, which when confirmed, can lead to optimal medical care and appropriate genetic counseling. **Study Design:** Patients aged 5-30 years with an hEDS diagnosis from a Genetics clinic were referred to our Multidisciplinary EDS program and evaluated by a comprehensive team of specialists, with carefully annotated history and physical findings. Genetic testing included Next Gen Sequencing panels and Whole Exome Sequencing were performed when possible. **Results:** More than 3,100 patient visits representing almost 600 patients have been seen since Nov 2019. Genetic testing was performed in a subset of patients. Genetic variants have been identified in 26 different genes, including connective tissue disorders (*COL1A1/2,*

COL3A1, COL5A1/2, FBN1, TGFBR2, PLOD1, ADAMTS2 and TNXB), neuromuscular disorders (*COL6A1, COL12A1, CACNA1S, MYH2, RYR1 and TTN*), and genes associated with Thoracic Aneurysm (*MYH11, MYLK, LOX*). Variants in OGDHL and *FLG* were identified in patients with hEDS, Eosinophilic Esophagitis (EoE) and rashes. **Interpretation:** Pediatric patients with suspected hEDS present differently than adults. Joint hypermobility may be more marked in younger patients and improve with age. Young children with an hEDS phenotype may present with hypotonia, allergies and EoE. Identification of specific genetic disorders allowed more appropriate medical, GI and cardiac care. Patients with Neuromuscular disorders received specialized management from neuromuscular programs. Patients with TAAD genes are being more closely followed by Cardiology and underwent appropriate vascular imaging. Careful vetting of symptoms, family history and physical findings tied to molecular testing has allowed the discovery of molecular variants which may have an associated lifelong need for management and medical interventions.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2039W: Genetic Architecture of 22,125 Deafness Subjects across China

Authors:

F. Bu[1], H. Yuan[1], J. Cheng[2], Y. Lu[1], M. Zhong[1], Q. Zhang[1], J. Geng[1]; [1]West China Hosp., Chengdu, China, [2]Weat China Hosp., Chengdu, China

Abstract:

Large-scale studies of individual Mendelian diseases involving tens of thousands of cases have been notably absent. In the Chinese Deafness Genetics Consortium (CDGC) study, we utilized a three-stage genetic testing strategy to conduct a genetic cohort study involving 22,125 subjects with deafness across mainland China to comprehensively assess genetic architecture and attributes. The CDGC cohort primarily consisted of individuals with early-onset (<7 years, 96.0%), severe/profound (>70 dB, 94.7%), and nonsyndromic (98.8%) deafness. We developed a framework for iterative analysis (FIA) and a recessive genotype network (RGnet) to aid in variant classification and curation. An overall diagnostic rate of 49.9% was achieved, encompassing 109 deafness genes. The top ten genes accounted for 88.6% of all diagnoses. The pan-China sample recruitment strategy revealed a latitude gradient distribution of *GJB2* p.V37I carriers, estimating the risk of early-onset severe/profound deafness to be less than 1% in isolated p.V37I homozygotes. A significant correlation between the increased count of recessive genotypes (RGs) and variant pathogenicity was identified and presented through the RG network. Cochlear

implantation outcomes for genetically diagnosed subjects were evaluated using the MAIS and MUSS scales. Outcomes for patients diagnosed with 75 deafness genes were provided to aid preoperative evaluation in clinical practice. In conclusion, the results of the CDGC project underscored the significance of a substantial Mendelian disease cohort in enhancing the capacity, efficiency, and precision of variant interpretation in genetic diagnosis. From a macroscopic perspective, this study provides fundamental data for formulating a cost-effective strategy for genetic screening and the implementation of public health policies aimed at preventing deafness, both in China and worldwide.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2041W: Genetic models and *ASTN1* variant contributions to neurodevelopmental disorders via deficits in astrotactin-mediated neuronal migration

Authors:

J. Levine[1], D. Calame[2], H. Du[2], A. Saad[2,3], A. Mushiba[4], N. Almontashiri[5], P. Gawlinski[6], W. Wiszniewski[7], E. Karaca[8], M. Zaki[3], D. Pehlivan[2], D. Marafi[9], L. AlAbdi[10], F. Alkuraya[11], J. Gleeson[12], J. Lupski[2]; [1]Baylor Coll. of Med./Texas Children's Hosp., Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Natl. Res. Ctr., Cairo, Egypt, [4]King Fahad Med. City, Riyadh, Saudi Arabia, [5]Taibah Univ., AL-MADINAH, Saudi Arabia, [6]Inst. of Mother and Child, Warsaw, Poland, [7]Oregon Hlth.& Sci. Univ, Lake Oswego, OR, [8]Baylor Scott & White Hlth., Dallas, TX, [9]Kuwait Univ., Jabriya, Kuwait, [10]King Saud Univ., Riyadh, Saudi Arabia, [11]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [12]Univ California, La Jolla, CA

Abstract:

Introduction: Astrotactin 1 (ASTN1) functions as a neuronal-glial ligand in the developing brain to promote neurons to migrate along radial glia in neuronal structures with laminar organization (e.g., cortex, hippocampus, cerebellum). In mouse models, disruption of ASTN1 results in neuronal migration deficits and a mild reduction in cerebellar volume. In humans, biallelic variants in *ASTN1* have been reported in a few patients with neurodevelopmental disorders (NDD) with or without brain malformations.
Objective: To characterize an evolving disease-gene association through deep phenotyping of patients with biallelic *ASTN1* variants and to identify three models of multilocus pathogenic variation (MPV) that contribute to disease in this cohort.
Methods: Individuals with biallelic *ASTN1* variants were identified through family-based genomics in a research NDD cohort. Additional cases were identified through GeneMatcher and

international collaborations.

Results: Fifteen patients with NDD from 9 unrelated families were found to have biallelic, ultra-rare, predicted damaging variants in *ASTN1*. Five families had homozygous predicted loss-of-function (LOF) variants, including a recurrent frameshift (NM_001286164.2: c.3159_3160del; p.(Gln1053HisfsTer13)) found in two unrelated Arabic families from Saudi Arabia, suggestive of a founder allele. Three models of MPV were present in this cohort, including triallelic (homozygous *ASTN1* variant in addition to a *de novo* heterozygous LOF variant in candidate gene *KDM3A*), double heterozygous (heterozygous variants in *ASTN1* and *ASTN2),* and double homozygous (homozygous variants in *ASTN1* and candidate gene *TRAK2*) due to distributive absence of heterozygosity (AOH). All patients had developmental delay and/or intellectual disability. Other common features include autism spectrum disorder, attention-deficit/hyperactivity disorder, dysmorphic facial features, seizures, and hypotonia. Regarding brain abnormalities, 8/15 patients had structural abnormalities, including polymicrogyria, lissencephaly, malformation of the hippocampus, and dysgenesis of the corpus callosum and cerebellum.

Conclusion: Pathogenic *ASTN1* variants likely contribute to recessive NDD by disrupting the radial migration of neurons along radial glial cells during neurogenesis, given the role of ASTN1 in the neuron-glia interaction, which is necessary for appropriate neuronal migration. Our cohort characterizes the phenotypic spectrum of *ASTN1*-related NDD and highlights the contribution of MPV to the genetic basis of NDD, including triallelic, double heterozygous, and distributive AOH models.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2042W: Genetic variants in *DDX53* contribute to Autism Spectrum Disorder associated with the Xp22.11 locus.

Authors:

M. Scala[1,2], C. A. Bradley[3,4], J. Howe[3], B. Trost[5], N. Salazar[3], C. Shum[3], M. Reuter[3], J. MacDonald[3], S. Y. Ko[3], P. W. Frankland[4,3], L. Granger[6], G. Anadiotis[7], V. Pullano[8], A. Brusco[9], R. Keller[10], S. Parisotto[11], H. Pedro[11], L. Lusk[12], P. Pojomovsky McDonnell[12,13], I. Helbig[14,13], S. Mullegama[15], U. D. Network[16], E. Douine[17], B. Russell[18], S. Nelson[19], C. Graziano[20], M. Schwab[11], L. Simone[11], F. Zara[21], S. Scherer[3,4]; [1]Univ. of Genoa, Genoa, Italy, [2]IRCCS Giannina Gaslini, Genoa, Italy, [3]The Hosp. for Sick Children, Toronto, ON, Canada, [4]Univ. of Toronto, Toronto, ON, Canada, [5]Hosp. for Sick Children, Toronto, ON, Canada, [6]Randall Children's Hosp., Portland, OR, [7]Randall Children s Hosp, Portland, OR, [8]Univ. of Turin, Torino, Italy, [9]Univ. of Torino, Torino, Italy, [10]ASL Città di Torino, Torino, Italy, [11]Hackensack Univ. Med. Ctr., Hackensack, NJ, [12]Children's Hosp. of

Philadelphia, Philadelphia, PA, [13]Univ. of Pennsylvania, Philadelphia, PA, [14]Children s Hosp. of Philadelphia, Philadelphia, PA, [15]GENE DX, Willis, TX, [16]NIH, Bethesda, MD, [17]UCLA, Los Angeles, CA, [18]UCLA David Geffen Sch. of Med., Los Angeles, CA, [19]UCLA Med Ctr, Los Angeles, CA, [20]AUSL della Romagna, Cesena, Italy, [21]Inst G Gaslini, Genova, Italy

Abstract:

Autism Spectrum Disorder (ASD) exhibits an ~4:1 male-to-female sex bias and is characterized by early-onset impairment of social/communication skills, restricted interests, and stereotyped behaviors. Disruption of the Xp22.11 locus has been associated with ASD in males. This locus includes the three-exon *PTCHD1* gene, an adjacent multi-isoform long noncoding RNA (lncRNA) named *PTCHD1-AS* (spanning ~1Mb), and a poorly characterized single-exon RNA helicase named *DDX53* that is intronic to *PTCHD1-AS*. While the relationship between *PTCHD1/PTCHD1-AS* and ASD is being studied, the role of *DDX53* has not been examined, in part because there is no apparent functional murine orthologue. Through clinical testing, here we identified 7 males and 1 female with ASD from 7 unrelated families carrying rare, predicted-damaging or loss-of-function variants in *DDX53*. Additionally, we identified a family consisting of a male proband with ASD and his affected mother, both harboring a gene deletion exclusively involving *DDX53*. Then, we examined databases, including the Autism Speaks MSSNG and Simons Foundation Autism Research Initiative, as well as population controls. We identified 24 additional individuals with ASD harboring rare, damaging *DDX53* variations, including the same variants detected in two families from the original clinical analysis. In this extended cohort of 24 participants with ASD (21 male, 3 female), we identified 19 mostly maternally-inherited variations in *DDX53*, including 10 missense changes, 2 truncating variants, 2 in-frame variants, 2 deletions in the 3' UTR and 3 copy number deletions. Our findings in humans support a direct link between *DDX53* and ASD, which will be important in clinical genetic testing. These same autism-related findings, coupled with the observation that a functional orthologous gene is not found in mouse, may also influence the design and interpretation of murine-modelling of ASD.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2043W: Genotype/phenotype correlation of MESP2 gene: A 20 years Cohort

Authors:

A. Cornier[1,2,3], E. Albino[4], S. Carlo[5,3,1,6], F. Velez - Bartolomei[7], J. Pascual[8], Y. Algarin[3], W. De Jesus[9]; [1]Genetic Section, San Jorge Children's & Women's Hosp., San Juan, PR, [2]Ponce Hlth.Sicences Univ., Ponce, Puerto Rico, [3]Genetic Diagnostic Group, San Juan, PR, [4]Univ. of

Puerto Rico - MSC, San Juan, PR, [5]Ponce Hlth.Sci. Univ., Cabo Rojo, PR, [6]Univ. Pediatric Hosp., UPR Med. Sci. Campus, San Juan, PR, [7]San Jorge Children s & Women s Hosp., San Juan, PR, [8]Univ. Hosp., Cleveland, OH, [9]Ponce Hlth.Sci. Univ., Ponce, PR

Abstract:

We are presenting the genotype/phenotype analysis of 38 patients with Spondylothoracic dysostosis (STD) due to MESP2 gene mutations and variations. We have followed up this cohort for over 20 years and have been able to document complications and outcomes including pulmonary function tests, developmental assessment and quality of life instruments, morbidity and mortality and its association to their respective MESP2 gene changes. STD, also known as Jarcho-Levin syndrome (JLS), is an autosomal-recessive disorder characterized by abnormal vertebral segmentation and defects affecting spine formation, with complete bilateral fusion of the ribs at the costovertebral junction producing a "crab-like" configuration of the thorax. The condition is prevalent in the Puerto Rican population, although it is a pan ethnic disorder. We have analyzed 38 patients with STD including 22 form Puerto Rico and 16 from other countries in Latin America. Twenty patients were homozygote for the most common (founder effect) mutation in Puerto Rican population, C.307G>T (p. glu103*), nine patients were compound heterozygous for a second nonsense mutation, E230X, or a missense mutation, L125V, which affects a conserved leucine residue within the bHLH region. We identified six additional patient who were combine heterozygotes for the C.307G>T / C.258_262del mutations Patient who were compound heterozygotes including the c.258_261 del, showed the least amount of clinical complications including a mild restrictive pulmonary disease resulting in a better clinical outcome in the neonatal and childhood periods with a mortality rate of less than 6% when compared to the patients who were homozygote for the c.307G>T mutation which have a mortality rate above 40%. Our findings will provide a better comprehension of the clinical outcome and prognosis of patient with MESP2 mutations not only in Puerto Rico but in Latin America.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2044W: Genotypic and Phenotypic Spectrum of maple syrup urine disease: A single center Study

Authors:

X. Yang[1,2], R. Yang[2], T. Zhang[2], D. Tan[2], R. Pan[2], Z. Chen[2], D. Wu[2], C. Chen[2], Y. Xu[2], L. Zhang[2], X. Li[2], Q. Shu[2], L. Hu[2]; [1]Emory Univ., Atlanta, GA, [2]Children's Hosp. Zhejiang Univ. Sch. of Med., Hangzhou, China

Abstract:

Background: Maple Syrup Urine Disease (MSUD) is an autosomal recessive metabolic disorder originating from defects in the branched-chain α-ketoacid dehydrogenase (BCKDH) complex encoded by BCKDHA, BCKDHB, and DBT. This condition presents a spectrum of symptoms and potentially fatal outcomes. Although numerous mutations in the BCKDH complex genes associated with MSUD have been identified, the correlation between specific genotypes remains to be fully elucidated. Aim: Our objective was to unravel the mechanistic intricacies interlinked with these genetic mutations and establish potential links between genotypic alterations and the clinical phenotypes of MSUD. Methods: We analyzed 20 MSUD patients recorded from January 2010 to May 2023. Patients' blood samples were collected by heel-stick through neonatal screening, and amino acid profiles were measured by tandem mass spectrometry. In silico methods were employed to assess pathogenicity, stability, and biophysical properties. Various computation tools were utilized for assessment, namely PredictSNP, iStable, and Align GVGD.Results: We detected 25 distinct mutations, including 12 novel mutations. The BCKDHB gene was the most commonly affected (53.3%) compared to the BCKDHA gene (20.0%) and DBT gene (26.7%). In silico webservers predicted all novel mutations were disease-causing.Conclusions: This study highlights the genetic complexity of MSUD and underscores the importance of early detection and intervention. The integration of neonatal screening with advanced sequencing methodologies is pivotal in ensuring precise diagnosis and effective management of MSUD, thereby significantly improving the prognosis for individuals afflicted with this condition.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2045W: Homozygous*CAPZA2*start-loss variant associated with developmental delay, epilepsy, and infantile hypertrophic cardiomyopathy

Authors:

M. Gandhi[1], R. Duan[1], A. K. Saad[1], H. Du[1], J. Posey[1], R. Gibbs[2], N. Al-Otaibi[3], D. Marafi[4], D. G. Calame[1], J. R. Lupski[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX, [3]Univ. of Ottawa, Ottawa, ON, Canada, [4]Kuwait Univ., Jabriya, Kuwait

Abstract:

Background: *CAPZA2* encodes a subunit of an F-actin capping complex CapZ. Newly formed actin filaments elongate rapidly. This elongation is arrested by the CapZ complex at the fast-

growing end of the actin filament, which allows for the cell to maintain its fine temporal and spatial control over its F-actin distribution. *Capza2* is required for organismal viability in mice, and knockout mice show defects in multiple organ systems. Recently, *de novo* missense variants in *CAPZA2* were identified in non-syndromic neurodevelopmental disorders (NDD) in human. *Drosophila* studies demonstrated the *CAPZA2* missense variants act as partial loss-of-function (LoF) and dominant negative alleles. **Methods:** Proband exome sequencing (ES) was performed on a consanguineous family within a Middle Eastern North African (MENA) NDD cohort. Ultra-rare, predicted damaging homozygous variants were tested for segregation within the family via Sanger dideoxy sequencing. **Results:** We identified a homozygous *CAPZA2* start-loss variant, NM_006136.3:c.3G>C: p.(Met1Ile) in the proband and her two affected siblings. The variant is absent from gnomAD v4.1 and has a CADD v1.7 phred score of 24.7. The proband's four healthy siblings and parents were either heterozygotes or wild type for this nucleotide position. All other candidate variants failed to segregate according to Mendelian expectations. The three affected siblings all exhibited severe developmental delay, normal head circumference, axial hypotonia, appendicular hypertonia, epilepsy, vision impairment, and hypertrophic cardiomyopathy. Brain magnetic resonance imaging (MRI) was either normal or showed prominence of the cerebral sulci, extra-axial fluid spaces, and ventricular system. **Conclusions:** Our studies suggest that biallelic *CAPZA2* LoF variants may cause a neurocardiac syndrome characterized by developmental delay, epilepsy, and hypertrophic cardiomyopathy. Functional validation of the impact of *CAPZA2* start-loss is currently ongoing using patient-derived fibroblasts. These studies will lead to a better understanding of the various allelic variants in *CAPZA2* and their impact on phenotype expression.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2046W: *HUWE1* gene, a prominent player in intellectual disability: a case-series report

Authors:

V. Plaiasu[1], A. Tutulan-Cunita[2], M. Stratan[2], I. Focsa[3], D. Prepelita[3], A. Pavel[2], D. Stambouli[2]; [1]INSMC Alessandrescu-Rusescu, Bucharest, Romania, [2]Cytogenomic Med. Lab., Bucharest, Romania, [3]Univ. of Med. and Pharmacy Carol Davila, Bucharest, Romania

Abstract:

HUWE1 related disorder known also as intellectual developmental disorder, X-linked syndromic, Turner type,
represent a very rare neurodevelopmental condition. Typically males are affected, while females

are asymptomatic carriers, or have a milder phenotype. In generally affected individuals show global developmental delay, dysmorphic features with relative macrocephaly, variably impaired intellectual development and poor or absent speech.
We describe the phenotypic and molecular characterization of 3 unrelated afected cases, two males and one female, each of them carrying a rare variant involving the sequence of *HUWE1* gene identified during whole exome sequencing approach. The authors noted similarities to other reported patients. But two patients differs from those previously described in the literature as they presented few additional particular findings: choroid plexus cyst, arrachnoid cyst in opto-chiasmatic cistern needed to be corrected by surgery. Also the female case shown to be carrier for a rare *HUWE1* variant was severe affected. Additional genetic testing were not performed for her family, but her mother presented with a mild skeletal phenotype and maternal grandmother with hypothyroidism.
A solid phenotype pattern among the affected patiens, both males and females, is still remaining to be established. The study expands the phenotypic spectrum in females with *HUWE1* mutations from X-linked intellectual disability.
<!--
/data/user/0/com.samsung.android.app.notes/files/clipdata/clipdata_bodytext_240605_143305_015.sdocx-->

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2047W: Hyperinsulinism associated with non-coding *HK1* variants

Authors:

A. Ganguly[1], K. Boodhansingh[2], **C. Monahan**[1], V. Sanders[2], W. Sigal[2,1], K. Lord[2,1], C. Stanley[2,1], D. De Leon[2,1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Approximately 30% of children with congenital hyperinsulinism (HI) are negative for pathogenic variants in any of the 9 genes currently associated with non-syndromic forms of HI. Non-coding variants in *HK1* were associated with HI in a large family with a dominant diazoxide-responsive form of HI, first reported by Pinney, et al. in 2008. Recently, additional cases have been described with non-coding variants in *HK1* associated with variable HI phenotypes (Wakeling, et al. 2022).
Under a research protocol, we sequenced a 350bp region in intron 2 of *HK1* in 250 children with genetics negative HI in an effort to identify additional cases related to non-coding *HK1* variants

and to characterize the clinical features of these cases. We identified 12 unique non-coding variants in intron 2 of *HK1* in 13 children with genetics negative HI (13/250, 5.2%). In five cases, the *HK1* variant was inherited from a parent (3 maternal, 2 paternal), one of whom is known to be affected with HI. In eight cases, the *HK1* variant was de novo. The age of presentation of HI ranged from day of life one to 21 months of age. Six of the cases had diazoxide-responsive HI; one of these had resolution of their HI at six years of age while the rest have persistent HI. Seven of the 13 cases were diazoxide unresponsive (54%); five underwent pancreatectomy at ages ranging from 6 months to 3 years of age.
Non-coding variants in intron 2 of the *HK1* gene have now been associated with HI in 28 probands. This data shows that a significant proportion of children with negative genetics in genes currently known to be associated with HI may harbor *HK1* intron 2 variants. Identifying these cases is important for clinical care as well as for assessing recurrence risk for families. Therefore, we have recently established CLIA-certified testing for variants in intron 2 of *HK1* and have screened six additional patients to date, either as targeted gene testing or as part of a comprehensive HI gene panel. One of these was found to have a paternally inherited *HK1* intron 2 variant that had been previously reported by Wakeling, et al. This child presented with HI at 11 months of age and was diazoxide responsive.
The suspected mechanism by which these variants result in HI is due to a disruption in the normal silencing of *HK1* resulting in aberrant postnatal expression in beta-cells. The *HK1* intron 2 variants are located in a region of open chromatin in EndoC-βH1 and possibly interact with the promoter region of *HK1*, based on published datasets. Work is ongoing to elucidate the underlying mechanism by which these variants may result in HI.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2049W: Long-read RNA sequencing reveals Parkinson's disease-associated transcription in oligodendrocytes ⭐

Authors:

J. Evans[1,2], E. Gustavsson[2], K. D'Sa[1,2], C. Anderson[2], M. Ryten[3], S. Gandhi[2,1]; [1]The Francis Crick Inst., London, United Kingdom, [2]Univ. Coll. London, London, United Kingdom, [3]Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

The pathological hallmark of Parkinson's disease (PD) is the loss of midbrain dopaminergic neurons and the aggregation of the protein alpha-synuclein (αSyn), encoded by *SNCA*. However, genetic data and single-cell analyses highlight a significant role for oligodendrocytes in the

pathogenesis of PD. Despite this, oligodendrocytes remain understudied due to the challenges of generating human *in vitro* models. Consequently, the molecular role of oligodendrocytes in PD onset and progression is poorly understood. Here, we optimized an experimental paradigm to generate enriched populations of oligodendrocytes derived from human induced pluripotent stem cells (hiPSC) and characterized their molecular and functional identity. Using enriched oligodendrocyte cultures derived from healthy donors and patients with *SNCA* mutations, we performed deep short- and long-read RNA sequencing to reveal the transcriptional diversity of these cells. We identified novel oligodendrocyte-specific transcripts that are not present in existing annotations, including transcripts mapping to PD-associated genes. We also identify disease-associated transcripts and changes in alternative splicing, with differential transcript usage between control and *SNCA*-mutant oligodendrocytes. Together our paired functional and transcriptomic analyses of oligodendrocytes demonstrate significant increases in the activation of the innate immune response in *SNCA*-mutant oligodendrocytes, characterised by interleukin and interferon signalling pathways. Critically, this inflammatory phenotype in PD iPSC-derived oligodendrocytes is associated with significant splicing dysregulation and occurs at the expense of a myelinating phenotype, with critical myelinating genes significantly downregulated. Thus, we demonstrate not only considerable oligodendrocyte-specific transcriptomic diversity, but show that *SNCA*-mutant oligodendrocytes have an inflammatory phenotype that could suggest a new role in PD pathogenesis.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2050W: Maternal Gonosomal Mosaicism Causing Nail-Patella Syndrome: A Case Report and Implications for Genetic Counseling

Authors:

J. Jang[1], H. Im[1], H. LEE[1], H. Sung[1], S. Cho[1], J-S. Lee[2], M-W. Seong[1]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Jee-Soo Lee, Seoul, Korea, Republic of

Abstract:

Background
Nail-patella syndrome (NPS) is an autosomal dominant disorder caused by abnormalities in the LMX1B gene. This study presents the first reported case of NPS resulting from maternal gonosomal mosaicism, characterized by a partial deletion of LMX1B.
Methods
The proband, a 10-year-old girl, exhibited characteristic symptoms of NPS. Familial whole-exome sequencing (WES) was conducted using peripheral blood samples from the patient, her

sibling, and their parents. Gap-PCR and Sanger sequencing were utilized to confirm the deletion and mosaicism.
Results
WES identified an approximately 2-kb deletion encompassing exons 4-6 of LMX1B in both the patient and her sibling. The mother's data revealed the same deletion in a mosaic form. Gap-PCR and Sanger sequencing verified the heterozygous deletion in both the patient and her sibling and the mosaic deletion in the mother. The variant was classified as pathogenic according to ACMG guidelines. Sequence analysis inferred that the deletion likely arose from a double-strand break followed by microhomology-mediated end joining.
Conclusion
This report documents the first case of familial NPS caused by a heterozygous deletion of exons 4-6 of LMX1B inherited from an asymptomatic mother with gonosomal mosaicism. These findings suggest that gonosomal mosaicism may play a significant role in disease transmission, even among asymptomatic parents. Therefore, there is a critical need to develop more sophisticated genetic counseling techniques that incorporate considerations of gonosomal mosaicism to enhance the management and

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2051W: Molecular insights into the pathogenesis of Chediak-Higashi syndrome and the biology of Lysosomal Trafficking Regulator

Authors:

M. Morimoto, J. Serra-Vinardell, P. Sharma, W. J. Introne, W. A. Gahl, M. V. Malicdan; NIH, Bethesda, MD

Abstract:

Introduction: Lysosome-related organelles (LROs) are cell type-specific cellular compartments, each with a unique morphology, composition, and function; they contribute to diverse processes such as pigmentation, blood coagulation, immunity, and neurological function. Dysfunction in the biogenesis and trafficking of lysosomes and LROs lead to LRO-related disorders including Chediak-Higashi syndrome (CHS), a rare autosomal recessive disorder characterized by partial oculocutaneous albinism, a bleeding diathesis, immunological dysfunction, and neurological impairment. While CHS was first described in the 1940s and Lysosomal Trafficking Regulator (*LYST*) was implicated as the genetic cause in 1996, molecular insights into how LYST deficiency leads to CHS have been elusive. The key challenges include the exceptionally large size of the gene (11.4 kb coding sequence) and protein (429 kDa), and the

absence of a validated antibody.
Methods: To further our understanding of the pathogenesis of CHS and the biology of LYST, relative *LYST* mRNA expression was assessed in several cell types from individuals with CHS and unaffected controls by quantitative PCR. To advance the protein studies of LYST, rabbit polyclonal antibodies against two LYST peptides were developed and tested in parallel with 3 commercially available antibodies. LYST-interacting candidates were identified in disease-relevant cell types, including melanocytes, NK cells, THP-1-derived macrophages, and iPSC-derived i^3Neurons, by LYST immunoprecipitation (IP)-liquid chromatography tandem mass spectrometry (LC-MS/MS).
Results: *LYST* mRNA expression varied widely among different cell types; there was no detectable expression in iPSCs, low expression in fibroblasts and iPSC-derived i^3Neurons, moderate expression in NK cells, and high expression in melanocytes. CHS patient fibroblasts, NK cells, and melanocytes generally had decreased *LYST* mRNA expression compared to unaffected controls. Three of the 5 antibodies exhibited a high molecular weight band that was present in control but absent in melanocytes from individuals with CHS by immunoblot analysis. LYST IP-LC-MS/MS revealed several novel LYST-interacting candidates, including those involved in RNA metabolism, translation, mitochondrial function, and stress granule assembly.
Conclusions: The analysis of LYST expression among different cell types, including those from individuals with CHS, and the preliminary detection of LYST-interacting proteins and the identification of disease-relevant pathways will serve as the bases for understanding the mechanism of disease and the identification of therapeutic targets.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2052W: Multiple approach modeling of Hypermobile Ehlers-Danlos Syndrome to elucidate a genetic etiology

Authors:

M. Kraus[1], B. Laraway[2], K. Velmurugan[1], M. Pavlova[1], A. Rozhok[1], I. Kogut[1], E. Elias[3], G. Bilousova[1], M. Haendel[2]; [1]Univ. of Colorado Anschutz, Aurora, CO, [2]Univ. of North Carolina, Chapel Hill, NC, [3]Children's Hosp. Colorado, Aurora, CO

Abstract:

Hypermobile Ehlers-Danlos Syndrome (hEDS) is an inherited connective tissue disorder that primarily manifests in generalized joint hypermobility, joint instability, and chronic pain; however, many other conditions occur alongside, which negatively affect the quality of life, sometimes in very severe presentations. The molecular mechanism that causes hEDS is

currently unknown together with the heterogeneous presentation of many hEDS patients makes diagnosis very challenging. A combinatorial approach through computable phenotyping and in vitro modeling was utilized to elucidate the etiology of hEDS. The computational phenotyping of hEDS patients leveraged observation codes, clinician notes, and laboratory results found within electronic health records housed in the National COVID Cohort Collaborative Enclave containing over 22.6 million patients. The symptom profile of hEDS using the Human Phenotype Ontology was compared to all other patients in the enclave to identify candidate hEDS patients based on semantic similarity regardless of diagnosis status. Computable phenotypes were transferred to the All Of Us Researcher Workbench to identify candidate causal or modifier variants using genomic data. A biorepository of hEDS patient samples was established from punch biopsies. Somatic cells from patients were reprogrammed into iPSCs and differentiated into skin organoids. Reprogramming cells was necessary for rejuvenation to rid samples of naturally occurring epigenetic contributions from aging and the environment. Primary fibroblasts, iPSCs, and organoid-derived fibroblasts from hEDS patients and control family members were sequenced using Nanopore technology to elucidate genetic mutations and methylation profiles. A combinatorial approach of these computational methods and in vitro modeling aims to unveil the etiology of hEDS to allow patients to receive molecular confirmation at diagnosis and target therapies.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2053W: New candidate genes associated with the development of Pulmonary Arterial Hypertension

Authors:

J. Tenorio[1], L. Miranda-Alcaraz[2], N. Gallego[3], A. Cruz-Utrilla[4], D. Maria Jesús[5], M. López Meseguer[6], A. Moya[7], N. Ochoa[4], A. Parra[3], M. Cazalla[3], C. Silvan[8], P. Arias[3], J. Nevado[2], P. Lapunzina[2], P. Escribano-Subias[4]; [1]INGEMM-HULP, ADNTRO Genetics and BITGENETIC, Madrid, Spain, [2]INGEMM-HULP, Madrid, Spain, [3]INGEMM-HULP-CIBERER, Madrid, Spain, [4]12 de Octubre Hosp., Madrid, Spain, [5]Ramón y Caja Univ. Hosp., Madrid, Spain, [6]Hosp. Univ.rio Vall d'Hebron, Barcelona, Spain, [7]Hosp. Univ.rio La Fe, Valencia, Spain, [8]INGEMM-HULP-, Madrid, Spain

Abstract:

Pulmonary Arterial Hypertension (PAH) is an infrequent disorder characterized by increased blood pressure in the pulmonary arteries, leading to progressive heart failure and premature death if not treated. This study aims to select and validate variants in genes previously associated with PAH and investigate new candidate genes potentially involved in the

development of PAH. A total of 55 clinically diagnosed PAH patients and 21 unaffected family members were analyzed by whole exome sequencing (WES). Variant prioritization was conducted through a bioinformatic custom algorithm. Pathogenic and variants of uncertain significance (VUS) were identified in 30.9% of genes previously linked to PAH. Additionally, three VUS variant were identified in the following four new candidate genes: ATF2, HDAC5, VASH1, and UACA. Variants in ATF2, VASH1, and UACA may influence the hyperproliferation and resistance to apoptosis of pulmonary vascular cells, contributing to PAH development. Furthermore, haploinsufficiency in the post-translational protein acetylation regulatory mechanisms involving the HDAC5 gene may also contribute to PAH development, resulting in an aberrant epigenetic signature that exacerbates the characteristic vascular remodeling process. In summary, we have identified several variants in four candidate genes that can be potentially involve in the pathogenesis of PAH through different molecular mechanisms, although further research is needed to confirm the potential role of these genes in the development of PAH.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2054W: New HSD17B4 gene mutation associated with D-bifunctional protein deficiency in a Puerto Rican patient

Authors:

N. Arciniegas[1], S. Carlo[1,2,3], E. Albino[4], J. Pascual[5], Y. Algarin[3], A. Cornier[2,3,6]; [1]Ponce Hlth.Sci. Univ., Cabo Rojo, PR, [2]Genetic Section, San Jorge Children's and Women's Hosp., San Juan, PR, [3]Genetic Diagnostic Group, San Juan, PR, [4]Univ. of Puerto Rico - MSC, San Juan, PR, [5]Univ. Hosp., Cleveland, OH, [6]Dept. of Publ. Hlth., Ponce Hlth.Sci. Univ., Ponce, PR

Abstract:

Case of a 17-year-old Puerto Rican male who has been under genetic evaluation since age 5 years. Was initially referred to genetics to rule out a mitochondrial disorder due to motor delay and a presumptive diagnosis of myopathy, abnormal EEG and attenuation of basal ganglia at Brain CT scan. Initial evaluation was remarkable for nonspecific elevation of hexadecanoylcarnitine and lactic acidemia. Organic acids, amino acids and carnitines profile as well as chromosome analyses were within normal values. The patient was lost to follow up and retuned to Genetic clinic at age 12 with progression of his delays including speech, motor and intellectual and learning disabilities. Neuromuscular molecular testing was normal as well as the mitochondrial whole exome sequencing studies. Several long chain and very long chain acylcarnitines started to increase after age 13 years. Whole exome sequencing (WES) was requested but took three years for the medical insurance to approve it. WES showed that the

patient is homozygote for HSD17B4 c.1767G>T variant. This variant was reported as a variant of unknown significance (VUS). The HSD17B4 gene is associated to D-bifunctional protein deficiency. Two of the three prediction In silico programs predicted the change to be possibly damaging and deleterious respectively. Very long chain profile confirmed the elevation of docosahexanoic acid. D-bifunctional protein deficiency is a disorder causing progressive deterioration of the central nervous system functions, high forehead, hypertelorism, high arched palate which are consistent with this patient phenotype. Elevations of the very long chain fatty acids and acylcarnitine species is due to the deficiency of the 17-beta-hydroxysteroid dehydrogenase IV that is involved in the peroxisomal beta-oxidation pathway of fatty acids. HSD17B4 c.1767G>T change is indeed pathogenic variant in its homozygous state, presenting most of the classical phenotype associated to D-bifunctional protein deficiency.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2055W: Noncanonical endosomal trafficking mediates GLUT10 ER to mitochondrial targeting to regulate vitamin C homeostasis ⭐

Authors:

Y-C. Lee; Academia Sinica, Taipei, Taiwan

Abstract:

Solute carrier (SLC) transporters, characterized by multiple transmembrane domains and comprising more than 400 members, mediate essential physiological functions by transporting diverse substrates across cellular membranes. However, the regulatory mechanisms governing their transport functions and physiological roles remain largely unclear. Our study discovered a novel pathway regulating the intracellular trafficking of the *SLC2A10*-encoded protein, glucose transporter 10 (GLUT10), which regulates intracellular vitamin C levels. Loss-of-function mutations in GLUT10 cause arterial tortuosity syndrome (ATS), which affects nearly all connective tissues. Our studies reveal that GLUT10 localizes to both ER and mitochondria and transports dehydroascorbic acid (DHA), an oxidized form of ascorbic acid (AA), to maintain AA levels. AA acts as a potent antioxidant and enzyme cofactor essential for collagen structure, contributing to connective tissue development and maintenance. Traditionally, protein targeting to mitochondria and the endomembrane system involves distinct molecular processes. Our investigation into the intracellular trafficking of GLUT10 reveals that it is trafficked from ER-Golgi to mitochondria via the plasma membrane and endosomes. Specifically, GLUT10 is targeted to ER via a signal peptide, undergoes N-glycosylation, and translocates to the plasma membrane. Subsequently, it is incorporated into early endosomes through a tyrosine motif-based pathway

and ultimately arrives in mitochondria via endosome trafficking. Disruption of GLUT10 intracellular trafficking, caused by trafficking domain mutants and disease-causing missense mutations, disrupted intracellular AA homeostasis. Our findings uncover a unique pathway for intracellular trafficking of an N-glycosylated transmembrane protein to mitochondria and highlight the role of GLUT10 subcellular redistribution in intracellular AA homeostasis and the pathogenesis of arterial tortuosity syndrome (ATS).

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2059W: Novel *TAF8* variants in three patients with neurodevelopmental disorder with severe motor impairment, absent language, cerebral hypomyelination and brain atrophy expand the genotypes and phenotype associated with this disorder

Authors:

R. Clark[1], W. Dobyns[2], S. Ramanathan[3], J. Bartley[1], D. Michelson[1], D. Cherukuri[1]; [1]Loma Linda Univ. Children's Hosp., Loma Linda, CA, [2]Univ. of Minnesota, Minneapolis, MN, [3]Loma Linda Univ., San Bernardino, CA

Abstract:

TAF8 encodes 1 of 13 TAFs (TATA-binding protein associated factors) that binds with TBP (TATA-binding protein) to comprise the transcription initiation factor TFIID. TAF8 forms a heterodimer with TAF10 that is necessary for TFIID assembly. TFIID assembles at gene promoters and recruits transcription factors to initiate transcription by RNA Polymerase II for all mRNA-coding genes. Pathogenic variants in *TAF1*, *2*, *4*, *8* and *13* cause intellectual disability (ID). Biallelic deleterious *TAF8* variants have been reported in only 8 patients (5 families) with a neurodevelopmental disorder consisting of severe motor impairment, absent language, cerebral hypomyelination and brain atrophy (OMIM #619972). A recurrent homozygous splice acceptor variant was reported in 6/8 individuals: c.781-1G>A in intron 7 generated a frameshift and mutant protein was undetectable in fibroblasts. We report novel *TAF8* variants and variable phenotypes in 3 males from 2 families. Patients 1 and 2, 18 and 17 year old brothers, have homozygous likely pathogenic ~8.8 kb deletions that include *TAF8* exons 8 and 9, with breakpoints in intron 7 and downstream of the gene. They presented as infants with congenital microcephaly, failure to thrive and poor feeding. Pt 1: HT Z= -3.84, WT Z= -4.47; Pt 2: HT Z= -3.70; WT Z= -4.99. Brain MRI/MRS in Pt 1 (4 mo) showed delayed myelination, mild generalized atrophy and reduced NAA/Cre peak in thalami and basal ganglia. Both boys are nonverbal,

nonambulatory, G-tube dependent, with profound ID, spastic quadriplegia and scoliosis. Pt 1 has glandular hypospadias, myopia, strabismus, nystagmus, metopic synostosis, acute GI bleed, spinal fusion (scoliosis), and febrile seizures. Pt 2 has right hip dysplasia, bilateral radial head dislocation, copious drooling and inconsistent head control. Patient 3, a 6-year-old boy, the smaller of fraternal twins, has mild developmental delay with his first words and walking at 3 yrs, and mild spastic diplegia with toe-walking. Growth parameters at 3 yrs were normal except for mild microcephaly: WT Z= -1.16, HT Z= -1.12, OFC Z= -2.05. At 1 yr brain MRI changes were subtle: borderline large lateral ventricles and thin posterior periventricular white matter. He has *TAF8* compound heterozygous truncating variants: (1) c.806_809delACAC (p.Asn269Thr*29) in exon 8, which is predicted to escape nonsense-mediated decay (NMD) and (2) c.719_720del, (p.Thr240Argfs*7), which is expected to generate a frameshift in exon 7 and undergo NMD. The novel *TAF8* variants and features of our 3 patients expand the known genotypes and phenotypes of this condition. The milder phenotype in Pt. 3 may reflect expression of a partially functioning protein.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2060W: Phelan-McDermid syndrome caused by a single nucleotide deletion: A case study

Authors:

J. Chapin[1], J. Black[2], N. Patel[3], R. Lebel[2]; [1]Norton Coll. of Med., SUNY Upstate Med. Univ., Syracuse, NY, [2]Ctr. for Dev., Behavior and Genetics, SUNY Upstate Med. Univ., Syracuse, NY, [3]Pediatric Endocrinology and Diabetes at Joslin, SUNY Upstate Med. Univ., Syracuse, NY

Abstract:

The *SHANK3* gene (OMIM #606230) encodes a scaffolding protein that is critical for synaptic function and propagation of signals to postsynaptic cells. Disruption of *SHANK3* function results in a rare neurodevelopmental disorder known as Phelan-McDermid syndrome (PMS, OMIM #606232). PMS is characterized by neonatal hypotonia, global developmental delays, and moderate to severe intellectual disability. PMS has also been associated with autism spectrum disorder, aggressive behaviors, seizures, kidney abnormalities, gastrointestinal problems, and mild dysmorphic features. Most cases result from a heterozygous deletion of chromosome 22q13.3, including the *SHANK3* gene. Rarely, heterozygous intragenic loss-of-function *SHANK3* pathogenic variants have been reported. We report a patient with an uncomplicated pregnancy born full term by C-section, with a length of 53 cm (75th percentile) and weight of 3.8 kg (70th percentile). Motor and speech delays were noted by two years old. At

15 years old, the patient presented with developmental disability, autism spectrum disorder, sensory processing disorder, anxiety, aggression, short stature, obesity, ankyloglossia, phimosis, absence seizures, celiac disease, pes planus, and minor dysmorphic features. Genetic testing for Fragile X syndrome and a chromosomal microarray were both normal. Whole exome sequencing revealed a *de novo* heterozygous pathogenic variant in the *SHANK3* gene denoted c.4065_4066del (p.V1357Gfs*4). Only 3% of reported PMS cases (7 individuals) result from *SHANK3* pathogenic variants. This specific variant has been reported in monozygotic twins and one other patient (De Rubeis et al., 2018). This case study expands its phenotype spectrum.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2061W: Population-level biobank sequencing reveals insights into *C9orf72*-associated repeat expansion disorder epidemiology, penetrance and clinical presentation ⭐

Authors:

G. del Angel[1], Z. Fairhurst-Hunter[2], K. Smith[2], K. Carss[2], J. Harrow[2], S. Deevi[2], F. Hu[2], K. Megy[2], Q. Wang[3], S. Wasilewski[2], S. Petrovski[2]; [1]AstraZeneca, Boston, MA, [2]AstraZeneca, Cambridge, United Kingdom, [3]AstraZeneca, Chapel Hill, NC

Abstract:

GGGGCC repeat expansions in gene *C9orf72* are the most common genetic cause of Amyotrophic Lateral Sclerosis (ALS) and related disorders of the ALS/Frontotemporal Dementia (FTD) spectrum. There is large clinical heterogeneity among patients with this expansion, with incomplete penetrance and large age of onset variability.
We used Whole Genome Sequencing (WGS) data from 462,054 European-ancestry UK Biobank participants to study the clinical and genetic heterogeneity of *C9orf72* repeat expansion carriers. We show that the rate of FTD or motor neuron disease symptoms is significantly enriched in the presence of germline *C9orf72* repeat expansions above 60 hexanucleotide units (OR=17.62, p=1.49e-138). We find, however, no statistical evidence of symptom enrichment in repeat expansion bins below 60 hexanucleotide in this cohort, supporting this expansion length value as a likely threshold for pathogenicity.
In accurately diagnosed ALS/FTD patients, the frequency of *C9orf72* repeat expansions closely follows previously reported literature. Out of 936 diagnosed ALS/FTD patients in the UK Biobank, *C9orf72* repeat expansions were present in 89 (9.5%), with other patients presenting known Pathogenic/Likely pathogenic (P/LP) variants in GRN (n=1), TARDBP (n=2), SOD1 (n=5) and VCP (n=3).

We performed a phenome-wide association study, comparing symptom enrichment of *C9orf72* repeat expansion carriers across the phenotypic spectrum. We show that, aside from expected associations with motor neuron disease and FTD, *C9orf72* repeat expansion carriers are commonly misdiagnosed with other related neurological and mental disorders, the most common being Alzheimer's Disease and Vascular Dementia.
C9orf72 repeat expansions in ALS/FTD patients are also associated with earlier recorded symptom onset (mean difference 5.0 years) as well as a shorter survival time from first recorded symptoms (median difference 2.0 years, log-rank p=2.35e-14).
We estimate the genetic prevalence of *C9orf72* repeat expansions at 120.1 per 100,000 individuals (95% CI: 109.9-131.1) in European ancestry individuals. In contrast, we estimate the diagnosed prevalence in cohort at 2.1 per 100,000 individuals if only accurate ICD10 diagnosis codes are considered. This is consistent with literature reported values of ~1.5 to 4 cases per 100,000. However, once other dementia misdiagnoses are considered, the symptomatic prevalence in this population jumps to 9.5 per 100,000 (95% CI: 6.9-13.09).
These results underscore the clinical heterogeneity of *C9orf72* repeat expansion patients and point to the need to refine diagnosis algorithms and develop treatment options for these patients.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2062W: Prevalence of heart disease among adults with recurrent pathogenic copy-number variants

Authors:

H. Shimelis[1], M. T. Oetjens[1], A. S. Berry[1], L. K. Walsh[1], S. A. krueger[1], Y. K. Khalil[2], M. Matsumura[2], C. L. Martin[1], D. L. Ledbetter[3]; [1]Geisinger, Danville, PA, [2]Geisinger Wyoming Valley, Wilkes-Barre, PA, [3]U. Florida Coll. of Med.-Jacksonville, Fernandina Beach, FL

Abstract:

Recurrent pathogenic copy-number variants (pCNVs) that can cause neuropsychiatric disorders, are found in about 1% of the general population. pCNVs also contribute to 10-15% of congenital heart defects (CHD) in children. However, their impact on adult cardiac diseases is not well known. The purpose of this study is to estimate the prevalence of different cardiac diseases in adults with a pCNV (pCNV-positive) as compared to those without a variant (pCNV-negative). The study population (N=127,239), a subset that excluded first-degree relatives, was identified from participants in Geisinger's MyCode Community Health Initiative with linked genomic and electronic health record (EHR). Individuals with one of the 12 pCNVs (1q21.1 deletion (del) and

duplication (dup),15q13.3 del, 16p11.2 proximal del and dup, 16p11.2 distal del, 16p13.11 del, 17p12 del, 17q12 del and dup, 22q11.2 del and dup) were identified through exome sequencing. Diagnosis with one of the five serious heart conditions: ischemic heart disease, angina, myocardial infarction (MI), cardiac arrest, and heart failure was determined using selected ICD9 and ICD10 codes in the EHR. The prevalence (% of pCNV-positive or pCNV-negative individuals with cardiac diseases) and odds ratio (OR) of each of the five or any one or more of the cardiac diseases were calculated using logistic regression after controlling for age, sex, and Body Mass Index (BMI). Compared to the pCNV-negative group, individuals with 1q21.1 del (30.2% vs. 18.9%, p = 0.008), 16p11.2 distal del (29.5.% vs. 18.9%, p = 0.04), or 17p12 deletion (27.3% vs. 18.9%, p = 0.03) had a greater prevalence of cardiac diseases. Of note, two CNVs were strongly associated with new and clinically important disease risk: 1q21.1 del was associated with MI (OR = 5.3; 95% CI: 2.7-9.8; p = 4.13 x 10-7), and 16p11.2 distal del was associated with ischemic heart disease (OR = 8.8; 95% CI: 3.7-18.8; p = 1.30 x 10-7). 16p11.2 distal del was also associated with heart failure (OR = 3.3; 95% CI: 1.4-7.0, p = 3.02 x 10-3). 17p12 del was associated with cardiac arrest (OR=8.7; 95% CI: 1.4-28.2; p=2.75 x 10-3), heart failure (OR=3.2; 95% CI: 1.5-6.5; p=1.53 x 10-3), and ischemic heart disease (OR=3.2; 95% CI: 1.3-7.2; p=6.56 x 10-3). These associations remained significant after correcting for multiple testing (12 tests). A subset of recurrent pCNVs are associated with an increased prevalence of serious heart diseases in adults. More research is needed to validate these associations after controlling for established cardiovascular disease risk factors and investigating the potential effect of congenital heart defects linked to the pCNVs.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2063W: Rare maternal inheritance of *ARFGEF1* variant associated developmental disability: A case report

Authors:

K. Renna, J. Black, N. Brescia, S. Tatum, R. Lebel; SUNY Upstate Med. Univ., Syracuse, NY

Abstract:

Background: Developmental delay, impaired speech, and behavioral abnormalities, with or without seizures (DEDISB) (OMIM #619964) is a neurodevelopmental condition associated with heterozygous pathogenic variants in *ARFGEF1* (OMIM #604141). Previously reported cases have occurred both *de novo* and via autosomal dominant inheritance. Of the inherited cases, most have been passed down from the father. Case Presentation: This male was born to a 22-year-old G2P1>2 mother via a classic C-section. Prenatal ultrasound revealed hydrocephalus, Dandy-

Walker malformation, macrocephaly, and reduced fetal activity. Ventriculoperitoneal and cystoperitoneal shunts were placed at 2 days. At 2 months, he experienced multiple seizures, resulting in a diagnosis of focal epilepsy. At 16 months, he presented to the genetics clinic with motor and speech delays, strabismus, frontal bossing, low-set ears, hypertelorism, down-slanting palpebral fissures, high arched palate, and inguinal hernia. Family history revealed that his mother had been diagnosed with epilepsy, hydrocephalus, and a brain cyst in childhood. His mother also has pulmonary stenosis and numerous psychiatric conditions. The maternal grandmother took coumadin throughout her pregnancy. The mother's symptoms, such as dysmorphic features and hydrocephalus, were initially attributed to Warfarin embryopathy, although this finding is not reported in the medical record. Exome sequencing detected a pathogenic heterozygous variant of maternal origin in *ARFGEF1* (c.4208+2T>C, p.?), corresponding with a diagnosis of DEDISB, also known as autosomal dominant *ARFGEF1*-related neurodevelopmental spectrum disorder. Approximately 19 cases of DEDISB have been reported in the literature. Fine and gross motor impairment, dysmorphic features, strabismus, abnormal brain findings, and epilepsy are features associated with DEDISB, and are seen in this patient. Additional findings in DEDISB include intellectual disability, speech delay, and behavioral challenges, which are also emerging in this patient. This case is only the second report of maternal inheritance of DEDISB. Most reported patients are notable for *de novo* variant (68%) or have paternally inherited variants (26%). While the mother presented with some features of Warfarin embryopathy, it is likely that her clinical presentation is better explained by DEDISB. Conclusion: The novel findings in this case highlight and suggest the need to follow both the patient and the mother for a better assessment of genotype-phenotype correlation in a heritable case of DEDISB.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2066W: Retrospective Analysis of Diagnostic Yield of Genetic Testing for Hereditary Connective Tissue Disorders

Authors:

M. Muriello, C. Griffis, S. Rose, L. Katzmark, S. Ginde, D. Basel; Med. Coll. of Wisconsin, Milwaukee, WI

Abstract:

Hereditary connective tissue disorders (HCTDs) are a heterogeneous group of conditions that affect multiple organ systems, primarily the skin, bones and joints, eyes, and blood vessels. The most serious complications of some HCTDs are vascular aneurysms and dissections, especially

of the aorta. The majority of patients referred for HCTD evaluation have a chief complaint of joint hypermobility, and the vast majority of these patients will be diagnosed with hypermobility spectrum disorder (HSD) or hypermobile Ehlers-Danlos syndrome (EDS) which has no known genetic etiology. Who among this patient population benefits from genetic testing is unclear. One approach is to test individuals who have “red flags” for non-hEDS HCTDs, like vascular aneurysms or dissection or features of Marfan syndrome. We hypothesize that in the absence of “red flags” for non-hypermobile HCTDs the diagnostic yield of molecular genetic testing is low. A retrospective analysis of the results of molecular genetic testing ordered for the evaluation of HCTDs at Children’s Wisconsin between December 2012 and July 2017 was performed. A total of 185 unique probands underwent genetic testing during the timeframe of the analysis. For the entire cohort 10.8% were diagnostic, 46.5% were non-diagnostic (at least one variant reported), and 42.2% were negative (Table 1 and Figure 1). Of the 20 individuals with diagnostic tests, 18 had either a dilated aortic root, family history of a specific HCTD diagnosis or vascular event, or a clinical diagnosis of Marfan syndrome. Clinical information for the 2 individuals who had none of these features is shown in Table 2; both had significant musculoskeletal involvement that is atypical for HSD/hEDS that would be included as “red flags” by current clinical practice. Therefore, the diagnostic yield of genetic testing for patients being evaluated for HCTDs who did not have a personal or family history of “red flags” for non-HSD/hEDS HCTDs was zero. The results support the hypothesis that genetic testing has a very low diagnostic yield for patients presenting for evaluation of HCTDs who do not have “red flag” signs or symptoms on exam, in personal or family history for a non-HSD/hEDS HCTD. Referral to clinicians and clinical centers with expertise in HCTDs is recommended to identify which patients should undergo genetic testing. While this study supports the idea that not all patients referred for HCTD evaluation benefit from genetic testing, it does not show exactly what phenotypes and history increase the diagnostic yield. Future studies should aim to clarify what specific non TAAD and non-Marfan red flags should be used to select patients for genetic testing.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2067W: RNF2-dependent ubiquitination is critical for transcriptional plasticity and neural development

Authors:

J. Sheingold[1], C. Ryan[1], S. Regan[1], M. Mulhern[2], J. Bain[2], J. E. Ploeger[3], S. J. Huang[3], V. Hartill[4], A. Rippert[5], E. Bhoj[5], S. Bielas[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Columbia Univ. Irving Med. Ctr., New York, NY, [3]Marshfield Clinic, Marshfield, WI, [4]Univ. of Leeds, Leeds, United Kingdom, [5]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Proper regulation of gene expression is critical for embryonic development, particularly in the brain. Histone post-translational modifications are one mechanism by which gene expression is regulated to affect developmental and terminal cell fate. Pathogenic variants in genes that encode the proteins that add, remove, and recognize specific histone marks are a genetic basis of neurodevelopmental syndromes. *De novo* heterozygous missense variants in *RNF2* cause the syndromic neurodevelopmental disorder Luo-Schoch-Yamamoto Syndrome (LUSYAM), which is characterized by developmental delay, intellectual disability, facial dysmorphisms, and seizures. RNF2 is a component of Polycomb repressive complex 1 (PRC1) and catalyzes the placement of histone H2A lysine 119 monoubiquitin (H2Aub1). Using a human *in vitro* model of neural development, we show that *RNF2* missense variants disrupt H2Aub1 and fail to repress developmentally important transcription factors that alter the trajectory of neural differentiation and cell fate. We observed abnormal induction of the epithelial-to-mesenchymal transition and differential expression of proneural genes critical for cortical development. These changes provide important molecular and developmental insights into the pathogenesis of clinically relevant features of LUSYAM. These findings also highlight the importance of proper H2Aub1 regulation for cell fate determination during fetal development.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2068W: The analysis of the *MCPH1* gene across curated databases suggest phenotypic expansion

Authors:

O. Oluwole[1,2]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Univ. of Cape Town, Cape Town, South Africa

Abstract:

Background
Pathogenic variants in the *MCPH1* gene are often linked to microcephaly. One should expect that since *MCPH1* is involved in the response to DNA damage in various cells, there would be an association with some other severe congenital disorders. Between 2021 - 2022, pathogenic variants in the *MCPH1* gene in African cohorts of non-syndromic hearing impairment were reported, also, evolutionary changes in *MCPH1* across 150 taxa with unique markers were detected. Recently, the allelic frequency spectrum determination in the gnomAD v2.1.1 dataset for recurrent variants in *MCPH1* revealed rare variants likely to produce unique traits. This follow-up study aims to establish facts that *MCPH1* has phenotypic expansion beyond

microcephaly.
Methods
This study performed SNP density and pharmacogenomics targets analyses from publicly avail *MCPH1* sequences, and investigated the landscape gene expression and transcription factor activity of *MCPH1* in the brain by using the ABI Brain Atlas, Bgee and GTEx RNA-seq data to perform statistical analyses with computational tools to identify functions and patterns with the gene enrichment.
Results
The *MCPH1* SNP density and in silico pharmacogenomics markers showed statistically significant differences in hydrophobicity and hydrophilicity pharmacogenomics targets within 100 - 200kb window adjusted by the active domains based on the previous phylogeny cluster that showed similar substitution rates, divergent lengths, and positive selections in 28 taxa. The functional enrichment revealed gene sets related to DNA binding, neurogenesis, and chromatin condensation. The over-expression of *MCPH1* is noticed in the brain tissue of a smoker vs non-smoker in the ABI microarray data, supported by transcription factor activity that may suggests tissue-specific regulation and sensitivity to somatic or germ-line deleterious variants.
Conclusion
The results from previous studies and the latest findings produced evidence to prioritise *MCPH1* multi-systemic abnormalities beyond microcephaly. Additional pivotal findings on *MCPH1* could reveal new insights and phenotypic discoveries in undiagnosed traits.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2069W: The impact of copy number variants in diagnosis and severity of autosomal dominant polycystic kidney disease

Authors:

S. Heneghan[1,2], E. A. E. Elhassan[3,4], H. Ghani[1,2,5], K. Collins[1,2], J. Hentschel[6], S. Schubert[6], M. Sigg[7], G. L. Cavalleri[1,2,5], P. Conlon[3,4], K. A. Benson[1]; [1]Sch. of Pharmacy and Biomolecular Sci., Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [2]SFI, Ctr. for Res. Training in Genomics Data Sci., Dublin, Ireland, [3]Dept. of Nephrology and Transplantation, Beaumont Hosp., Dublin, Ireland, [4]Dept. of Med., Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [5]FutureNeuro SFI Res. Ctr., Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [6]Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [7]BioMarin Pharmaceutical Inc., Novato, CA

Abstract:

INTRODUCTION: Autosomal dominant polycystic kidney disease (ADPKD) is characterized by cyst development and kidney enlargement, leading to progressive kidney function failure. While most cases are due to short variants in *PKD1* or *PKD2*, some families remain genetically unresolved and have wide intrafamilial variation in disease severity. We set out to determine to what extent copy number variants (CNVs) contribute to 1) genetically unresolved diagnosis and 2) variation in disease severity.
METHODS: Targeted next generation sequencing (NGS) data of 378 ADPKD patients from the Irish Kidney Gene Project were analyzed using a ClinCNV pipeline to identify CNVs impacting cystic kidney genes. CNVs were validated from array comparative genomic hybridization (aCGH) and multiplex ligation-dependent probe amplification (MLPA) of *PKD1* and *PKD2*. Regression models assessed the association of the presence of CNVs and CNV burden (in kilobases) to age at onset of kidney failure and height-adjusted total kidney volume (htTKV).
RESULTS: Disease-causing CNVs were identified and confirmed by MLPA in 13 individuals across 7 families, increasing the diagnostic yield by 3.2%, from 89.2% to 92.4%. ClinCNV detected *PKD1* or *PKD2* CNVs with 100% (25/25 true negative) specificity and 60.00% (15/25 true positive) sensitivity based on MLPA data from 50 samples. aCGH of 5 samples showed 83% (5/6 CNVs) sensitivity for regions with sufficient probe coverage (>=10 probes).
After quality filtering, 9.7% (36/371 individuals) had at least one additional CNV in a cystic kidney gene, identifying sixteen deletions and twenty-four duplications. Twenty-one of these individuals had a pathogenic PKD variant in addition to a non-diagnostic CNV. Regression analyses showed no significant impact of CNVs in cystic kidney genes on age to kidney failure and htTKV, but this is likely due to low sample size of individuals with an additional CNV with data available (n=21 and n=10 for age to kidney failure and htTKV data, respectively).
CONCLUSIONS: ClinCNV effectively identifies disease-causing CNVs in ADPKD, suggesting its clinical utility when initial targeted NGS analysis is inconclusive. Additional non-diagnostic CNVs do not appear to contribute to variation in disease severity of ADPKD. This study underscores the clinical relevance of repurposing targeted NGS using bioinformatic tools for CNV detection to enhance diagnostic utility for ADPKD.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2070W: The p.T1163M variant in *TTBK2* causes ciliary dysfunction

Authors:

K. Kume[1], Y. Tada[1], K. Hosoba[2], T. Miyamoto[3], Y. Izumi[4], H. Kawakami[1]; [1]Res. Inst. for Radiation Biology and Med., Hiroshima Univ., Hiroshima, Japan, [2]Hiroshima Univ., Hiroshima, Japan, [3]Yamaguchi Univ., Ube, Japan, [4]Tokushima Univ., Tokushima, Japan

Abstract:

Background:Spinocerebellar ataxia type 11 (SCA11) is an autosomal dominant neurodegenerative disease caused by *TTBK2* (Tau Tubulin Kinase 2) variants. *TTBK2* gene encodes a serine-threonine kinase that regulates ciliogenesis. Although a number of *TTBK2* variants are reported in SCA11 patients, their molecular pathophysiology remains unclear.
Purpose:The aim of this study is to clarify the association of the variant in *TTBK2* and ciliary function in the pathogenesis of SCA11.
Methods: Exome sequencing was performed for a patient with an autosomal-dominant SCA. The identified variant was validated by Sanger sequencing. RPE cells transfected wild-type, p.T1163M, and 443 amino acids truncated TTBK2 (Halo-tagged) were incubated in serum-free medium for 48 hours. Immunohistochemistry was performed using anti-Halo, anti-Ninein, anti-acetylated tubulin, and anti-Smo antibodies. Smo signal intensity in acetylated tubulin positive cilial area was measured.
Results: We identified a variant *TTBK2*: c.3488C>T,p.T1163M. Compared to WT, the proportion of ciliated cells was decreased in T1163M mutant and truncated TTBK2 overexpressing cells. In addition, Smo localization in the cilia was decreased in cells expressing T1163M and truncated TTBK2.
Conclusions: The T1163M variant in *TTBK2* inhibits ciliogenesis and cilia function as a Shh receptor. Primary cilia may play an important role in the pathogenesis of SCA11.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2071W: *TNXB* mutations in Hypermobile Ehlers Danlos Syndrome and Hypermobility Spectrum Disorder

Authors:

M. Galindo[1], R. Handsaker[2,3], V. Janoušek[4,5], C. Laukaitis[6,7,8]; [1]Genetics Graduate Interdisciplinary Program, Univ. of Arizona, Tucson, AZ, [2]Program in Med. and Population Genetics, Broad Inst. of Massachusetts Inst. of Technology (MIT), Cambridge, MA, [3]Harvard Univ., Boston, MA, [4]Biodviser Ltd, Oldham, Manchester, United Kingdom, [5]Dept. of Pediatrics and Inherited Metabolic Diseases, First Faculty of Med., Charles Univ. and Gen. Univ. Hosp. in Prague, Prague, Czech Republic, [6]Carle Illinois Coll. of Med., Urbana, IL, [7]Carle Hlth., Urbana, IL, [8]Inst. for Genomic Biology, Urbana, IL

Abstract:

Introduction: Pathogenic variants in both copies of *TNXB* cause recessive classic-like Ehlers Danlos Syndrome (EDS), one of the 13 EDS types. Heterozygous *TNXB* mutations are suggested to cause hypermobile EDS (hEDS), the most common type, constituting 90% of EDS cases. Investigating *TNXB* mutations in hEDS is limited with next-generation sequencing (NGS) due to the presence of *TNXA*, a short pseudogene of *TNXB* with distinct deleterious mutations. Short reads cannot reliably map to either gene or pseudogene due to their similarity. The paralogous sequence increases the risk of DNA misalignment, potentially resulting in a dysfunctional *TNXB/A* chimera. We employed long-read sequencing to study *TNXB* variants in the region confounded by *TNXA*. **Methods**: The hEDS Gene Study enrolled individuals diagnosed with hEDS. (n=100), hypermobility spectrum disorder (HSD; n=29) and unaffected family members (n=25). Unaffected and unrelated control samples were also analyzed (n=62). Using a 3' primer unique to *TNXB*, long range PCR was performed on all samples, amplifying the 5.5kb segment of *TNXB* that is paralogous with *TNXA.* The PCR product was sequenced using Oxford Nanopore platform. Alignment was completed with minimap2, and variants were called with Haplotypecaller. **Results**: A heterozygous splice variant (c.12463+2T>C) previously characterized as deleterious was observed in 2 individuals with hEDS, 1 with HSD, 2 unaffected family members and 1 control. This variant did not segregate with hEDS in 2 families. Mutations used to characterize *TNXB/A* chimeras (p.C4058W, p.R4073H, p.S4175N, p.D4172N, and a 120 base pair deletion) were not observed in hEDS/HSD, but p.C4058W was observed in 1 control. Rare variants with deleterious in silico predictions were observed in 2 individuals with hEDS and 1 with HSD and not in controls: p.D3741N, p.G3845S, and p.R3915W. **Conclusion**: This study is the first, to our knowledge, to investigate *TNXB* variants in hEDS/HSD using long read sequencing. Rarely, heterozygous *TNXB* mutations are observed in individuals with hEDS/HSD and may account for a rare EDS subtype that would require further phenotypic characterization. The data from this study identifies trends in *TNXB* variants that will be translated to methods that utilize widely available NGS whole genome and exome sequencing data. This will enable identification of *TNXB* mutations in larger hEDS studies like the hEDS Genetic Evaluation (HEDGE).

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2073W: Understanding neurobehavior in TANGO2 Deficiency Disorder and the role of B vitamins ⭐

Authors:

A. Wilderman[1], J. R. Rideb[1], A. Bajic[2], Y. Cao[1], N. Li[1], C. Miyake[3], L. Zhang[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurobiological Res. Inst., Houston, TX, [3]Texas Childrens' Hosp., Houston, TX

Abstract:

TANGO2 deficiency disorder (TDD) is a rare multisystem disease caused by bialleleic pathogenic variation in the transport and Golgi organization protein 2 homolog (*TANGO2*) gene. It is marked by developmental delay, seizures, ataxia, dystonia, dysarthria as well as acute metabolic and cardiac crises. The neurodevelopmental aspects of TDD appear before 2 years of age, are chronic and have the largest impact on quality of life for TDD patients and their families. The molecular mechanism of TDD is largely unknown and no specific treatment is currently available. Recent retrospective patient natural history study and a Drosophila model have suggested using B vitamin may alleviate metabolic crises and improve neurological symptoms. However, TDD has not been studied in a mammalian model or human relevant cell types, which hampers the use of Vitamin B as a treatment in patients. Here we combined the TANGO2 deficient mouse model and human iPSC neurons with *TANGO2* KO and its isogenic controls to understand the cellular and organismal neurological defects in TANGO2 deficiency and test the role of Vitamin B in these models directly. We have performed neurobehavioral tests on *Tango2* knockout (T2KO) mice and littermate controls under B-vitamin rich (chow) or B-vitamin depleted diets (VBD) as well as VBD supplemented with Vitamin B5. We found that T2KO mice recapitulated impaired motor coordination and balance as TDD patients even on chow diet. Both WT and T2KO mice had significant weight loss on VBD beyond 2 weeks, which ultimately lead to death, however T2KO showed faster decline and earlier death, which can be reversed by Vitamin B5 supplementation, supporting a role of Vitamin B5 in the treatment of TDD. Vitamin B5 improves the general health of both WT and T2KO mice and delay their death without specifically affect any particular neurobehavior test. In a parallel effort, we created TANGO2 deficient human induced pluripotent stem cells (iPSC) by CRISPR editing and differentiated them into neurons (iN) together with isogenic controls. We performed comprehensive transcriptomic, metabolic and functional assays on these neurons. *TANGO2* knockout iN show aberrant expression of genes associated with central nervous system development and neurotransmitter receptor complexes and are enriched for associations with the cerebellum and cingulate gyrus. Using these complementary in vivo model as well as human relevant cell types, we aim to understand the molecular, cellular and organismal pathology of TDD and the role of Vitamin B in its treatment.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2074W: Understanding the genetic contributions to Cerebral Palsy: A comprehensive analysis of the complement system in neuronal development and plasticity pathway

Authors:

C. Zhang[1], Q. Zhou[2], T. Gong[1], J. Ma[1], H. Deng[1], Z. Zhong[1], C. Sun[1], X. Lei[1]; [1]The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China, [2]The First People's Hosp. of Fuzhou, Fuzhou, China

Abstract:

Cerebral palsy (CP) is a complex neurological disorder characterized by motor and postural impairments. Despite extensive studies, the precise genetic mechanisms underlying CP remain elusive. Here we employed genome-wide summary-data based MR (SMR) and Mendelian randomization (MR) analysis to investigate the potential causal role of genes within the complement system in neuronal development and plasticity (CSNDP) pathway in CP pathogenesis. Leveraging summary-level data from large-scale GWAS and QTL studies, we assessed the associations of CSNDP-related gene expression, DNA methylation, and protein abundance with CP susceptibility. We identified several putative causal genes associated with CP risk, including CX3CL1 and TYRO3, both implicated in neuroinflammation and synaptic modulation. Co-localization analysis provided strong evidence for shared genetic variants driving the association of CX3CL1 and TYRO3 with CP risk. Furthermore, druggability assessment revealed the potential therapeutic targets of CX3CL1 and TYRO3, supporting their relevance in CP treatment strategies. Phenome-wide association studies demonstrated no significant adverse effects of drugs targeting CX3CL1 and TYRO3 on other disease traits, suggesting their safety profile. Our findings shed new light on the molecular underpinnings of CP and highlight the potential of targeted interventions within the CSNDP pathway. Further study is warranted to validate these findings and explore personalized treatment approaches for individuals with CP.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2075W: Unlocking Diagnoses: Integrating Whole Genome Sequencing, RNA-Seq and collaborative multi-site research in the Indiana University Undiagnosed Rare Disease Clinic (URDC) cohort

Authors:

K. Liaqat[1], K. Treat[1], L. Mantcheva[1], M. Abreu[1], T-H. Schwantes-An[1], K. Booth[1], R. Ly[1], A. Breman[1], M. Tayeh[1], B. Helm[1], B. Graham[2], S. Ware[2], F. Vetrini[2], E. Conboy[2]; [1]Indiana Univ. Sch. of Med., Indianapolis, IN, [2]Indiana Univ., Indianapolis, IN

Abstract:

Abstract:In the United States, rare disease (RD) is defined as a condition, which affects fewer than 200,000 individuals, or about 1 in 1,600 people in the current population. However, collectively, there are about 7,000 of these disorders affecting 6-8% of the US population, thereby causing a significant public health and socioeconomic impact, in addition to the personal medical, financial, and psychological toll on patients. People living with rare diseases often struggle for years before receiving an accurate diagnosis, and often remain undiagnosed for a decade or longer. The Undiagnosed Rare Disease Clinic (URDC) established in January 2020, is a multidisciplinary collaborative research clinic that focuses on providing genetic diagnosis for patients and families who are dealing with the uncertainty of an undiagnosed rare disease and diagnostic odyssey. Recently (September 2023), URDC has been designated as Indiana University Diagnostic Center of Excellence as the clinical site of Undiagnosed disease network (UDN). The URDC team enrolls only those patients who have been suffering from genetic disorders and are undiagnosed even after extensive genetic testing. Following enrollment of patients at URDC samples are collected during the clinic visit for reanalysis of whole exome Sequencing (WES), research-based whole genome sequencing (WGS), RNA-Seq, and targeted metabolomic testing. We reanalyze the clinically negative ES data on a research basis. Additionally, the GS is also performed, and data is analyzed to identify the causative variant in the intronic region that is not captured by ES. The RNA-Seq data is analyzed to validate the intronic variant. The Artificial Intelligence (AI) based genome analysis platform EMEDGENE is utilized for the analysis of ES, GS, and RNA-Seq analysis. To date, 110 patients have been enrolled in the URDC, with 20% solved or likely solved cases. The remaining undiagnosed patients present a unique opportunity for further research and novel gene discovery. The team developed multi-site collaborations with various research and clinical groups to study the functional effect of genes and variant candidates. The teamwork of URDC is especially important in the state of Indiana as it has been markedly underserved regarding genetic services. Here, we describe the experience of URDC, the analytical strategy and the results obtained in the last 4 years, experience, and learned lessons with URDC as related to ultra RD patients. Out of 110 enrolled patients, 22 cases have been solved, with a 20% diagnostic rate. We summarize the results of this pilot project and emphasize the importance of a sustainable institutional URDC program.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2076W: Unraveling the brain-imaging predictors for cerebral palsy: insights from mendelian randomization analysis of brain imaging-derived phenotypes

Authors:

Q. Zhou[1], T. Gong[2], H. Deng[3], J. Ma[2], Z. Zhong[4], C. Sun[3], X. Lei[3], C. Zhang[3]; [1]The First People's Hosp. of Fuzhou, Fuzhou, China, [2]The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China, [3]The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ.,, Nanchang, China, [4]The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nachang, China

Abstract:

Cerebral palsy (CP) is the primary cause of motor disability in children, with its complex etiology involving genetic and environmental factors. Neuroimaging techniques provide valuable insights into brain abnormalities associated with CP, yet the precise mechanisms remain poorly understood. Investigating causal relationships between brain imaging-derived phenotypes (IDPs) and CP risk could shed light on understanding the pathophysiology and aid in early detection strategies. In this study, a Mendelian Randomization (MR) analysis was conducted on 419 IDPs to explore their causal associations with CP risk. Genetic instruments derived from genome-wide association studies were used to assess the influence of IDPs on CP. Sensitivity analyses employing various MR methodologies ensured the robustness of the findings and mitigated biases. Significant causal relationships were uncovered between eight specific IDPs and CP risk. Increased volume in the left superior frontal gyrus, enlarged cortical surface area of the left precentral gyrus, and increased thickness in the right rostral anterior cingulate cortex were associated with heightened CP risk. Conversely, higher fractional anisotropy (FA) in the left corticospinal tract was linked to reduced CP risk. Elevations in mode of the diffusion tensor (MO) in the fornix, intracellular volume fraction (ICVF) in the genu of the corpus callosum, and isotropic or free water volume fraction (ISOVF) in the left and right inferior longitudinal fasciculi were also tied to altered CP risk profiles. Sensitivity analyses consistently validated these findings, highlighting the utility of IDPs as critical imaging markers for predicting CP. Our findings underscore the importance of IDPs in predicting CP risk and provide valuable insights into its pathophysiology. This study laid the foundation for future research aimed at understanding the mechanisms underlying CP and developing novel diagnostic strategies.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2077W: Unraveling the pathogenesis of SAVI-associated pulmonary fibrosis: transcriptional investigation of iECs and the role of EndMT

Authors:

C. McNinch[1,2,3,4], D. Yang[4,5,6,7], M. Boehm[4,5,6,7], R. Goldbach-Mansky[3,4,8,9]; [1]Bioinformatics and Computational BioSci.s Branch, Bethesda, MD, [2]Office of Cyber Infrastructure and Computational Biology, Bethesda, MD, [3]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [4]NIH, Bethesda, MD, [5]Natl. Heart, Lung, and Blood Inst., Bethesda, MD, [6]Translational Vascular Med. Branch, Bethesda, MD, [7]Lab. of Cardiovascular Regenerative Med., Bethesda, MD, [8]Translational Autoinflammatory Diseases Section, Bethesda, MD, [9]Lab. of Clinical Immunology and Microbiol., Bethesda, MD

Abstract:

Background: STING-associated vasculopathy with onset in infancy (SAVI) is a type I interferonopathy characterized by early-onset pulmonary fibrosis. The underlying mechanism by which STING activating mutations contribute to interstitial lung disease and fibrosis, unique to SAVI among other type I interferonopathies, remains poorly understood. This transcriptomics study characterizes signaling pathway signatures of SAVI patient-derived endothelial cells.
Methods: Three RNA-seq experiments were used to investigate dysregulated gene expression in SAVI patients and STING1 signaling in target cells. First, induced endothelial cells (iECs) from n=4 SAVI patients and n=4 non-affected volunteers were compared. Next, n=2 patient-derived cell lines carrying STING1-N154S/+ were isogenically corrected using CRISPR/Cas9 and iECs were compared to parent line controls. Finally, healthy human lung microvascular endothelial cells were treated with 2,3 cyclic guanosine monophosphate (2,3 cGAMP) to mimic STING1 signaling. Differential gene expression, pathway enrichment and upstream regulator analyses were conducted to identify dysregulated pathways and activated or repressed transcription factors. Gene expression signatures from cell line-derived data were validated using nanoString's GeoMx Digital Spatial Profiler on diseased and non-affected lung tissue.
Results: Expression profiles of SAVI and healthy control iECs were initially similar but diverged during tissue culture propagation. Type 1 interferon response genes accounted for much of the expression differences between early and late passages of SAVI iECs, and between SAVI and healthy control iECs. Notably, pathways related to pulmonary fibrosis were also upregulated in SAVI patient iECs. Additionally, numerous genes involved in the Endothelial to Mesenchymal Transition (EndMT) had altered gene expression, characterized by downregulation of endothelial genes and upregulation of mesenchymal genes. This transcriptional shift towards mesenchymal-like gene expression was also observed in HLMECs treated with STING1 ligand.

Upstream regulator analyses revealed a network of transcription factors involved in EndMT, with STAT3 playing a prominent role.
Conclusions: SAVI patient-derived iECs can be used to capture transcriptional alterations specific to this disease. Transcriptional profiles of these iECs strongly suggest the involvement of a STING-dependent endothelial-to-mesenchymal transition, with STAT3 playing a crucial role. Consequently, targeting the STING/TBK1-STAT3 pathway holds promise as a potential treatment strategy for SAVI.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2078W: Untargeted metabolomics as a potential screening tool for 3-methylglutaconic aciduria syndromes

Authors:

C. Gijavanekar[1], C. R. Difalco[1,2], A. N. Grace[1,2], J. D. Odom[1,3], K. Machol[1,2], L. Emrick[1,2], H. Dai[1,4], Y. Wang[1,4], L. Vossaert[1,4], F. Xia[1,4], S. H. Elsea[1,4], F. Scaglia[1,2,5]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]CHRISTUS Children's Hosp., San Antonio, TX, [4]Baylor Genetics, Houston, TX, [5]Prince of Wales Hosp., Hong Kong, Hong Kong

Abstract:

The 3-methylglutaconic aciduria (3-MGA-uria) syndromes comprise a heterogeneous group of inborn errors of metabolism defined biochemically by detectable elevation of 3-methylglutaconic acid (3-MGA) in the urine. In primary (or type 1) 3-MGA-uria, distal defects in the leucine catabolism pathway directly cause this elevation. Secondary 3-MGA-uria syndromes, however, are unrelated to leucine metabolism-specific defects but share a common biochemical phenotype of elevated 3-MGA. It is currently thought that this accumulation is due to an underlying buildup of acetyl-CoA in the mitochondria from impaired oxidation in the TCA cycle with ensuing formation of trans-3-methylglutaconyl CoA and its subsequent byproducts, including 3-MGA. In these disorders, urine 3-MGA levels are known to fluctuate and at times, are undetectable by standard urine organic acid analysis (UOA), thereby reducing the utility of this biochemical screening method. Here, we retrospectively evaluated a cohort of seven patients with confirmed 3-MGA-uria syndromes. Out of eight total urine samples obtained in this cohort, two UOA analysis reports from two unique participants did not detect 3-MGA. This inherent limitation highlights the need for a more sensitive clinical modality. Untargeted metabolomic profiling is a rapidly emerging technology that is being used to detect and characterize biochemical abnormalities in many inborn errors of metabolism. Untargeted metabolomic profiling performed on plasma samples in this cohort identified significant

elevations of 3-MGA in all seven individuals. This high level of clinical sensitivity demonstrates the promising potential for untargeted metabolomics analysis as an effective biochemical screening tool for 3-MGA-uria syndromes in plasma.

Session Title: Mendelian Phenotypes Poster Wednesday Session

Board 2080W: Von-Hipple Lindau Syndrome - A case report from India

Authors:

H. Singh; Post Graduate Inst. of Med. Ed. and Res., Chandigarh, India

Abstract:

Von-Hipple Lindau Syndrome - A case report from INDIA Dr.Harjot Singh, Dr. Parminder Kaur, Prof. Inusha Panigrahi PGIMER,Chandigarh, INDIA Von-Hipple Lindau (VHL) is an autosomal dominant inherited condition characterized by tumors of various organs such as retinal, CNS hemangioblastoma, renal carcinoma or pancreatic tumors etc. The prevalence of this disorder is 1 in 36000 live births. Tumors may be benign or malignant including various organs. VHL, usually show its symptoms by 65 years of age. The disease is due to VHL gene present on chromosome 3 which is mainly a tumor suppressing gene. The VHL gene consists of three exons and code for VHL protein which is a glycan anchored protein helps in signal transduction. A 23 year old female presented to our department with complaints of visual disturbances and abdominal pain for which she underwent ophthalmological evaluation and was diagnosed to have retinal angiomas and ultrasonography abdomen revealed multiple pancreatic cysts largest 3.7*4.2 cm9s and cystic lesion in left kidney. In view of bilateral retinal angiomas and pancreatic cyst clinical suspicion of VHL syndrome was kept and magnetic resonance of brain was done which show right cerebellar hemangioblastoma, renal function test were normal with blood urea of 20 and serum creatinine of 0.7 mg/dl, with uric acid level of 5.1 mg/dl. For cerebellar hemangioblastoma she was operated with mini-supraorbital craniotomy and excision, medium pressure ventriculoperitoneal shunt placement was done. There was positive history of renal cell carcinoma in her father. With the above constellation of findings and presence of a positive family history clinical diagnosis of VHL syndrome was kept and targeted NGS was done which showed a heterozygous missense variation in exon 1 of VHL gene on chromosome 3: VHL:c.333C>G:p.(Ser111Arg), this variant has been previously reported as pathogenic for VHL syndrome. The lady is in our follow up and screening in children for the same variant was planned. Periodic screening of eyes, hearing evaluation, abdominal screening, neuroimaging and screening for other cancers like pheochromocytoma has been been recommended in all cases of VHL.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3041W: A *de novo* deep intronic variant in *PKD1* adjacent to a common polymorphism creates a novel splice site in an individual with polycystic kidney disease

Authors:

T. Han[1], A. Rehman[1], K. Bergstrom[2], J. Bassetti[2], V. Jobanputra[1]; [1]New York Genome Ctr., New York, NY, [2]Weill Cornell Med. Coll., New York, NY

Abstract:

Utilization of whole genome sequencing (GS) in molecular diagnostics continues to increase, especially for rare diseases. However, despite significant advances in computational methods to predict the effects of non-coding variants, interpretation of intronic variants remains largely speculative, particularly for those located outside of canonical splice sites. Thus, pairing GS with RNAseq has the potential to resolve the molecular consequences of a large number of intronic and noncoding variants detected by GS. We present here a case of autosomal dominant polycystic kidney disease (ADPKD) caused by a *de novo* deep intronic variant in *PKD1* adjacent to a common polymorphism, the combination of which generated a novel dinucleotide splice site and a pathogenic allele. The proband here was clinically diagnosed with polycystic kidney disease in his 20s as an incidental finding during a whole-body scan for an unrelated reason. In his 30s, the patient underwent an abdominal MRI, results of which were consistent with ADPKD with Mayo Classification 1C. Standard genetic testing, including targeted *PKD1*/*PKD2* testing and two different kidney disease panels, returned negative results. GS performed by the Molecular Diagnostic Laboratory of the New York Genome Center identified a *de novo* deep intronic variant adjacent to a paternally inherited common SNP. Individually, neither the *de novo* variant nor the common SNP are predicted to impact splicing, but as a dinucleotide substitution, generate a novel splice donor site deep within intron 16 of *PKD1*. Evaluation of RNAseq data from patient fibroblasts compared to fibroblasts from his father, who carried only the common SNP, revealed inclusion of a 114bp pseudoexon into the *PKD1* mRNA transcript. Although this new exon preserved the reading frame of the transcript, it included an in-frame stop codon that is expected to result in loss of function. RNA data allowed us to upgrade this deep intronic variant to a full pathogenic allele using ACMG classification criteria. This case highlights the utility of RNAseq to clarify the effects of a deep intronic variant, which allowed us to return a

molecular diagnosis to a patient with a clear clinical diagnosis who had undergone years of inconclusive testing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3042W: A deep dive comparing the diagnostic yield and cost of clinical actionable findings from exome sequencing and multigene panel testing for epilepsy

Authors:

T. Williams, D. Pineda-Alvarez, Y-L. Ting, K. Agre, S. Hamilton, E. Chen, F. Facio, S. Aradhya; Invitae, San Francisco, CA

Abstract:

Genetic testing (GT) for epilepsy has diagnostic, prognostic, and therapeutic implications. Despite recent guidelines for GT in patients with epilepsy favoring exome sequencing (ES) over multigene epilepsy panel testing (MGPT), MGPT is still widely used by clinicians. This study compares the molecular diagnosis (MolDx) rate, variant types (VT), clinical actionability (CA), and the cost of clinically actionable findings (CAF) in patients with epilepsy tested by ES or MGPT.

Patients were referred for diagnostic MGPT or ES for epilepsy between December 2015-December 2023. Patients in the ES cohort were selected based on clinician-provided ICD-10 and HPO terms related to epilepsy. Variants were classified using a validated variant classification framework based on the ACMG/AMP guidelines. CA was defined as having specific treatment or actions that could impact disease management or prevention in patients with a diagnostic finding in certain genes. This list of genes was curated from published literature and/or expert opinion. List prices of MGPT and ES were obtained for several commercial and academic laboratories and were averaged. Health insurance medical policies (MP) covering WES and MGPT were reviewed. Odds ratios (OR) and p-values were calculated using G-Tests; p-values<0.05 were considered statistically significant.

The overall MolDx rate in patients was higher in the ES cohort (459/1,582; 29.0%) versus MGPT (9,054/76,575; 11.8%) (OR 3.0, $p<2.2x10^{-16}$). Diagnostic findings were identified by ES and MGPT in 303 and 209 genes, respectively, with 103 (25.2%) genes overlapping. Single nucleotide variants and small indels were more commonly identified by ES than MGPT (64.6% vs 52.0%, p=0.0000001, and 21.9% vs 16.4%, p=0.0115, respectively), while CNVs and low-complexity variants were more common in MGPT than ES (16.5% vs 9.1%; p=0.00001, and 8.6% vs 0.9%; p<0.0000001, respectively). Among patients with a MolDx identified by ES and MGPT, 175/459

(38.1%) and 4,977/9,054 (55.0%), respectively, had results in genes with CA (OR 1.98, $p=1.62x10^{-12}$). The cost associated with identifying one CAF was $31,233.20 and $32,354.09 when using ES and MGPT, respectively. Most MPs for WES require some form of pre-test genetic counseling (PTGC), while MGPT MPs do not require pre-test genetic counseling.
The MolDx in patients with epilepsy tested using ES is higher than those tested via MGPT. Compared to ES, MGPT had a higher rate of results with CA at a similar healthcare cost, and most MPs do not require PTGC. This study may help clinicians choose appropriate GT for patients by considering the variability in VT detection, MolDx rate, CA, cost of CAF, and the need for PTGC.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3043W: Classes of variants contributing to the increased diagnostic yield of whole genome and RNA sequencing

Authors:

G. Pitsava[1], I. De Dios[2], L. Auriga[1], S. Marmolejos[3], V. A. Fusaro[1], C. King[1], J. Lotempio[1], M. Almalvez[1], UCI-GREGoR Consortium, A. Ko[4], E. Délot[3], C. Xiao[1], S. I. Berger[3], E. Vilain[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]Univ. of California Irvine, Irvine, CA, [3]Children's Natl. Hosp., Washington, DC, [4]Children s Natl. Res. Inst., Washington DC, DC

Abstract:

Background: Over the past decade, the extensive use of sequencing technologies has transformed genetic testing and has tremendously accelerated the discovery of new disease-related genes. Despite this progress, however, the diagnostic rate of genetic testing for rare disease is still under 50%. Objective: To evaluate whether previously unsolved cases of individuals suspected to have a genetic disorder but without a molecularly confirmed diagnosis can be solved by deploying new technologies and/or analytical approaches. Methods: We analyzed 355 families without a prior molecularly confirmed genetic diagnosis using whole-genome sequencing. 42.3% (n =150) consisted of proband-only families, 13.2% (n =47) consisted of duos/duos+, 43.4% (n =154) consisted of trios/trios+ and 1.1% (n=4) consisted of proband and sibling. Following variant calling, variant interpretation was performed using the MOON software combined with manual curation and filtering. Results: We found a diagnosis in 12.1% of cases, while in an additional 2.3% of cases we detected a VUS that could explain the phenotype based on the phenotypic association of the gene in which it occurred. The major reasons that initial genetic testing had failed to uncover the diagnosis were: a) the causative gene was not included in the panel of genes initially tested; b) variant reclassification; c) the

variant occurred outside regions covered by whole-exome sequencing. In 6% of the families the causative variant was a deep intronic variant, whose effect on splicing was confirmed via blood RNA-sequencing, while 0.8% of families had a variant in the new disease-causing non-coding gene RNU4-2. Among the cases that received a diagnosis, the most prevalent phenotype was neurodevelopmental abnormalities. Conclusion: Whole-genome sequencing and RNA-sequencing increased the diagnostic yield in patients that remained unsolved after clinical testing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3044W: Classification of meiotic/mitotic origin of chromosome 21 nondisjunction error in Down syndrome using proband-only sequencing: Methodology and application in a cohort with acute lymphoblastic leukemia

Authors:

Z. Li[1], W. Yang[1], G. Wu[1], M. Devidas[1], M. Shago[2], A. J. Carroll[3], N. A. Heerema[4], J. Gastier-Foster[5], B. L. Wood[6], L. Sanclemente[5], E. A. Raetz[7], S. P. Hunger[8], M. L. Loh[9], E. Feingold[10], T. C. Rosser[11], E. G. Allen[11], S. Sherman[11], K. R. Rabin[5], P. J. Lupo[5], J. J. Yang[1]; [1]St. Jude Children's Res. Hosp., Memphis, TN, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]Univ. of Alabama at Birmingham, Birmingham, AL, [4]Ohio State Univ., Columbus, OH, [5]Baylor Coll. of Med., Houston, TX, [6]Children's Hosp. Los Angeles, Los Angeles, CA, [7]New York Univ. Langone Med. Ctr., New York, NY, [8]Children's Hosp. of Philadelphia, Philadelphia, PA, [9]Univ. of Washington, Seattle, WA, [10]Univ. of Pittsburgh, Pittsburgh, PA, [11]Emory Univ., Atlanta, GA

Abstract:

Background: Down syndrome (DS), the most common chromosomal disorder, is caused by nondisjunction (NDJ) of chromosome 21 that occurs during meiosis I (MI) or meiosis II (MII) in parental gametogenesis, or during post-zygotic mitosis. Tracing of the origin of chromosome 21 NDJ to MI, MII or mitotic error is essential for the understanding of the etiology and epidemiology of DS and may yield insights into associated medical conditions. Conventionally, this classification is performed using proband-parent trio genotyping, limiting its applicability to DS proband-only data. **Methods:** Here, we developed a chromosome 21 NDJ error classification method that uses only genotyping of DS probands, by combining informative rare single-nucleotide polymorphisms (SNPs) and rare haplotypes in hidden Markov models. Rare SNPs and rare haplotypes are defined based on gnomAD and 1000 Genomes Project, respectively. Our

method segments chromosome 21 into blocks with two or three grandparental origins, upon which the NDJ errors and crossover events can be identified at high resolution. **Results:** When compared to chromosome 21 NDJ error classification based on an analysis of 152 DS proband-parent trios who underwent whole genome sequencing (WGS), our method achieved an accuracy of 95.4%. In crossover events identification, our method demonstrated a recall of 93.7% and a precision of 100%. In a dataset of 764 DS probands who underwent WGS, frequencies of MI, MII, and mitotic errors identified by our method are consistent with prevalences estimated using trios in prior reports. Further applying the method to 519 DS probands with acute lymphoblastic leukemia (ALL), we found that the type of NDJ error was associated with overall survival, with patients with mitotic error demonstrating significantly poorer outcomes (5-year overall survival 68% vs 90%, P=0.0043). The type of NDJ error was also associated with leukemia subtype, with MII errors making up a significantly greater proportion of *CRLF2*-rearranged *BCR*::*ABL1*-like ALL than of other genetic subtypes (32% vs 17%, P=0.026), possibly due to the link between this subtype and the risk of specific regions on chromosome 21 being reduced to homozygosity as a result of the MII error. **Conclusion:** We propose a novel method for chromosome 21 NDJ error classification and crossover event identification that can be applied to study the etiology of aneuploidy in cohorts without availability of parental samples. We further demonstrate the potential for this method to generate insights into DS-associated medical conditions, reporting here on associations of NDJ error type with survival and genomic subtype in a cohort with DS and ALL.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3046W: Deciphering the genetic mystery of Congenital Nephrotic Syndrome

Authors:

A. Alshehri, S. Escudero, K. Delgado, L. Thompson; Nicklaus Children's Hosp., Miami, FL

Abstract:

Introduction

Congenital nephrotic syndrome (CNS) is a rare disorder characterized by proteinuria, hypoalbuminemia, and edema in infancy. Congenital Nephrotic Syndrome of the Finish type (CNF), a notable subtype often caused by biallelic NPHS1 variations, is associated with early onset nephrotic syndrome and elevated AFP in the prenatal period of both carriers and affected fetuses.

Case description

This case began prenatally with elevated AFP levels. After ruling out other causes, a distant maternal family history of CNF was noted. Due to this history, targeted molecular testing via amniocentesis was recommended. Testing revealed a heterozygous, maternally inherited, likely pathogenic mutation in NPHS1 (c.847>T p. (GIn283Ter)). No second mutation was found. These results suggested that the elevated AFP was due to the fetus being a carrier, not affected by CNF. However, almost immediately postnatally, the proband presented with features of nephrotic syndrome. Due to a consistent phenotype, further genetic evaluation including chromosome microarray and a nephrotic syndrome panel was recommended to both address gaps in testing and broaden evaluation for other inherited nephrotic syndromes. Results only noted the same NPHS1 variant. No other single gene variants nor copy number variants found. Given the ongoing concern for CNF, whole genome sequencing (WGS) was offered. WGS noted a 4.8kb terminal deletion of uncertain significance involving the 3' UTR of NPHS1. This deletion initially could not be confirmed due to limitations of the lab's validated reporting size threshold, so a follow-up targeted PCR was performed confirming it's presence. Father deferred testing to confirm phase, however mother was not found to carry this deletion.

Discussion

CNF is an autosomal recessive condition that presents with prenatal elevated AFP and postnatal CNS. Carriers of NPHS1 variants can have the prenatal findings, but they remain asymptomatic postnatally. Although many of the variations are known, there are always limitations to testing. The terminal deletion in this case was undetected by conventional testing as it was too small for routine chromosome microarray and too large for gene sequencing. This highlights the well-known gaps in genetic testing. New technologies, such as optical genome mapping, are currently being brought to clinical use to try and limit these testing gaps. Such new technologies should be considered in similar clinical situations in the future. This deletion remains a VUS, partly due to lack of paternal testing, however clinical presentation alongside variant interpretation raises high concern for a diagnosis of CNF for this patient.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3047W: Deciphering the genetic terrain of dilated cardiomyopathy: A study of 100 cases from the Indian sub-continent

Authors:

A. Goswami, N. Alam, V. Gupta, Y. Gautam, V. Ojha, S. Kumar, R. Ray, S. Seth, S. Arava; AIIMS, New Delhi, India

Abstract:

Background: Dilated Cardiomyopathy (DCM) is a genetic disorder involving a heterogeneous group of cardiac muscles and is characterised by ventricular dilatation, impaired systolic function, and reduced myocardial contractility with left ventricular ejection fraction (LVEF) being less than 40%. It may lead to sudden cardiac death too. Multiple environmental and genetic factors are known to cause DCM. Several candidate genes have been reported to be associated with DCM, with majority of the mutations being found in the TTN and LMNA genes.
Aim and Objective: The present study aims to unravel the genetic spectrum of variants implicated in RCM in India. **Method:** 100 DCM patients along with their first-degree family members, if available, were enrolled in the study after taking their written informed consent. Detailed clinical evaluation, ECG, Echocardiography screening and/or cardiac MRI were performed in the probands along with endomyocardial biopsy, if required, to identify the RCM phenotype. For molecular study, 5 ml peripheral blood in EDTA vial was collected. Whole exome sequencing was performed in 42 cases on the Illumina Novaseq 6000 platform. In silico analysis was used to make the predictions about the variant pathogenicity. Validation of the prioritised variants in probands was performed using Sanger sequencing. **Results:** Among the probands, the mean age of onset was 32.49 years with the majority (50%) of affected individuals being in the age group of 18-45 years ranging from 1-70 years. 40% of the patients belong to NYHA class II followed by class III (41%), IV (13%) and I (6%). Out of 100 cases, genetic analysis was carried out in 42 cases. Analysis revealed 43 variants in 24 genes in 33 cases, of which 42 were compound heterozygous and 1 was homozygous. Most of the variants were observed in the TTN (n=9) gene followed by MYH7 (n=4) gene. 19.04% of variants were predicted to be pathogenic/likely pathogenic and 80.9% were predicted to be variants of uncertain significance. Consequently, 3 probands with confirmed genetic diagnosis underwent heart transplantation. Upon screening of the family members (n=146) of 49 available families using clinical workup, 26.5% (n=13) families with disease were observed. Members of 9 families were unaware of the condition and were found upon screening whereas 4 families reported members with already known diseases of variable severity. These individuals are being followed up and have been educated about the implications, hence, providing preventative care. **Conclusion:** Comprehensive phenotype and genotype analysis of 45 cases reveals the spectrum of restrictive cardiomyopathy in India.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3048W: Delving into the abyss: Integrating multi-omics and targeted functional studies to inform variants of uncertain significance

Authors:

S. Mokashi, J. A. Cooley Coleman, A. Kandhadai, N. Sahajpal, C. Jones, J. Lee, M. Friez; Greenwood Genetic Ctr., Greenwood, SC

Abstract:

Although the recent widespread use of whole genome sequencing (WGS) as a first-tier diagnostic test has increased diagnostic yield, it has also resulted in larger numbers of variants of uncertain significance (VUS's) being reported that pose a challenge to clinical teams, patients, and their families. Potentially diagnostic variants in introns and untranslated regions often end up entering a black hole of variant classification, with few ways to resolve their clinical significance without investing in time-consuming and expensive functional studies. Here, we first show our laboratory's traditional strategy for resolving variants predicted to have effects on splicing through RT-PCR based studies on RNA extracted from patient blood samples. We demonstrate the efficacy of this strategy through representative cases. This strategy, however, has limitations. For example, a deep intronic variant not predicted to affect splicing would likely not be a candidate for such functional studies. RNA-seq based discovery of aberrant splicing events becomes invaluable in such cases. Therefore, we next use case examples to highlight situations in which we leveraged a range of advanced methodologies, including RNA-seq, optical genome mapping, long-read genome sequencing and targeted functional studies, to discover and resolve variants that affect splicing or expression levels. Genome sequencing identified two *OTUD6B* gene (associated with autosomal recessive intellectual disability) variants in a patient: a Likely Pathogenic deletion and a deep intronic VUS. By employing targeted cDNA sequencing, RNA-seq, transcript-specific qPCR, and western blots, we resolved the classification of the deep intronic variant, resulting in a definitive diagnosis for the patient. For another patient, direct analysis of RNA-Seq data for outlier events identified exon 2 of the *PSMB1* gene as a hotspot of significant exon skipping. A rare variant located 21 bp from the acceptor site was likely responsible. This variant was not believed to be diagnostic for the patient due to limited overlap between their clinical presentation and the *PSMB1*-associated syndrome; however, it is an example of how transcriptome-directed studies could identify candidate variants that were not prioritized by whole-genome analysis due to limited efficacy of currently available annotation tools. In the era of WGS, resolving VUS's poses a formidable challenge. However, by adopting a comprehensive, multidimensional approach encompassing diverse techniques and methodologies, we can navigate more readily through the complexities of genetic variation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3049W: Development and assessment of bioinformatics approach for whole genome sequencing-based newborns screening in the Generation Study ⭐

Authors:

D. Kasperaviciute[1], J. Kaplanis[1], P. Sivakumar[1], D. Deen[1], M. Gomes[1], I. Leong[1], J. Skelton[1], P. Quinn[1], A. Parrish[2,3], E. Baple[4], A. Need[1], K. Stone[1], A. Pichini[1], D. Bick[1], E. Thomas[1], R. Scott[1]; [1]Genomics England, London, United Kingdom, [2]Royal Devon and Exeter Hosp., Exeter, United Kingdom, [3]NHS Exeter Genomics Lab., Exeter, United Kingdom, [4]Univ. of Exeter, Exeter, Exeter, United Kingdom

Abstract:

Whole genome sequencing has the potential to screen for hundreds of rare genetic conditions in a single test providing opportunities for earlier treatment. The Generation Study explores the benefits, challenges, and feasibility of sequencing the genomes of 100,000 babies in England, complementary to current newborn screening; as well as use of babies' genomes for discovery research and over their lifetimes.

214 severe, early onset conditions caused by genetic changes in 472 genes where early, equitably accessible interventions could improve outcomes were selected for the study. Following operational feasibility and genome data quality studies in 40 adults and 600 babies, cord blood was chosen as a primary sample type and capillary blood as a secondary.

To achieve high positive predictive value and minimise the need for manual review of the genomic results, we have developed an automated variant prioritisation approach. Known pathogenic variants from ClinVar, a commercially procured dataset, curated variant lists, and predicted loss-of-function variants (where appropriate) will be prioritised. We have modelled this algorithm in a cohort of ~35,000 individuals of mixed ancestries representative of British population. Highest numbers of prioritised non-reportable variants were observed in genes associated with multiple phenotypes, multiple modes of inheritance, overlapping other disease genes and containing hypomorphic variants. After further optimisation of the algorithm and variant exclusion list, we reassessed specificity in 5,855 genomes processed through the bioinformatics pipeline that will be used in the Generation Study. Several genes were removed from the study due to predicted high number of false positives. Using this automated approach, we estimate ~3% of the genomes will need manual review of prioritised variants, while ~1% of cases are expected to have reportable variants.

Study-wide sensitivity, assessed in 229 patients from our internal dataset of patients with rare disease, reached 70%. The majority of non-prioritised variants were missense variants not

present in known variants datasets. We have also assessed variant callability in all study genes based on genome alignment quality and coverage.
In summary, recognizing the unique context of genome sequencing for population screening purposes in newborns, we have developed a bioinformatics pipeline with high expected specificity and acceptable sensitivity. We demonstrate the importance of analytical validity assessment before genes are added to the screening list, as well as flexibility to make rapid adjustments to gene and variant lists based on emerging evidence.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3050W: Development and Scaling of a Sensitive NGS-based Assay for Clinical Diagnosis of Acute Meningitis and Encephalitis

Authors:

P. Trefry[1], T. Desmet[1], N. Lennon[1], J. Abreu[1], M. Dasilva[1], W. Brodeur[1], E. McDaid[1], R. DePina[1], J. Harold[1], A. List[1], O. Conroy[1], C. Milbury[2], E. Crawford[2], A. Uehara[2], F. Wang[2], S. Miller[2], B. O'Donovan[2], N. Gimbrone[2], A. Carretta[2]; [1]Broad Clinical Labs, Cambridge, MA, [2]Delve Bio, Boston, MA

Abstract:

Meningitis and encephalitis are two forms of acute neurological illnesses that can lead to long-term disabilities or even death. Currently, these infections are diagnosed through a differential diagnostic approach which leverages the patient's history, clinical representation, imaging, and serial laboratory tests. This traditional approach can be time-consuming, costly, and only identifies the causative pathogen in approximately 50% of acute cases. Additionally, many of these tests often come back negative, resulting in prolonged time between admission to diagnosis and associated treatment course. Leveraging next-generation sequencing provides an opportunity to create a single, highly-sensitive pathogen-agnostic assay for routine clinical care, but attempts to do so have been restrained by difficulties scaling the workflow, high costs, long turnaround time, and challenges with clinical interpretation.
DelveBio and the Broad Clinical Labs have collaborated to develop a scalable and sensitive metagenomic NGS-based workflow that provides a diagnosis in a timeframe equivalent to conventional send-out testing. Cerebral spinal fluid is processed through a fully automated workflow, generating dual-indexed Illumina-compatible sequencing libraries derived from both DNA and RNA. Single-end sequencing is performed on the resulting libraries using the Illumina NovaSeq 6000 instrument on an SP flowcell. Data is then processed through Delve's analysis pipeline, which aligns non-human reads to a curated version of the Genbank database of all

known and potential pathogens.
Here Broad Clinical Labs and DelveBio present a sensitive, scalable, and fully automated diagnostic workflow which enables rapid pathogen-agnostic screening for clinically relevant and causal pathogens of acute neurological illnesses. Leveraging NGS has expanded the number of pathogens screened for in a single test, reducing the number of tests required to correctly diagnose patients, while maintaining a turnaround time similar to or better than current send-out testing. By connecting physicians with this technology, patient outcomes are improved via shortened hospital stays and reduced time between admission and treatment, both of which reduce overall costs.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3051W: Diagnostic yield of rare genetic disorders applied genome sequencing as a first-line test among 2,188 rare disease patients in Korean population

Authors:

Y. Kim[1], **Y. Kim**[2], H-Y. Chae[3], B-J. Kim[1], H-Y. Park[2], M-H. Park[1]; [1]Div. of Genome Sci., Dept. of Precision Med., NIH, Cheongju, Korea, Republic of, [2]NIH, Cheongju, Korea, Republic of, [3]Div. of Bio Bigdata, Dept. of Precision Med., NIH, Cheongju, Korea, Republic of

Abstract:

The falling price of sequencing tests and advancement of bioinformatic technologies are increasing the diagnostic yield for undiagnosed rare diseases by analyzing the genome sequence of patients. Large-scale genomic studies of rare diseases have been carried out in previous studies, however, there is lack of research in Asian populations. First pilot study of National Project of Bio Big Data planned to recruit 5,000 rare disease probands and families for clinical diagnosis of rare diseases using genome sequencing in Korea, a total of 4,978 genomes from 2,188 probands and 2,790 family members were sequenced. The two bioinformatic pipelines, panel-based analysis and untargeted analysis were used to screen candidate causal variants of disorders. Prioritized candidate causal variants were represented by considering genotypes, clinical symptoms, and phenotypes (Human Phenotype Ontology terms) through expert clinicians' review and interpretation. Overall diagnostic yield was 35.0% including 491 diagnostic genes in the 19 categories of rare disorder in combining the panel-based and untargeted analysis. The overall diagnostic yield was influence by probands age, enrolled number of family members, as well as the diagnostic yield was different across the disease categories. In addition, the untargeted analysis represented higher diagnostic yield compared to the those of panel-

based analysis, the increasing yield was come from the differences that types of identified variants, candidate gene prioritization, and variant classification between the two analytic pipelines. In conclusion, our results showed the diagnostic yield including diagnostic genes and variants in the participants from the National Project of Bio Big Data in Korean. These results can support the next large-scale genomic studies on the rare diseases and give an insight to the genomic medicine.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3052W: Evaluation of Illumina Complete Long Reads for Genome-Wide Variant Discovery

Authors:

M. Kelly, E. A. Werren, J. M. Ewald, H. Tjong, P. Vats, M. A. Comerford, R. Gorham, M. Pechulis, M. D. Adams; The Jackson Lab. for Genomic Med., Farmington, CT

Abstract:

INTRODUCTION:Whole genome sequencing is revolutionizing rare disease discovery and diagnosis, but standard short-read (SR) technologies are limited in their ability to accurately and reliably detect structural variants >50bp (SVs). The advent of long-read (LR) sequencing from PacBio and Oxford Nanopore has brought further advancements. This study aimed to assess the efficacy of the newly released Illumina Complete Long Reads (ICLR) for variant discovery across multiple variant classes.
METHODS:Four reference DNA samples (NA12878, NA24385, NA02944, NA14724) and two paired venous blood and dried blood spots (DBS) were used. DNA and venous blood underwent PacBio HiFi sequencing to ≥30X mean coverage (DBS yield was insufficient for HiFi). Data were processed using read-based (pbmm2, pbsv, Sniffles2, DeepVariant) and assembly-based (hifiasm, PAV) methods. All specimens were prepared using the Illumina DNA PCR-Free library prep to generate “unmarked” reads as well as the ICLR and ICLR Human Comprehensive Panel (HCP) prep for “marked” reads. Standalone SR-WGS (≥30X coverage), SR-WGS+ICLR, and SR-WGS+ICLR-HCP were processed using DRAGEN. All data were annotated and analyzed in Emedgene using hg38.
RESULTS:Using Genome in a Bottle (GIAB) truth set data (n=2), both ICLR methods performed similarly to SR-WGS for SNVs and indels. HiFi outperformed both ICLR methods on SNVs (F-measure=99.92% vs 99.53%), while ICLR performed slightly better than HiFi on indels (98.89% vs 98.79%). Compared to HiFi (n=6), ICLR methods improved performance over SR-WGS alone for SNVs (95.08% vs 89.74%), indels (85.44% vs 84.92%), and SVs (39% vs 36.74%). Relative to

venous blood performance for SR-WGS and ICLR, DBS (n=2) demonstrated similar or reduced concordance to paired venous blood run on HiFi for SNVs (SR-WGS=90.16%, ICLR=90.14%), indels (SR-WGS and ICLR=85%), and SVs (SR-WGS=35.65%, ICLR=37.96%). Within Emedgene, the number of candidate variants prioritized was significantly higher for HiFi than ICLR (12 vs 7.5 for most likely candidates and 223.7 vs 95.4 for candidate variants).
CONCLUSION:ICLR outperformed SR-WGS alone. ICLR and HiFi perform similarly compared to GIAB truth sets. Concordance of SVs for ICLR was low; however, it is unclear to what extent this is due to a high false positive rate for HiFi and/or low true positives rate for ICLR, demonstrating a need for high-confidence SV truth sets. Overall, ICLR provides a viable approach to LR-WGS, with particular application for low-input specimens.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3053W: Genome sequencing versus exome sequencing in undiagnosed pediatric genetic diseases: An updated systematic literature review and meta-analysis of diagnostic and clinical utility

Authors:

J. Belmont[1], R. Pandey[2], N. Brennan[2], S. Peng[3], K. Trachana[3], S. Katsandres[2], O. Bodamer[4], D. Veenstra[5]; [1]Baylor Coll. of Med., Houston, TX, [2]Curta Inc, Seattle, WA, [3]Illumina Inc, San Diego, CA, [4]Boston Children's Hosp., Dedham, MA, [5]Univ of Washington, Seattle, WA

Abstract:

Background: While previous meta-analyses of studies of pediatric genetic disorders have found a greater diagnostic yield for genome-wide sequencing (GWS), including whole genome and whole exome sequencing (GS/ES), compared to chromosomal microarray, additional recent studies merit the evaluation of GS vs ES by clinical indication and line of testing. The aim of this systematic literature review (SLR) and meta-analysis was to evaluate the diagnostic and clinical utility of GS vs ES in undiagnosed pediatric genetic diseases. **Methods:** The SLR and meta-analysis were conducted per the PRISMA and MOOSE guidelines. We searched MEDLINE and EMBASE databases to identify studies published in English language between January 2013 and May 2023. Outcomes included diagnostic utility (proportion patients identified with pathogenic/likely pathogenic [P/LP] variants per ACMG guidelines) and clinical utility (proportion of patients with a positive diagnosis after GS/ES who experienced a change in clinical management). We conducted a meta-analysis on the diagnostic yield of GS vs ES by line of testing and clinical indication. We report results from the random-effects model given the substantial heterogeneity observed (>60%). **Results:** We identified 108 studies including a total

of 24,631 probands (≤18 years) with diverse clinical indications. Pooled diagnostic yield for GWS was 34.2% (95% CI 27.6, 41.5) vs 18.1% (95% CI: 13.1, 24.6) for non-GWS. Studies reporting within-cohort comparisons suggested 2.4-times higher odds of diagnosis by GWS vs non-GWS (95% CI: 1.4, 4.0). Based on the different clinical indications for testing, studies reporting within-cohort comparison between GS suggested 1.7 times odds over ES (95% CI: 0.94, 2.92). Pooled diagnostic yield was 30.6% (95% CI: 18.6, 45.9) for GS and 23.2% (95% CI: 18.5, 28.7) for the ES. In the first line, diagnostic yield was qualitatively higher for GS than ES among critically ill (38.7% and 32.8%), not-critically ill (39.9% and 38.6%), neurologic conditions (56.9% and 40.0%), multiple congenital anomalies (48.6% and 38.8%), and infants ≤1 year of age (37.7% and 28.1%). Clinical utility for GS was 58.7% (95% CI: 47.3, 69.2) and ES was 54.5% (95% CI: 40.7, 67.6). **Conclusion:** Overall, the results of our meta-analysis indicates that GWS has a superior diagnostic utility over non-GWS in pediatric patients with suspected genetic diseases and suggests a higher diagnostic yield for GS than ES. Particularly, the evidence favors GS over ES in first-line testing for critical illness, congenital anomalies, neurologic condition, and in infants ≤1 year of age, advocating for its consideration as a primary genomic diagnostic tool.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3054W: How to gain a sex chromosome: Monosomy X to the rescue

Authors:

J. O'Shea[1], K. Rudd[2], E. Andersen[3], B. Hong[3]; [1]Univ. of Utah / ARUP Lab., Salt Lake City, UT, [2]ARUP Lab., Durham, NC, [3]ARUP Lab., Salt Lake City, UT

Abstract:

Turner syndrome (TS) is characterized by one X chromosome with the absence of all or part of a second sex chromosome. Approximately half of individuals with TS have complete monosomy X (45,X; MX) whereas half have 'variant karyotypes' with deletions, isochromosomes, or rings involving the sex chromosomes.
Mosaicism is common in TS, where there is one MX cell line and another cell line containing a second normal or abnormal sex chromosome(s). ~7% of TS individuals have a cell line with one normal X chromosome and one normal or abnormal Y chromosome (e.g. 45,X/46,XY; 45,X/46,X,del(Y); 45,X/46,X,der(Y)). These individuals are at increased risk of developing gonadoblastoma. In cases where a second X chromosome is detected, the presence of Y chromosome material is essentially ruled out, since the most likely explanation for the MX cell line is postzygotic loss of an X chromosome from disomy to monosomy. This informs potential

risk for gonadoblastoma and is important for clinical management. The most common variant karyotype observed is 45,X/46,XX mosaicism, which accounts for ~13% of TS. Postzygotic nondisjunction of 46,XX would lead to 45,X/46,XX/47,XXX mosaicism, which could be consistent if a 47,XXX line was lost during development. An alternative mechanism that could also lead to 45,X/46,XX mosaicism is monosomy X rescue, resulting in uniparental isodisomy (isoUPD) for the X chromosome, though this has rarely been reported. We reviewed patterns of prenatal (n=29; amniocytes=18, villi=8, hygroma fluid=3) and postnatal (n=18; peripheral blood=17, cord blood=1) MX cases identified by SNP microarray testing, performed over a six-year period at a clinical reference laboratory, to examine the frequency and characteristics of cases suggestive of MX rescue. We identified 2 probands (1 prenatal and 1 postnatal; 4% of MX cases), with MX and a 46,XX cell line with complete isodisomy X, visible as homozygosity across the entire X chromosome. The two cell lines were confirmed by karyotyping (prenatal - 45,X(18%)/46,XX(82%)) or FISH (postnatal - X(94%)/XX(6%)) analysis. The presence of isoUPD X does not preclude the presence of cryptic Y chromosome material, so gonadoblastoma risk should still be considered. Understanding the mechanism of monosomy X rescue could also explain high risk monosomy X cell-free DNA screens that resolve to 46,XX by the time of diagnostic testing. Patients with features of TS and a normal 46,XX karyotype may benefit from SNP microarray testing to uncover an isoUPD X, consistent with monosomy X rescue early in development or in another tissue.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3055W: Identification of Highly Pathogenic Avian Influenza Clade 2.3.4.4b

Authors:

E. Wendlandt[1], D. Tsang[1], E. Tretter[2], A. Britten[2], G. Frandsen[2], C. Vakulskas[1], A. Gissot-Boens[1]; [1]Integrated DNA Technologies, IOWA CITY, IA, [2]Udder Hlth.Systems, Meridian, ID

Abstract:

Highly pathogenic avian influenza viruses (HPAI) activity has intensified globally since 2021, with the first reported incidence in North America late 2021 (Eastern Canada) and early 2022 (United States). HPAI is an extremely contagious disease within poultry caused by Influenza H5 and H7 subtypes affecting multi-organ systems, resulting in high mortality rates within bird populations. While primarily confined to avian populations, in early 2024 reports began to surface of a novel, flu-like infection, circulating within bovine populations. Late March 2024, the United States Department of Agriculture confirmed the presence of HPAI in dairy cattle from farms in Kansas

and Texas. Since the initial identification, the outbreak has spread to at least nine states with the causative virus being a novel strain of influenza A H5N1, clade 2.3.4.4b. Identification of the virus within infected cattle can be difficult, as the milk presents with an abnormal appearance, similar to colostrum, which can inhibit traditional extraction methods. Here we present a novel quantitative real-time PCR (RT-qPCR) assay capable of detecting the clade 2.3.4.4b Influenza H5N1 virus. To bypass the need for problematic extraction workflows, we present a novel, extraction-free, workflow for the identification of H5N1 in infected milk. The presented methods were able to detect H5N1 2.3.4.4b in infected cattle as well as milk in large holding tanks. In addition, the use of synthetic templates demonstrated excellent sensitivity, with a limit of detection below ten copies per reaction. With the increased incidence of novel influenza H5N1 variants circulating globally, here we demonstrate a novel method for identification of H5N1 in infected dairy cattle.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3056W: Impact of Genome boards on the rate of molecular diagnosis for patients with neurodevelopmental and neurological disorders: An update and further developments.

Authors:

E. Hammar, M. Abramowicz; Univ. Hosp. of Geneva (HUG), Geneva, Switzerland

Abstract:

In 2012, clinical application of whole exome sequencing (WES) followed by targeted bioinformatics analysis of specific gene panels for the diagnosis of Mendelian disorders was introduced in our hospital. Whole genome sequencing and RNA analyses have since been developed to resolve unsolved cases. To further develop the benefits of these technologies for our patients and increase the synergy between laboratory, medical genetics and other medical specialties, Genome Boards for twelve groups of medical specialties have been created since 2019, including neurodevelopmental, neurology, gastrointestinal and renal disorders and ENT patients. During these weekly meetings, clinical cases and molecular results are presented, the class of variant pathogenicity is debated in view of anatomical, pathological and radiological aspects, and the laboratory reports are finalized. Financial access to the analyses and ethical issues such as informed consent, disclosure of incidental findings and/or of VUS are also addressed.
The aim of this work is to analyze the clinical utility of Genome Boards for neurological and neurodevelopmental disorders with or without epilepsy, through the measure of their impact

on the number of patients with molecular diagnosis.
Between 2021 and 2023, a total of 349 cases with developmental delay disorders (DD) and 245 cases with neurological diseases (Neuro) were analyzed. Overall, we found pathogenic variants (class 4 or 5) in 107/349 (31 %) of DD patients and in 65/245 (27 %) of Neuro patients. Among these patients, 82 DD and 57 Neuro patients were presented at multidisciplinary Genome boards, resulting in molecular diagnoses with class 4 or class 5 variants for 43/82 (52%) DD patients and 29/57 (51%) of Neuro patients. Several urgent cases could be resolved thanks to the multidisciplinary interactions, leading to an optimization of clinical management. In conclusion, these results show that Genome boards, through interactions between medical and laboratory specialists, lead to an enhanced ratio of molecular diagnoses for patients with neurodevelopmental and neurological disorders. Further developments are aimed at fostering collaborations for the development of precision treatments including the use of antisense oligonucleotides.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3057W: Increasing epilepsy-related diagnostic yield from short-read genome sequencing

Authors:

E. Higginbotham[1], J. Nguyen[1], K. Bell[2], N. J. Chandler[3], L. S. Chitty[3], J. Coleman[2], M. Coleman[2], J. Cross[3], A. Griffiths[2], B. Hanson[3], M. Lachgar-Ruiz[3], P. Lombard[3], C. Magro[3], C. Marshall, R.[1], B. Paternoster[3], A. Pritchard[3], I. E. Scheffer[2], W. Shao[4], J. Sidhu[1], S. Stephenson[2], B. Trost[1], Gene-STEPS Study Group, V. Chau[1], K. B. Howell[2], A. McTague[3], A. Poduri[4], Z. Stark[2], G. Costain[1], A. M. D'Gama[4]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Murdoch Children's Res. Inst., Melbourne, Australia, [3]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom, [4]Boston Children's Hosp., Boston, MA

Abstract:

Background: Neonatal and infantile-onset epilepsies often have genetic etiologies. Early genetic diagnosis is crucial for improving outcomes by facilitating personalized treatments and timely interventions. Gene-STEPS is a multi-center study of the diagnostic yield and clinical utility of trio rapid genome sequencing (GS) in infants with unexplained seizures. We have reported a clinical diagnostic yield of 43% for the first 100 infants. However, clinical laboratories' bioinformatics pipelines have limited ability to detect some variant types, including non-coding, mosaic, structural, and mitochondrial variants. Here, we describe initial results of advanced analysis strategies that further increase diagnostic yield.

Methods: For Gene-STEPS participants without genetic diagnoses after clinical GS (n=170), we conducted the following analyses: 1) re-evaluation of variants of uncertain significance, 2) manual review of GS data for a second variant in cases with a single pathogenic variant in an autosomal recessive gene, 3) reanalysis of all coding SNVs and indels, including in a consensus epilepsy gene list (n=1,045), 4) assessment of additional variant types, such as mosaic, structural, and mitochondrial variants, and 5) targeted RNA-sequencing.
Results: We established (likely) diagnoses for 15 infants, including candidate variants in both known disease-causing genes and genes not previously associated with disease like *ATP6V0B* and *HECTD1*. Further investigations are ongoing to establish associations between these candidate genes and epilepsy. The remaining diagnoses included a complex structural variant (2.3 Mb inversion disrupting *KMT2A*) and a mosaic SNV (*PIK3R2* present in 8.5% of reads).
Conclusion: Reanalysis of short-read GS increases the diagnostic yield for infants with unexplained seizures, with considerable anticipated impact on clinical management and reproductive counseling. Continuing improvements in bioinformatics analysis strategies, updates to variant interpretation guidelines, and discovery of new epilepsy-associated genes emphasize the importance of ongoing reanalysis of GS data.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3058W: Integrating emerging data into genomic testing: Outcomes from an evidence-based reanalysis initiative ⭐

Authors:

K. Radtke, M. Towne, C. Horton, S. Saliganan, K. Webb, B. Gasser, K. Hagman, B. Wayburn, J. Huang; Ambry Genetics, Aliso Viejo, CA

Abstract:

As evidence on gene-disease relationships (GDR) and variant pathogenicity grows, incorporating new data into variant interpretation requires a comprehensive approach. Sources of evidence can vary and include reports in the literature which may define new GDRs or expand phenotypic spectrums, public databases with more diverse representation, and variant-level functional data, such as RNA analysis.
We reviewed exome sequencing (ES) reported at a clinical lab between 2011- 2021 and subsequent reclassifications through 2023. All cases were involved in the Patient for Life Program, an evidence-driven reanalysis initiative that incorporates emerging evidence into clinical testing. A team of scientists proactively monitor scientific literature and other data

sources. As new information is identified, historical ES cases are reviewed, and reclassifications are issued when appropriate. We identified the outcome (negative, uncertain, positive [pathogenic/likely pathogenic]), evidence category resulting in reclassification (gene, variant, or clinical overlap) and directionality (upgrade v. downgrade). Within each evidence category we grouped evidence source. There was a 19% relative increase in diagnostic yield (21% vs. 25%). Overall, 9% (963/10,921) of cases received a reclassification report; some underwent multiple reclassifications, totaling 993. New evidence related to genes was the most impactful, accounting for 64% (637/993), followed by variant (29%; 285/993) and clinical overlap (7%; 71/993). The most vital source of evidence (64% of reclassifications) was literature describing new patients, often establishing new GDRs. Other common sources of evidence were new proband phenotypes (7%), updated MAF data from population databases (6%), family cosegregation studies (6%), and improvements to lab classifications and reporting procedures (5%). Most reclassifications (82%; 811/993) were triggered from evidence from external sources, such as publications and updated proband phenotypes, while 18% (182/993) were the result of a laboratory follow-up studies including family cosegregation studies, RNA analysis, and additional structural assessment. Collectively, 45% (449/993) had clinically significant upgrades, moving from uncertain or negative to positive, underscoring the impact of new evidence. Implementing evidence-driven reanalysis to proactively monitor, create, and incorporate emerging data improves ES accuracy, clinical utility, and diagnostic yield. Clinical laboratories should invest resources in proactive reclassification based on new evidence to maximize the diagnostic yield of genomic-based testing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3059W: International nomenclature for genomics

Authors:

R. Hastings[1,2], C. Morton[3], N. Chia[4], J-M. Dupont[5], L. Conlin[6], J. McGowan-Jordan[7], M. Kim[8], J-Y. Han[9], J. den Dunnen[10], S. Moore[11]; [1]GenQA, Edinburgh, United Kingdom, [2]GenQA, Oxford, United Kingdom, [3]Harvard Med. Sch., Boston, MA, [4]Queensland Univ. of Technology, Brisbane, Australia, [5]APHP.Ctr.-Université Paris Cité, Paris, France, [6]Children's Hosp. of Philadelphia, Philadelphia, PA, [7]CHEO, Ottawa, ON, Canada, [8]Seoul St. Mary s Hosp./The Catholic Univ Korea, Seoul, Korea, Republic of, [9]Dong-A-Univ Col Med, Busan, Korea, Republic of, [10]Leiden Univ. Med. Ctr.,, Leiden, Netherlands, [11]SA Genomics Hlth.Alliance, Adelaide,, Australia

Abstract:

With the advent of new technologies and deepening understanding of the nature of genomic aberrations and/or pathogenic variants, there is an increasing need for an international nomenclature to report the findings. The International System for Human Cytogenomic Nomenclature (ISCN) provides a standard approach to describe any genomic rearrangement identified by techniques ranging from karyotyping to FISH, microarray, genomic mapping, various region-specific assays, and DNA sequencing. The nomenclature also incorporates, as applicable, whether the abnormality is inherited or *de novo*, the proportion of the sample with the aberration and the variant allele frequency (VAF). Additional nomenclature for chimeras, moles and chromoanasynthesis and chromothripsis is included in ISCN 2024 and the region-specific assay nomenclature is expanded to include nomenclature for targeted chromosome analysis, targeted microarray analysis, fusion genes, expansion repeats and methylation assays for imprinting. Genome mapping technology is rapidly being adopted as a powerful cytogenomic tool which can detect aneuploidy, balanced and unbalanced structural variation as well as copy number changes. The latest release of ISCN 2024 will include new nomenclature for genome mapping which includes elements of karyotyping, microarray and region-specific nomenclature to describe the observed structural and copy-number variation. A brief overview of the changes to ISCN 2024 will be presented with constitutional and neoplasia examples illustrating the general principles of the nomenclature along with specific genome mapping examples for the different types of detectable numerical and structural rearrangements.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3060W: Long-read genomic sequencing reveals expanded GAA-GGA chimeric alleles in Friedreich ataxia

Authors:

S. Bidichandani[1], M. Tackett[1], C. Lam[1], D. Lynch[2]; [1]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Friedreich ataxia (FRDA) is the most prevalent inherited ataxia. It is typically caused by inheriting an expanded GAA triplet-repeat in intron 1 of the *FXN* gene from both parents. Expanded alleles range in size from 100 to 1500 triplets, and most alleles have >500 triplets. 15-20% of FRDA patients are genetically diagnosed as having expanded alleles of the same size (e.g., 800 and 800 triplets), because only a single expanded allele is visualized. We hypothesized that such patients may be hemizygous, i.e., for some reason only one *FXN* allele is being detected. In a

prospective cohort of 105 FRDA patients, we identified 21 (20%) unrelated individuals with a single expanded allele via two different long-range PCR amplicons. Long-read genomic sequencing (Oxford Nanopore, PromethION) in 12 of these patients revealed proximal *FXN* gene deletions in two. However, in the remaining 10 patients we discovered an expanded GAA-GGA chimeric repeat ranging from 800-1100 triplets. The conventional expanded GAA repeat was also detected in all patients, which matched the single expanded allele detected by long-range PCR. All 10 expanded GAA-GGA chimeric repeats were resistant to PCR amplification by the standard long-range PCR test. The current assumption that visualizing a single expanded allele is tantamount to a patient being homozygous for the same sized allele is frequently incorrect. Approximately 16% of FRDA patients have a new class of pathogenic allele that has been (and continues to be) missed by standard genetic diagnostic testing. Such individuals will have relatives whose heterozygous carrier status has been mis-assigned. Proximal *FXN* deletions are also likely under-diagnosed in FRDA.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3062W: Molecular diagnosis of patients with optic atrophy<:> a multimodal approach increased the diagnostic rate.

Authors:

E. Ullah, A. Naik, C. Bender, R. Al Rawi, R. B. Hufnagel, D. Blain, A. Agather, L. A. Huryn, W. M. Zein, B. P. Brooks, B. Guan; Natl. Eye Inst., NIH, Bethesda, MD

Abstract:

Optic atrophy (OA) is clinically characterized by a pale optic disc that indicates damage to the anterior visual pathway. Numerous conditions may cause optic atrophy including increased intracranial pressure, inherited conditions affecting the neuroretina, inflammatory diseases, vascular conditions, space-occupying intracranial lesions, trauma, infections, and toxicities. In the United States, in one study OA has been reported to cause blindness in 0.8% of the population, with a higher prevalence of the condition found in African Americans (1.9%). The genetics of hereditary conditions that are associated with OA are not well understood as only about 25% of patients receive a molecular diagnosis. Pathogenic variants are commonly reported in the genes *OPA1, WFS1, MT-ND6, MT-ND4, MT-ND1, UCHL1, AFG3L2, ACO2, RTN4IP1, NR2F1,* and *OPA3*. Notably, 20% of positive results are accounted for by mitochondrial genes (mtDNA). This study shares the experience of the Ophthalmic Genomics Lab with the genetic diagnosis of patients presenting with optic atrophy at the National Eye Institute (NEI) Ophthalmic Genetics clinic. Patients were clinically evaluated and were enrolled in an exome-

genome protocol (NCT02077894). Clinical genetic testing was first performed on (29/30) cases using a virtual panel of 19 nuclear genes on an exome backbone. The exome data was re-analyzed after processing through an in-house pipeline. Negative samples (11/30) underwent research genome sequencing using the same pipeline. Select cases (03/30) also underwent targeted mitochondrial genome sequencing. This cohort included 30 cases that underwent exome-based panel or targeted mtDNA testing. Genetic testing identified pathogenic variants in *OPA1, MT-ND6, WFS1, MFN2, COL18A1,* and *FDXR*, providing a molecular diagnosis to six patients (20%). Exome or genome analysis further identified disease-causing pathogenic variants in *ACO2*, *OFD1*, and *NBAS* in four additional patients (13%). This study also identified several candidate genes, among which a few are particularly intriguing and require further investigation. This systematic analytical study increased the diagnostic rate for optic atrophy by utilizing a panel of nuclear genes, mtDNA sequencing, exome re-analysis, and genome sequencing. This demonstrates the effectiveness of a holistic approach to genetic testing for OA. The comparatively low diagnostic yield suggests additional OA genes have yet to be identified. Identification of the candidate genes will hopefully expand our understanding of the pathogenesis of OA, improve testing yield, and, in the long term, inform treatment options.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3063W: Molecular testing for patients with congenital limb anomalies: yield analysis and new candidate genes

Authors:

A. Mokhtari[1], J. Charbonneau[2], V. Miranda[3], K. Jizi[3], M-A. Delrue[3], P. Egerszegi[3], I. Thiffault[4], P. Campeau[5]; [1]Montreal Univ., Montreal, QC, Canada, [2]St Justine Hosp., Montreal, QC, Canada, [3]St Justine Hosp., Montreal, QC, Canada, [4]Children s mercy, Kansas City, MO, [5]CHU Sainte Justine, Montreal, QC, Canada

Abstract:

Introduction Congenital limb anomalies are a common and heterogenous group of disorders. In the literature, up to 65% remain without an etiological diagnosis. To help close this gap, in this study, we describe the diagnostic outcomes of a historical cohort of patients presenting with limb anomalies skeletal genetics clinic of the St-Justine hospital major Canadian pediatric center. Methods All patients presenting with a predominant complaint of limb anomaly between 2014 and 2024 were included. Severe polymalformative cases and patients for which the main referral reason was not a limb anomaly were excluded. Variant interpretation was done according to ACMG criteria. A solved case was defined as a case in which a pathogenic or likely

pathogenic variant was identified in a phenotypically corresponding gene. Demographic, investigation and diagnostic information were extracted and described. The solved and unsolved sub-groups' characteristics were compared using standard statistics (t-test, chi-square). Results 131 patients were included in the final cohort, with a diagnostic yield of 35%. The most common conditions were polydactyly (24%), syndactyly (19%) and radial anomalies (19%). 50% of the cohort had syndromic features. 23% of solved cases (12% in unsolved) had a family history of limb anomalies. 7 (5%) patients underwent CMA only, 81 (63%) at least one panel and 43(32%) also underwent exome sequencing. Among a subset of patients with negative CMA and panel, exome sequencing yielded a diagnosis in 11 (25%). We identified 24 novel mutations in known disease-causing genes including *TBX3* (3 cases), *PRKAR1A* (2 cases) and *HOXD13* (2 cases). Further, we identified a frame-shift mutation in *HNRNPH2* in a male. This contributed to a better understanding of the condition as, until recently, it was thought that the disease was lethal in males and only associated with missense variants. The phenotype associated with other genes was expanded including *BMP4* and *PHF6*. We also identified two new candidate genes/regions for limb anomalies: *HOXA11* and a small 2q31.1 deletion involving notably *HOXD10*. Interpretation of the variants in these genes was facilitated by previously conducted studies in mice where mouse phenotypes correlated with our patients' phenotypes. Conclusion Limb anomalies remain a diagnostic challenge. According to our study, Exome sequencing likely represents an opportunity to identify novel etiologies in this heterogenous group of disorders, including the two candidate loci we identified: *HOXA11* and *HOXD10*. Model organism studies represent an important avenue to gain a greater understanding of the genetics of limb development and new disease genes.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3064W: mRNA-based analysis of non-coding variants increases the diagnostic yield of clinical whole genome/exome sequencing for rare diseases

Authors:

J-L. Blouin[1,2], A. VANNIER[1], C. Marconi[1], F. Masclaux[1], E. Hammar[1], M. Guipponi[1,2], R. van Heurck[1], T. Nouspikel[1], S. Suchet[1], G. Van Winckel[1], S. Fokstuen[1,2], M. Abramowicz[1,2], T. Rio Frio[1]; [1]Univ. Hosp. of Geneva, Geneva, Switzerland, [2]Univ. of Geneva-Sch. of Med., Geneva, Switzerland

Abstract:

Background: Whole genome/exome sequencing (WGS/WES) in clinical practice have significantly improved genetic diagnosis. With the sequencing of partial or complete intronic regions, numerous identified rare non-coding variants that could be of clinical significance, i.e. with a disruptive impact on mRNA splicing, are overlooked. Based on current guidelines for variant interpretation (ACMG), these rare non-coding variants are usually of uncertain significance (VUS) as *in silico* tools cannot accurately predict a potential effect on mRNA splicing. Implementation of a cost-, labor- and time-effective functional assay in a clinical diagnostic setting is required to decipher the pathogenicity of these variants and to provide a genetic diagnosis. **Methods**: Rare intronic variants identified by WGS/WES in patients were selected for targeted molecular investigations on RNA extracted from patients' blood. Variant selection relied on their location in strong candidate genes, their phase with an identified pathogenic variant for recessive diseases and/or in silico prediction of splice site creation or impact on canonical or weaker splice sites. **Results**: Over a 2-year period 22 rare non-coding VUS identified in patients' WES/WGS data were selected for mRNA analysis. Nineteen of these variants were localized in the vicinity of splicing sites (-20 to +20) and 3 were deep intronic. Analysis revealed diverse and unpredictable effects of variants on mRNA splicing. Notably, a deep intronic variant in *MFN2* gene without in silico supporting predictions was shown to disrupt a binding site for an alternative splicing factor repressor, causing the inclusion of a normally repressed out-of-frame cryptic exon. A deep intronic variant in *MYBPC3* created a strong donor splice site that caused the inclusion of a partial intronic retention. A homozygous variant in the acceptor site of the last intron of *ELMO2* caused the use of several weak alternative intronic and exonic acceptor sites and ultimately to the production of 5 transcripts with aberrant 3' ends. Overall, 15 of the 22 (68%) selected VUS were reclassified as likely pathogenic/pathogenic based on results of targeted mRNA analysis, according to their impact on splicing and in accordance with the pathogenic mechanism of mutations of the gene they lie in. **Conclusion**: Our results emphasize the importance of characterizing at the mRNA level rare non-coding VUS identified by WGS/WES and highlight the limitation of in-silico predictions. This analytical pipeline allowed the reclassification of 68% of non-coding VUS as likely pathogenic/pathogenic and significantly improved the diagnostic yield of WGS/WES.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3065W: On the (ir)reproducibility of EpiSign, a methylation-based clinical assay

Authors:

L. Ferreira, R. Hopkin, L. Martin; Cincinnati Children's Hosp., Cincinnati, OH

Abstract:

Establishing the diagnosis of a genetic condition is oftentimes challenging. The development of assays that interrogate the methylation status of 100,000's of genomic *loci* has allowed researchers to explore the influence of epigenetics in human disease. Recently it has been suggested that methylation signatures can be used to assist diagnosis. EpiSign is a leading methylation assay developed for such purposes. These signatures have not been independently reproduced or verified, and the data has not been publicly shared. A formal attempt to replicate them independently is long overdue. This study fills this gap by reproding the results of the largest EpiSign publication. We secured 496 methylation profiles: 365 controls (including syndromes with fewer than 20 samples each) and 131 cases for syndromes with more than 20 samples (CHARGE, Kabuki, Sotos and Williams). We could not perform a thorough and complete replication since most of the data used was not publicly shared. By employing a 75%-25% sample split for training and testing and repeating the analysis 1,000 times, we obtained the following results (in decimals.) Mean training accuracy of [1, 1, 0.99, 1] for CHARGE, Kabuki, Sotos and Williams, respectively, in line with reported results. Metrics obtained with the test set include 95% confidence intervals for accuracy, sensitivity and specificity (the latter two metrics have not been reported.) CHARGE: mean accuracy 0.96 [0.92,0.98], mean sensitivity 0.81 [0.69,0.93], mean specificity 0.96 [0.93, 0.98]; Kabuki: mean accuracy 0.93 [0.89,0.95], mean sensitivity 0.70 [0.58,0.82], mean specificity 0.93 [0.89, 0.96]; Sotos: mean accuracy 0.94 [0.92,0.97], mean sensitivity 0.79 [0.66,0.92], mean specificity 0.95 [0.92, 0.97]; Williams: mean accuracy 0.995 [0.99,0.998], mean sensitivity 0.99 [0.95,0.995], mean specificity 0.995 [0.989, 0.999]. Though some of these results mirror the ones in the publication, our efforts identified concerns about the robustness of these estimates. Our data shows there is a strong batch effect across all batches and it was not possible to perform any correction as some of the cases are perfectly correlated with the batches. There is also a serious concern of data leakage between feature selection and model training as the same samples were used for both. Also, accuracy estimates are likely inflated given the testing set was highly imbalanced. Finally, the EpiSign methodology has changed several times and it remains unclear how a prediction model could have been estimated for conditions with fewer than 20 samples. All in all, further analysis is warranted to validate EpiSign signatures.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3066W: Project FIND-OUT: Protocol and pilot study for whole genome sequencing in symptomatic infants

Authors:

G. Pierobon Mays[1], A. Terala[2], A. Baxter[3], K. Schmidt[2], E. Rountree[4], B. Liebers[3], S. Berger[5], T. Bichell[2]; [1]COMBINEDBrain, Verona, WI, [2]COMBINEDBrain, Brentwood, TN, [3]Guide Genetics, Inc., Schenectady, NY, [4]COMBINEDBrain, Avon, CO, [5]Childrens Natl. Res. Inst., Washington, DC

Abstract:

There are many challenges to diagnosing rare genetic neurodevelopmental disorders (RGNDs) in children. Of particular concern is the latency between identifying symptoms and finding a genetic diagnosis, averaging 30 months; this results in the loss of critical time for intervention. Factors contributing to that delay include lack of guidelines for genetic testing in infants outside the NICU, long wait times and limited access to genetic specialists, and restrictive insurance policies. This study aims to determine the utility of whole genome sequencing (WGS) for infants at-risk for RGNDs outside of the NICU setting.

FIND-OUT participants were 3-12 months old, had not had previous WGS, and had symptoms in at least 2 of 7 categories: feeding, movement, NICU admission, developmental delay, other (structural anomalies, growth concerns, or specialist referral), unprovoked seizures, or tone. Infants were evaluated via caregiver surveys and assessments by trained project staff. FIND-OUT participants had proband-only WGS at a CLIA/CAP certified laboratory, with results disclosed by a genetic counselor.

From December 2023-May 2024, 22 enrolled, averaging 4.2 criteria; 56% had a pediatric neurology referral. Reasons families reported for joining FIND-OUT include insurance barriers to obtaining genetic testing, specialist (including geneticists) unwillingness to order testing, wait time to see genetics, and a desire to optimize care for their child.

Thus far, 19 samples have been submitted, with 11 returned results; 18% had diagnostic results, 36% had possibly diagnostic results, and 45% had non-diagnostic results. Two participants had secondary findings and 2 had incidental findings; 3 of these results had childhood health implications. The age at date of WGS results ranged from 4-15 months. Three participants had prior genetic testing and four had genetics evaluations.

The diagnostic rate for this study is similar to those in studies of WGS for other cohorts, supporting the utility of WGS for infants at-risk for RGNDs. Time to diagnosis was significantly shorter than the average of 30 months, allowing for interventions to be started much earlier. Receiving WGS through FIND-OUT removed barriers to testing, highlighting the need for changes in WGS ordering practices and insurance coverage. Understandably, genetics professionals have concerns about non-genetics providers ordering WGS; however, if genetics evaluation expands the phenotype, WGS reanalysis is available. Overall, these early data support the value of WGS in infants at-risk for RGNDs, even before a genetics consult, to shorten the time to diagnosis and increase the opportunity for intervention.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3067W: Resolving complex genotypes in residual clinical samples with long range-PCR and nanopore sequencing assay

Authors:

C. Edwards[1], B. Killinger[1], A. Laurie[2], L. Hsu[2], A-S. Lebre[3], T. Yammine[4], C. Ho[5], I. Volkova[5], M. Roberts[1], T. Markulin[1], M. Mihailovic[1], E. Priego[1], B. C. Haynes[1], B. Hall[1]; [1]Asuragen, Austin, TX, [2]Canterbury Hlth.Lab., Christchurch, New Zealand, [3]CHU Reims et Université Reims Champagne-Ardenne (URCA); Université Paris Cité, Inst. of Psychiatry and NeuroSci. of Paris (IPNP), INSERM, Reims et Paris, France, [4]CHU Reims, Reims, France, [5]Stanford Hlth.Care, Palo Alto, CA

Abstract:

Introduction:
Traditional methods struggle with detecting complex variants in key carrier screening genes, necessitating complicated workflows and producing results that often lack comprehensive detection of pathogenic variants. Six of the top ten highest risk genetic disorders have complex variants, and resolving the risk of heritability or disease across these genetically complex regions often requires the use of multiple workflows and/or ancillary methods.
Here we discuss results from a single workflow using long-range PCR amplification, long-read nanopore sequencing, and automated analysis software to detect multiple classes of genetic variation within a multigene panel (*SMN1/2*, *F8* intron inversions, *CYP21A2*, *GBA*), to identify structural rearrangements (such as fusions and inversions), single nucleotide variants (SNVs), insertion/deletions (INDELs), and copy number variants (CNVs). Paralogs are distinguished and variants are phased to enable comprehensive genotype analysis and differentiation of gene from pseudogene sequences.
Methods:
The prototype assay workflow was tested across four laboratories with >200 unique samples including clinical residuals and cell line material. External laboratories were trained to perform the prototype assay workflow and run automated analysis software. Target regions were enriched in PCR reactions, barcoded, pooled, and sequenced on MinION™ flow cells (R10.4.1) with a Mk1B (Oxford Nanopore Technologies). Results were compared with orthogonal methods including MLPA, Sanger sequencing, and long-range PCR.
Results:
Across all variant types, the assay was ≥95% concordant with orthogonal results. SNVs/indel concordance was >98%. *SMN1/2* CNV (0 to ≥3) showed ≥95%
agreement. *GBA* and *CYP21A2/TNXB* were differentiated from their pseudogenes, gene-

pseudogene fusions were identified, and SVs were resolved. *F8* intron inversions showed ≥95% agreement. We identified *F8* Intron 1 and 22 inversions, *CYP21A2* del/dups, SVs, and variant phasing, *GBA/GBAP* chimeras, SNVs, CNVs, as well as *SMN1/2* hybrids. The assay was tested against a diverse set of samples, and all expected classes of variation were detected.
Conclusion:
Long-Range PCR coupled with nanopore sequencing elegantly resolves difficult targets that are otherwise intractable by short-read sequencing and require ancillary reagents and equipment to accurately resolve. Our results highlight the utility of nanopore sequencing and dedicated bioinformatics software to consolidate workflows and automate both the analysis and interpretation of challenging variants relevant to hereditary genetic disease.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3068W: Revisiting the genetic basis of DFNB9 through cryptic variant exploration and haplotypic phasing

Authors:

P-H. Lin[1,2], C-Y. Tsai[2], Y-T. CHIANG[1], C-H. Ho[2], Y-H. Lin[2], J. Hsu[3], T-C. Liu[2], C-J. Hsu[4], P-L. Chen[3], C-C. Wu[2]; [1]Natl. Taiwan Univ., Coll. of Med., Taipei City, Taiwan, [2]Natl. Taiwan Univ. Hosp., Taipei City, Taiwan, [3]Natl. Taiwan Univ., Taipei City, Taiwan, [4]Dept. of Otolaryngology, Taichung Tzu Chi Hosp., Taichung, Taiwan

Abstract:

Background: The autosomal recessive deafness DFNB9, caused by pathogenic variants of the *OTOF* gene, is characterized by non-syndromic auditory neuropathy spectrum disorder (ANSD). Recent advances in animal studies and human clinical trials have highlighted the potential for gene therapy to rescue DFNB9. However, unresolved genetic diagnosis due to limitations in current genetic testing hampers available treatment options for affected individuals. This study aims to identify potentially pathogenic *OTOF* variants using both short-read sequencing (SRS) and long-read sequencing (LRS) methods, complemented by a prediction-to-validation pipeline using SpliceAI and minigene assays. **Methods:**Sixty-five unrelated non-syndromic ANSD patients were enrolled for genetic analysis. Targeted SRS was used as a first-tier diagnostic tool. In addition to coding DNA sequences, SpilceAI was used to detect potential aberrant splicing variants, and minigene assays were used to validate differential expression of spliced transcripts resulting from variants predicted to cause mis-splicing. Furthermore, haplotype phasing was performed using Oxford Nanopore Technologies platform for LRS. **Results:**Among the subjects, 33 were confirmed to harbor biallelic

pathogenic *OTOF* variants, while five were found to harbor monoallelic pathogenic *OTOF* variants. Three novel variants (c.3864G>A, c.5813+2T>C and c.4501G>A) were identified in this study. In addition, two cryptic variants (c.3864G>A and c.3894+5G>C) with high predicted score by SpliceAI were observed to induce alternative splicing as confirmed by minigene assays. Furthermore, a novel variant, p.K1992R, was found to be closely linked to the previously reported pathogenic variant p.E1700Q. Haplotype phasing with LRS revealed that the *cis* configuration between p.E1700Q and p.K1992R constitutes an independent pathogenic allele that, when in *trans* with the other pathogenic variants, can result in compound heterozygosity in *OTOF* and lead to DFNB9. **Conclusion:**The integration of SRS and LRS techniques, in conjunction with tools such as SpliceAI for predicting mis-splicing events and minigene assays, facilitates the elucidation of the genetic basis of DFNB9 in affected patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3069W: Specification of frequency criteria for secondary findings genes to improve classification concordance

Authors:

J. Johnston[1], K. Lee[2], D. I. Ritter[3,4], S. M. Harrison[5], L. G. Biesecker[1]; [1]NHGRI, NIH, Bethesda, MD, [2]Univ. of North Carolina, Chapel Hill, NC, [3]Baylor Coll. of Med., Houston, TX, [4]Texas Children's Hosp., Houston, TX, [5]Ambry Genetics, Aliso Viejo, CA

Abstract:

ACMG policy recommends that secondary findings (SF) from clinical sequencing be offered to individuals. Toward standardizing this process, the ACMG Secondary Findings Working Group curates a list of genes for variant return, v3.2 contains 81 genes. In 2015, the ACMG/AMP put forward criteria for classifying the pathogenicity of variants applicable to mendelian disease. Prior work has shown discordance in variant classification amongst laboratories using these classification criteria. To increase concordance, ClinGen is providing additional guidance through the Sequence Variant Interpretation group and Variant Curation Expert Panels (VCEPs). However, 57 genes suggested for SF return do not currently have published specified criteria. Having specified criteria for SF genes is important as laboratories are expected to classify variants in these genes for SF return whether or not they have specialized expertise in the gene. Classification concordance could be increased by providing basic specified criteria. Our goal is to provide thresholds for BA1 (stand-alone benign) and BS1 (allele frequency greater than expected for disorder) along with guidance for application of PVS1 (loss of function) prior to VCEP criteria publication. The Whiffin-Ware calculator was used to define BA1 (wBA1) with

disease frequency, gene contribution and penetrance values identified from the literature. It is reasonable, for genes without a VCEP, to define a more stringent BA1 (sBA1) at ten times the wBA1 value. We have compared calculated wBA1 and sBA1 values with popmax filtering allele frequencies (FAF) of the most common pathogenic/likely pathogenic (P/LP) variants in gnomAD v4.1.0. P/LP variants were defined as variants with a ClinVar classification of P/LP or novel loss of function variants in the MANE transcript for genes with evidence supporting haploinsufficiency. For 15 genes without published VCEP BA1 values, the popmax FAF of the most common P/LP variant was higher than the calculated wBA1, none were higher than the calculated sBA1. For BS1, we suggest using a value three times wBA1 (sBS1) when published VCEP criteria are not available. Four genes, *BTD*, *CACNA1S*, *RET*, and *TTR*, had a known P/LP variant with a popmax FAF higher than the calculated sBS1. One additional gene, *SDHAF2*, had a novel LoF variant with a FAF higher than the calculated sBS1. For PVS1, we have used the ClinGen dosage sensitivity data and variant information from HGMD and ClinVar to determine if loss of function variants contribute to disease. Specification of basic criteria should allow for improved classification concordance for the return of secondary findings.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3071W: The Face of *RNU4-2* Neurodevelopmental Disorder, a New, Common and Recognizable Syndrome

Authors:

J. Friedman[1], K. Barbour[1], K. Wigby[2], A. Besterman[3], L. Tobin[4], **J. Lenberg**[4], M. Del Campo[1], L. Bird[5], M. Bainbridge[5]; [1]Univ. of California San Diego, La Jolla, CA, [2]Univ. of California Davis, Davis, CA, [3]UCSD, San Diego, CA, [4]Rady Children's Inst. for Genomic Med., San Diego, CA, [5]Rady Children's Hosp., San Diego, CA

Abstract:

BACKGROUND: *RNU4-2* is a newly identified gene associated with a large proportion of individuals with neurodevelopmental disability. The gene encodes U4 small nuclear rna, a critical component of the spliceosome. Diagnosis of this condition is hampered by the inability of commonly employed and easily accessible clinical testing methods to detect variants in this non-coding gene. In addition, early studies report only aggregate phenotypic features without systematic, in depth prospective individual patient evaluations; this may lead to under-reporting of certain aspects, failure to represent the full phenotypic spectrum, and do not highlight aspects that may facilitate diagnosis. Identification of recognizable phenotypic features that may prompt further evaluation of individuals with neurodevelopmental disorders for this

condition are needed.We aimed to identify individuals with *RNU4-2* gene variants sequenced at our institution and to deeply phenotype these individuals to clarify the clinical profile of individuals with *RNU4-2* and identify key phenotypic features that may be a clue to diagnosis in order to highlight individuals for whom more specialized testing may be indicated.
METHODS: Whole Genome Sequencing (WGS) data from 6,734 individuals with any phenotype who underwent WGS at Rady Children's Institute of Genomic Medicine were queried for variants in the *RNU4-2* gene. Four children with *RNU4-2* variants followed clinically at Rady Children's Hospital were identified and examined in detail by a team including neurologists, clinical geneticists and dysmorphology experts. Phenotypic features were systematically collected and compared across our cohort and in relation to previously published aggregate data.
RESULTS: From our analysis, core and common aspects of the phenotype associated with *RNU4-2* gene variation emerged. In addition, we were able to define a recognizable facial pattern. **CONCLUSIONS:** Enhanced recognition of the clinical phenotype most notably the associated dysmorphic facial features associated with *RNU4-2* gene variation will inform patient selection for genome sequencing or targeted gene testing and will enable increased diagnosis of this common condition. Increased diagnostic rate will benefit patients by providing a diagnosis and will facilitate research evaluating, genotype/phenotype associations, natural history, incidence, underlying disease mechanism, and development of targeted therapies.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3072W: The impact of a machine learning framework in reducing uncertainty and increasing equity in clinical genetic testing

Authors:

F. Facio[1], E. Chen[1], B. Johnson[1], K. Ouyang[1], H. Metz[1], A. Colavin[1], R. Nussbaum[1,2], K. Nykamp[1], L. Fresard[1]; [1]Invitae, San Francisco, CA, [2]Univ. of California San Francisco, San Francisco, CA

Abstract:

Introduction: Variants of uncertain significance (VUS) remain a challenge in clinical genetic testing, especially in individuals from racial, ethnic and ancestral (REA) groups underrepresented in genomic research and clinical testing. This study describes the impact of a machine learning (ML) framework on variant classification (VC) and VUS reduction in a real world cohort of > 1 million individuals undergoing clinical genetic testing.
Methods: The development and validation of the ML framework, known as the Evidence Modeling Platform (EMP), is described by Fresard et al. in another abstract. DNA sequencing

results from de-identified unrelated individuals referred for genetic testing at Invitae® between Sep 2013 and Oct 2023 were included. Age, gender, and REA were provided by ordering clinicians. To assess the impact of EMP evidence on VC we identified all variants that used EMP for definitive classification (P, LP, LB, B) but would have been classified as VUS in the absence of EMP evidence. Clinical impact was assessed from all patients who had at least one variant with an EMP-dependent definitive classification, computed across 12 clinical areas (Hereditary Cancer, Preventive, Neurology, Pediatric Genetics, Cardiology, Metabolic Disorders and Newborn Screening, Immunology, Ophthalmology, Exome, Carrier, Hematology, Dermatology), and 11 REA categories (Ashkenazi Jewish, Asian, Black, Hispanic, Middle Eastern, Native American, Pacific Islander, Sephardic Jewish, White, Multiple/Mixed Ethnicity, Unknown/Other).
Results: A total of 1,052,843 individuals received at least one VC supported by EMP evidence, and of these 545,805 received a definitive classification: 95% (520,169) with one or more LB/B and 8% (41,876) with one or more LP/P classification. The proportion of individuals who benefited from at least one definitive EMP-dependent classification varied by clinical area, ranging from 48.3% in ophthalmology to 5.6% in Hematology. Across REA categories, one or more variants whose definitive classification depended on EMP were found in significantly more individuals of Asian (26%), Middle Eastern (23%), Sephardic Jewish (22%), and Black (21.6%) backgrounds compared to non-Hispanic white (12.7%) and Ashkenazi Jewish individuals (7.2%) (one-sided Fisher's exact test, $p < 10e-20$).
Conclusion: A systematic approach to integrating ML models into VC reduces VUS and leads to more definitive results. Importantly, EMP models led to more definitive classifications in individuals from underrepresented REA groups, bringing more equity to the VC process.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3073W: The parental origin of genomic imbalances found in patients with mental retardation and developmental delay

Authors:

S. Shim[1], H. Jang[1,2], J. Park[2], H. Kim[2], M. Go[1,2], S. Yang[2], J. Kim[2], S. Eom[2]; [1]Cha Univ., Seongnam, Korea, Republic of, [2]CHA Biotech Inc., Seoul, Korea, Republic of

Abstract:

Mental retardation and developmental delay (MR/DD) is a disorder that causes delayed motor development, speech delay, intellectual disability, and seizures. Chromosome aberrations are a well-known cause MR/DD and are found in 4-30% of MR/DD patients. Therefore, cytogenomic analyses such as chromosome microarray and multiplex ligation-dependent probe amplification

(MLPA), are primarily used to diagnose MR/DD patients. A total of 1,106 patients were examined and 42 patients (3.8%) from 40 families had genomic imbalances. As a result of parental testing, nine patients were of paternal origin, 16 were of maternal origin, and 23 were de novo variants. Six patients had two variants, one is inherited and the other is *de novo*. Genomic losses were found in 16 patients and genomic gains were identified in 15 patients. Twelve patients had both genomic gain or loss. In two families, genomic gains and losses resulted from submicroscopic balanced translocation of the mothers. Some variants were inherited from phenotypically normal parents and they might be normal copy number variants. And for some de novo variant, additional analysis is in progress to determine their pathogenicity. Genomic testing has become common, but genetic causes are still not found in most idiopathic MR/DD patients. In order to ensure the effectiveness of genome analyses for MR/DD patients, additional researches including building a more in-depth database are required.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3074W: The Pitfalls of Whole Exome Sequencing Analysis in Clinical : A Single Medical Center Experience

Authors:

C-H. Lee, Y-C. Liang, Y-N. Huang; Dept. of Pediatrics, Kaohsiung Chang Gung Mem. Hosp., Kaohsiung City, Taiwan

Abstract:

Whole exome sequencing (WES) is not only a massively parallel but also a high-throughput DNA sequencing for all coding regions in an individual's genome. As a result, WES has emerged as a powerful diagnostic tool to detect gene variants associated with disease in clinical. However, pitfalls can still occur during the annotation and interpretation of gene variants in WES data. In this discussion, we share the common but important perspectives that are often overlooked. The first case was a three-year-eight-month-old girl presenting with hypokalemia and metabolic alkalosis. Initially diagnosed with Bartter syndrome by the nephrologist and WES identified only the heterozygous pathogenic variant *SLC12A3* c.179C>T (p.Thr60Met). Subsequent analysis of the BAM file revealed a copy number variant, specifically a deletion at 16q13 (56879064-56880251), encompassing the *SLC12A3* gene. This finding confirmed the diagnosis of Gitelman syndrome, illustrating the importance of considering both single nucleotide variants and copy number variants in autosomal recessive disease. The second case was a three-year-eight-month-old girl with strong suspicion of neurofibromatosis. At first, WES did not reveal any

pathogenic variant in the *NF1* or *NF2* gene. Examination of the BAM file revealed absence of reads in the regions c.2326-5 to c.2409+5 and c.2991-5 to c.3113+5 of the *NF1* gene. Sanger sequencing of these regions identified the heterozygous pathogenic variant *NF1* c.2352 G>C (p.Trp784Cys). This is a technical limitation in WES due to ineffective capture and amplification. The third case was a ten-day-old female neonate who altered by the newborn screening center for Carnitine Palmitoyl transferase II deficiency. Prenatal WES analysis in amniotic fluid only identified a heterozygous pathogenic variant in the *SLC25A20* gene. Subsequent postnatal WES revealed compound heterozygosity in the *CPT2* gene: c.1054T>C (p.Phe352Leu) of uncertain significance and c.1799G>A (p.Gly600Glu) pathogenic. Notably, analysis of the BAM file indicated that the first variant was c.1054_1055delinsCG (p.Phe352Arg), which exhibited greater predicted damage to protein function and closer alignment with the phenotype. This underscores the importance of multi-nucleotide variant analysis in such scenarios. In conclusion, our experience demonstrates that the paramount importance of phenotype in diagnosis, even initial whole exome sequencing analysis fails to align with the observed phenotype. Additionally, BAM files represent valuable data resources that can assist to identify genetic variations beyond single nucleotide variant in WES.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3075W: Uncovering hearing impairment-related structural variants via integrating long-read and short-read whole-genome sequencing

Authors:

Y-T. Chiang[1], P-H. Lin[1], C-Y. Tsai[2], Y-S. Lu[3], T-L. Liao[4], H-Y. Weng[5], P-L. Chen[6], J. Hsu[7], S-F. Tsai[4], C-C. Wu[8]; [1]Natl. Taiwan Univ., Coll. of Med., Taipei City, Taiwan, [2]Natl. Taiwan Univ. Hosp., Taipei city, Taiwan, [3]Natl. Taiwan Univ. Hosp., Taipei City, Taiwan, [4]Inst. of Molecular and Genomic Med., Natl. Hlth.Res. Inst.s, Zhunan, Taiwan, [5]Natl. Yang Ming Chiao Tung Univ., Taipei City, Taiwan, [6]Natl. Taiwan Univ., Taipei, Taiwan, [7]Natl. Taiwan Univ., Taipei City, Taiwan, [8]Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Abstract:

Hereditary hearing impairment (HHI), the most common cause of permanent hearing loss in childhood, has been linked to a vast array of over 200 genes. Although advances in sequencing technology have greatly facilitated clinical diagnosis, more than 50% of hearing-impaired children remain undiagnosed, probably because the current sequencing strategy focuses primarily on variants smaller than 50 base pairs. Structural variants (SVs) larger than 50 bp are

often overlooked despite their potential importance. Long-read whole-genome sequencing (lrWGS) has demonstrated superior SV detection to short-read whole-genome sequencing (srWGS). To identify SVs in undiagnosed HHI patients, we collected Pacific Biosciences (PacBio) HiFi lrWGS data from 11 affected children from unresolved multiplex families and Illumina srWGS data from their family members. PacBio pbmm2 and pbsv were used for lrWGS SV detection, and Illumina DRAGEN was used for srWGS. To identify potential HHI-causing SVs from over 20,000 SVs in each patient, we applied specific SV filtering strategies. First, family-based SV joint calling and co-segregation analyses were performed on the srWGS data. Second, jointed SV sets of Human Pangenome Reference Consortium samples, generated by pbsv from PacBio (103 samples) and DRAGEN (43 samples), were used to exclude SVs that may result from technical bias. Third, SVs were excluded based on the following criteria: allele frequencies > 2% in the Taiwan Biobank or gnomAD, LOEUF scores > 0.6, CADD_SV scores < 10, and no overlap with pathogenic SVs in ClinVar as annotated by Variant Effect Predictor (VEP). Ninety percent of each SV segment was the required minimum overlap with SVs in the datasets. After the filtering, lrWGS identified an average of 5,761 SVs, and srWGS identified an average of 905 co-segregating SVs. On average, 64 SVs were detected by both srWGS and lrWGS for each affected child. Among these common SVs, four deletions located in non-coding regions and identified in four families were predicted to disrupt RNA splicing, making them potential SV candidates for further investigation. One notable deletion was a 25,617 bp deletion in *TAS2R14*, which encodes a G-coupled receptor. Although *TAS2R14* is not a well-known deafness gene, the deletion may represent a novel pathogenic SV, given the known association of G-coupled receptors with hearing loss. This study demonstrates an innovative SV analysis strategy that integrates lrWGS and srWGS, offering promising avenues for improving the clinical diagnosis of HHI and uncovering novel deafness-associated SVs, paving the way for future advances in the field.

Session Title: Molecular and Cytogenetic Diagnostics Poster Wednesday Session

Board 3076W: VariantMatcher: phenotypic and genomic data sharing to facilitate variant classification and disease gene discovery

Authors:

N. Sobreira[1], M. Szeremeta[2], D. Villela[2], R. Martin[1], S. Griffith[1], L. Vail[1], J. Posey[3], Z. Coban-Akdemir[4], K. Doheny[5], J. Lupski[3], D. Valle[1], A. Hamosh[1], C. Antonescu[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Diagnósticos da América S.A., DASA, Sao Paulo, Brazil, [3]Baylor Coll. of Med., Houston, TX, [4]UTHlth.Houston Sch. of Publ. Hlth., Houston, TX, [5]Ctr. for Inherited Disease Res. (CIDR), JHU, Baltimore, MD

Abstract:

VariantMatcher (https://variantmatcher.org) is a web-based tool that connects individuals interested in a specific variant worldwide. It enables sharing de-identified variant-level and phenotypic data from research participants to discover novel disease-causing variants and genes. The database contains rare (MAF < 1% in gnomAD) nonsynonymous SNVs identified in 6,827 VCF files (957,249 unique variants) of affected and unaffected individuals sequenced as part of multiple projects along with their detailed phenotypic information. Matches are based on genomic coordinates, although phenotypic features can be added and shared in the email notifying the match. VariantMatcher is also connected to theBeacon Network through the Beacon protocol. As of June 1st, 2024, VariantMatcher had 1,167 submitters from 60 countries; 19,913 variants had been queried, and 512 variants matched 4,662 individuals. To date, most matches have supported the more precise variant classification of variants initially classified as VUSs, enabling better, more precise patient care. To investigate the use of VariantMatcher and its benefit in variant classification, we analyzed 63 VUSs identified by Diagnósticos da América S.A., DASA, a laboratory in Brazil, between 2018 and 2021 and present in VariantMacher. Of these 63 VUSs, 23 were in genes that cause autosomal or X-linked dominant diseases. By analysis of the variant matches' phenotypes, four variants were reinterpreted as "benign" and 14 as "likely benign"; four remained classified as VUSs. Further analysis of the other variant matches is underway to determine the impact of these matches in the classifications of variants in genes related to autosomal and X-linked recessive diseases.To increase VariantMathcer's query capabilities, we now allow queries of specific groups of variants/genes. For every variant match, we also return metadata information that includes the number of samples with the variant in heterozygosity or homozygosity and the total number of individuals queried. Following the Matchmaker Exchange model for gene-level data sharing, we used the Beacon v2 API to connect VariantMatcher and Franklin by genoox. This allows VariantMatcher users to query Franklin by genoox and vice versa. VariantMatcher has been widely used globally and has fostered significant collaborative success stories. It enables researchers, healthcare providers, and patients to share their phenotypic and genomic data and build new connections. These collaborations have significantly increased the discovery rate of novel etiologies for diseases of unknown cause, inspiring and motivating potential users to join the community.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6002W: Genetic association of Glucose-6-phosphate dehydrogenase variants and Pre-eclampsia (PET) in Ghana ⭐

Authors:

N. Karikari Agyemang[1], C. Obirikorang[2], L. Ahenkorah Fondjo[2], M. Lawson Gakpey[1], A. Ghansah[1], D. Adu[3]; [1]Noguchi Mem. Inst. for Med. Res., Accra, Ghana, [2]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [3]Univ. of Ghana Med. Sch., Accra, Ghana

Abstract:

Pregnant African Americans carry more than a 2-fold higher risk of Preeclampsia toxemia (PET) as compared to non-Hispanic Whites (6.04% to 3.75%) and Hispanic Americans (6.04% to 2.58%)(P<0.0001 Recent studies suggest that Africans in SubSaharan Africa (African Blacks) also have a similarly high predilection to PET as do African Americans. In one study in Ghana, the incidence of PET was 7.9%. The high incidence of PET in African Americans and African blacks suggests that these two populations may share common genetic predispositions for such. One possible contributor to preeclampsia is variants in the G6PD gene, encoding for glucose-6-phosphate dehydrogenase (G6PD). G6PD variants evolved and rose high in some populations, including Africans, because they were protected against malaria. Some studies have suggested that G6PD deficiency may be associated with preeclampsia. We studied this in a case-control study of 154 Ghanaian women with preeclampsia and 169 pregnant healthy controls receiving antenatal care at the Korle-Bu Teaching Hospital. Isolated genomic DNA was uniplex-PCR amplified, and Restriction Fragment Length Polymorphism was assayed for single-base variants of both G6PD 202 and G6PD 376. The association between maternal G6PD (A and A-) variants and PET was determined using the chi-square test and logistic regression. 23.3% of maternal cases with preeclampsia were deficient for G6PD variant A-, as did 8.90% of controls, and for variant A, 36.4% of maternal cases with preeclampsia were deficient, while 21.9% of controls were deficient for variant A. We conclude that maternal G6PD variants A and A- showed an association with the development of preeclampsia [p =0.016, p = 0.005], respectively, with the A - increasing the risk of PET development by more than 4 fold(OR=4.20 [2.25-12.1] p = 0.002). These data suggest that G6PD deficiency may be associated with an increased risk of preeclampsia. This observation must be confirmed using a larger study set and a robust genotyping platform.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6003W: A data-driven map of human metabolism across the allele frequency spectrum ⭐

Authors:

M. Zoodsma[1], S. Yasmeen[1], C. Beuchel[1], L. Kohleick[1], A. Williamson[2], M. Pietzner[1], C. Langenberg[2]; [1]Berlin Inst. of Hlth.@ Charité, Berlin, Germany, [2]Queen Mary Univ. of London, London, United Kingdom

Abstract:

Rare and common genetic variation regulates the plasticity of human metabolism and its effect on human health and disease. Until recently, systematic, data-driven identification of metabolic regulators, including across-organs, tissues and cells has not been possible, and current knowledge is largely based on bespoke experimental work in well-defined model systems. Here, we provide a comprehensive, data-driven map of the genetic regulation of metabolites across the allele frequency spectrum covering a broad range of biochemical molecules with a specific focus on lipoproteins. We quantified 249 circulating metabolites by nuclear magnetic resonance spectroscopy (Nightingale Health platform) and assessed the full breadth of the allele frequency spectrum by performing genome-wide association studies of common variation (MAF>0.5%) in up to 460K UK Biobank participants across three ancestries, as well as whole exome sequencing rare variant burden testing in >400K individuals of European ancestry. We identified 394 novel associations between common regional sentinels and metabolic phenotypes ($P < 2 \times 10^{-10}$), including many loci linked to lipoprotein metabolism missed by global efforts only focussing on standard clinical measures of lipid metabolism, such as total cholesterol. We observe effects to be consistent across multiple ancestries, with only one pleiotropic region specific to British African-ancestry individuals. We integrate trans-ancestral finemapped loci with knowledge-driven machine learning classifiers to nominate effector genes that we corroborate with rare burden testing and exonic variants with strong effect sizes to establish allelic series at key metabolic genes. Finally, integrating genetic data on cardiovascular diseases, we showcase how our results can inform prioritization of targets for pharmacotherapy. Our results highlight the value of biobank-scale studies of refined metabolic traits to identify genetic regulators of human lipoprotein metabolism at scale and identify strategies for the prevention and treatment of human diseases.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6005W: A summary statistics-based toolset for identifying *trans*—regulation of proteomic pathways by cancer-related variants in the UK Biobank

Authors:

D. Dutta[1], K. Wang[1], H. Zhang[2]; [1]Natl. Cancer Inst., Rockville, MD, [2]Natl. Cancer Inst., Cambridge, MA

Abstract:

Introduction: Association of disease-related genetic variants identified in genetic studies with molecular phenotypes like gene expression and protein levels can identify mechanisms involved in disease etiology. Studies till date have primarily focused on the effects of disease-related variants on nearby genes (*cis*-associations). However, recent evidence suggest that a larger proportion of disease heritability is explained by distal (*trans*-) associations networks regulated by such variants. Since, identifying *trans*-associations is substantially underpowered due to weaker effect sizes, joint analysis of sets of genes, in a pathway, can improve power by aggregating weaker trans-associations and provide insights into mechanisms regulated by disease-related variants.**Methods:** Here, we aimed to develop a comprehensive toolset to identify the *trans*-associations of disease-related variants with gene sets or pathways through developing efficient summary statistics based multiple phenotype tests. We first propose multi-GHC, a multivariate extension of the generalized higher criticism (GHC) test, based on association summary statistics, and use a novel application of generalized pareto distribution to approximate p-values through a computationally efficient algorithm. We then developed summary statistics-based implementations for several other multiple phenotype tests, including MultiSKAT, Multi-Burden and ACAT, and constructed an integrated toolset called **Multi-*trans*** to identify *trans*-associations between given gene sets and variants associated with diseases.**Results:** Large-scale simulations showed accurate control of type I error as well as robust power under a spectrum of scenarios, by the efficient synthesis of several complimentary multiple phenotype methods. We applied Multi-trans to identify the regulation of more than 10,000 pathways curated from MSigDb, by variants associated with 18 different cancers using the publicly available summary statistics of 2940 proteins profiled in UK Biobank. We identified several pathways that were identified using aggregation of *trans*-associations with multiple variants only and could not be identified otherwise. Through detailed *in silico* analysis, we provide additional support for the identified pathways to be relevant to cancer pathogenesis.**Conclusion:** Multi-trans is a powerful and computationally efficient toolset that detects *trans* regulators of pathways and gene sets, which offers novel opportunities to understand the etiological impact of disease-associated variants on gene regulatory networks and biological processes.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6006W: Accurate HLA genotype inference in single cell RNA sequencing data by integrating allele transcript abundance with low-pass genotyping

Authors:

K. Lagattuta[1], **L. Sood**[1], S. Sakaue[1], S. Raychaudhuri[2]; [1]Harvard Med. Sch., Boston, MA, [2]Brigham and Women s Hosp., Brookline, MA

Abstract:

The Major Histocompatibility (MHC) locus, which contains the highly polymorphic Human Leukocyte Antigen (HLA) genes, accounts for more autoimmune disease risk than any other locus in the genome. Despite this overwhelming evidence, single cell RNA sequencing (RNAseq) studies rarely collect HLA genotype information. Thus, studying HLA biology at single cell resolution is currently limited by the accuracy of HLA genotype inference from RNAseq. Existing tools such as arcasHLA attempt to quantify the abundance of thousands of highly homologous HLA allele sequences, which results in poor accuracy due to severe read multimapping. Accuracy particularly suffers for class II HLA genes, and for single cell RNAseq, which captures only the 5' or 3' end of the transcript. However, it is well understood that HLA genotypes can be accurately imputed from genetic variants surrounding the HLA genes, due to extreme linkage disequilibrium within the MHC region. We hypothesized that we could call genetic variants surrounding the HLA genes, and use that information to improve HLA genotype inference in RNAseq data. To this end, we developed a new computational tool BRIGHT (Base calling from RNAseq to Infer Genotypes for HLA in Transcriptional data). In brief, BRIGHT calls genetic variants throughout the xMHC region based on psuedobulk pileups of single cell RNAseq reads. BRIGHT then estimates the probability of each HLA allele based on this surrounding genetic context, and uses the resultant probability vector to adjudicate reads mapping to multiple HLA alleles. In benchmarking analyses, BRIGHT matches arcasHLA accuracy for *HLA-A, HLA-B, HLA-C, HLA-DPA1*, and *HLA-DPB1* and notably outperforms arcasHLA for *HLA-DQA1, HLA-DQB1, and HLA-DRB1*. On average across *HLA-DRB1* residues, BRIGHT demonstrates 94.2% amino acid accuracy compared to 84.9% for arcasHLA (n=79 individuals). To our knowledge, BRIGHT is the first HLA genotype inference tool developed for single cell data, and is the first to integrate transcriptomic and genetic information. BRIGHT will equip the single cell researchers with a reliable method to infer the HLA genetic background for any individual, and thereby enable studies of how HLA genetic variation shapes disease risk through the single-cell transcriptome.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6007W: Ageing and age-related diseases are accompanied by longitudinal changes in alternative splicing dysregulation ⭐

Authors:

D. Wang[1], Y. Raza[1], A. Roberts[1], R. Bowyer[1], A. Ramisch[2,3], M. Tomlinson[1], C. Shore[1], X. Yan[1], E. Dermitzakis[2,3], C. Steves[1], J. Bell[1], C. Menni[1], J. S. E. Moustafa[1], K. Small[1]; [1]King's Coll. London, London, United Kingdom, [2]Univ. of Geneva, Geneva, Switzerland, [3]Inst. of Genetics and Genomics in Geneva, Geneva, Switzerland

Abstract:

Ageing is characterised by cellular and molecular dysregulation, including broad changes to gene expression. Previous cross-sectional studies highlighted the role of splicing regulation in ageing but failed to capture the trajectory of alternative splicing within individuals as they age and develop diseases. To fill this gap, we investigated the longitudinal relationship between alternative splicing and ageing using whole blood RNA-seq data from the MultiMuTHER study (a subset of TwinsUK) of 335 female twins (median age = 61 years) at three time points over a 6 - 8 year timespan. As a proxy for mis-splicing, intron retention (IR) exhibited longitudinal changes at 400 genes (FDR 5%), of which 88% increased over time. The median level of IR across all genes in a sample (global IR) also increased within individuals ($P = 5.2\times10^{-5}$). In contrast, time-varying gene expression levels and other types of alternative splicing (e.g., exon-skipping, alternative promoters) were relatively evenly split between increasing and decreasing. We further investigated the relationship between global IR and common age-related diseases, finding positive associations with multimorbidity ($P = 0.003$), diabetes ($P = 1.8 \times 10^{-4}$) and hypertension ($P = 0.002$). In all cases, the associations strengthened with time. Further supporting the disease associations, we observed positive interactions of diabetes ($P_{inter} = 0.016$) and hypertension ($P_{inter} = 0.020$) with time on global IR, indicating an acceleration of mis-splicing in cases vs controls. Integration with matched metabolomics revealed positive associations of global IR with metabolites involved in chronic kidney disease including p-cresol sulfate (FDR 5%). Next, we identified three RNA binding protein (RBP) motifs enriched in increasingly retained introns (FDR 5%), which are targeted by the RBPs SRSF4, ELAVL1 and AGO1. Expression levels of the genes encoding these three RBPs decreased over time and were negatively associated with global IR ($P < 0.001$) and 64% - 89% of individual IR events significant at FDR 5%, suggesting declines in these RBPs could be implicated in the age-related increase in mis-splicing. Our study provides insights into molecular mechanisms of ageing relating to the deterioration of the transcriptome. The identified RBPs could be potential targets for modulation of mis-splicing for ageing and age-related diseases. The strengthened disease associations over time suggest that

increased splicing dysregulation is a consequence of disease rather than a cause, and that living longer in ill health could lead to an accelerated deterioration of molecular regulation.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6008W: Allele specific transcription factor binding across human brain regions gives mechanistic insight into eQTLs ⭐

Authors:

A. Anderson[1,2], B. Moyers[1], J. Loupe[1], I. Rodriguez-Nunez[1], S. Felker[1], J. Lawlor[1], W. Bunney[3], A. Sequeira[3], H. Akil[4], E. Mendenhall[1], G. Cooper[1], R. Myers[1]; [1]HudsonAlpha Inst. for Biotechnology, Huntsville, AL, [2]Univ. of Alabama at Birmingham, Birmingham, AL, [3]Univ. of California, Irvine, CA, [4]Univ. of Michigan, Ann Arbor, MI

Abstract:

Transcription factors (TFs) regulate gene expression by facilitating or disrupting the formation of transcription initiation machinery at genomic loci. Given that TF occupancy is partially driven by DNA sequence recognition, genetic variation can impact TF-DNA interactions and subsequently gene regulation. Although TF binding has been extensively profiled in cell lines, TF binding in the human brain remains underexplored. In our previous work, we generated a comprehensive multi-omic resource of TF binding from bulk tissues and sorted nuclei of human post-mortem brain. Building on this resource, we conducted a comprehensive analysis to determine the influence of genetic variation on TF binding in the human brain. Utilizing graph genomes constructed from phased whole genome sequencing (WGS), we compared ChIP-seq signals between alleles to identify allele-specific binding (ASB) sites for 94 TFs across nine brain regions (DLPFC, CB, FP, OL, AMY, ANCG, DMPFC, HC, SCG). Our findings identified thousands of ASB variants, with rarer alleles more frequently reducing TF binding, while common variants were equally likely to increase or decrease binding. Motif analysis indicated that specific TFs had a higher incidence of ASB variants which alter motifs than others, and high occupancy target (HOT) sites were notably depleted for ASB variants which altered a motif. We identified ASB variants with prior evidence of allele-specific gene expression in the brain, showing correlation between ASB significance and eQTL effect direction concordance. Furthermore, 29.5% of ASB variants influenced the binding of multiple TFs, particularly for variants impacting cohesin complex members (CTCF, RAD21, SMC3). ASB variants for these TFs also exhibited a higher overlap with eQTLs and were enriched for purifying selection, suggesting that variants affecting three-dimensional genome interactions significantly contribute to gene expression variation. Neuron-specific cis-regulatory elements were markedly depleted for ASB variants. Additionally,

ASB variants demonstrated increased enrichment for neuropsychiatric disease heritability, underscoring the significance of studying TF binding within a biologically relevant context. Overall, our study provides a resource for future mechanistic studies into cis-regulatory variation in the human brain, offering insights into how genetic variation impacts gene regulation and contributes to neurological disease risk.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6009W: An atlas of the genetic basis of alternative tandem transcriptional initiation for human diseases ⭐

Authors:

L. Li, X. Zou, W. Wang, R. Ding, C. Yu; Shenzhen Bay Lab., Shenzhen, China

Abstract:

The majority of disease-associated variants were located in genomic non-coding regions. However, functional interpretation of these non-coding variants remains a significant challenge. In our previous study, we constructed the first atlas of human 3'-UTR Alternative polyadenylation Quantitative Trait Loci, which can be used to interpret approximately 16.1% of trait-associated non-coding variants. Besides variants in 3'UTR, around 50% of genes can also generate, shorten, or extend 5'UTRs by employing alternative transcription start sites (TSS), which can lead to altered functions of target mRNAs, such as mRNA stability and translational efficiency. However, genetic determinants of alternative tandem TSS usage in various human tissues and cell types and their association with human diseases remain elusive. Here we developed a new computational method (DATTSS) to identify tandem TSS usage from standard RNA-seq data. Applying this method to 28,618 RNA-seq datasets across 103 tissues and cell types reveals the widespread genetic regulation of tandem TSS usage. We identified approximately 0.41 million common genetic variants associated with tandem TSS usage of 5,436 target genes. By integrating with 90 well-powered genome-wide association studies datasets, we identified 173 and 680 disease susceptibility genes through colocalization analysis and transcriptome-wide association studies, respectively. As proof of principle validation, we show that alternative alleles facilitate 5'UTR lengthening of the *MAFF* gene, leading to an increased risk of breast cancer. Together, our findings show that alternative tandem TSS contributes substantially to the molecular mechanisms underlying human complex traits and diseases.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6010W: Annotation database updates and their impact on missense variant effect predictions

Authors:

V. Pejaver, **T. Bergquist**; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Recent work has suggested that computational/predictive tools that infer the functional and/or phenotypic impact of genetic variants can provide much stronger evidence towards the assertion of variant pathogenicity/benignity than previously stipulated by the American College of Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) guidelines for clinical variant classification, especially for missense variants (e.g., MutPred2, VEST4, AlphaMissense). For scalability and convenience, it is common practice for missense variant impact predictor developers to precompute scores for all theoretically possible variants in the human exome and make them available to end-users, often through variant annotation tools (e.g., VEP, ANNOVAR, dbNSFP) and bioinformatics resources (e.g., Ensembl, Uniprot, RefSeq). However, these resources are continually updated with the most current genetic and molecular information. These updates may involve changes to the mappings between nucleotide variant coordinates and amino acid variant positions, the actual protein sequence itself, the designation of "canonical" status to an isoform, among others. Thus, static predictions may not necessarily be in synchrony with updates to bioinformatics databases and may give discordant predictions if they were rerun using the latest information. Here, we quantify the extent to which this influences clinical variant classification by systematically predicting impact on the same variants across multiple releases of the Ensembl database. We found that among 70,566 variants from clinically relevant genes in ClinVar, 11,378 (~16%) were affected by changes in Ensembl between releases 90 and 110. Furthermore, when MutPred2 predictions on those variants were grouped into evidential strength categories as per the ACMG/AMP clinical variant classification guidelines, database changes often shifted the strength of predictive evidence of the affected variants from the evidence strength *Indeterminate* to *Supporting* pathogenicity/benignity and vice versa. Of the 11,378 affected variants, 1054 changed from *Supporting* pathogenic to either *Indeterminate* or *Supporting* benign, 3161 variants changed in their strength of evidence from *Indeterminate* to *Supporting* pathogenic or benign, and 2670 variants changed in their strength of evidence from *Supporting* benign to Indeterminate or *Supporting* pathogenic. As future ACMG/AMP guidelines envision a more prominent role for variant impact predictors in clinical variant classification, our results have direct implications for their proper incorporation into clinical workflows and pipelines.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6011W: Applying high throughput deep mutational scanning to study transcription factorsimplicated in the biology of Substance Use Disorders

Authors:

E. Orr[1], M. Grupe[1], D. Baldridge[2]; [1]Washington Univ. in Saint Louis, St. Louis, MO, [2]Washington Univ. in Saint Louis, SAINT LOUIS, MO

Abstract:

Substance use disorder (SUD) collectively affects >40 million people in the US and has significant morbidity and mortality. SUD has very high heritability (~ 40-60%) yet is a complex, polygenic, pleiotropic, and heterogenous trait. GWAS have been successful in untangling the complex genetic landscape of SUD, but determining the underlying genetic mechanisms remains elusive. In addition to alcohol and nicotine metabolizing genes and substance receptors, genes coding for transcription factors (TFs) have been implicated in SUD, most notably in reward pathways in response to drug stimulus (e.g. FosB, CRB, ELK1). However, very few genetic variants in these genes have been studied regarding their effects on TF function. Using a deep mutational scanning (DMS) approach, we are performing a direct functional assessment of the effects of thousands of missense variants in parallel. While DMS has been used in characterization of monogenic diseases, it has not yet been applied to genes implicated in SUD. We hypothesize that by performing high-throughput DMS on TFs that are part of important reward and stimulus pathways, we can not only identify distinct genetic variation that may be linked to SUD, but also provide evidence of the functional significance of key amino acid residues in the biological activity of these TFs. A comprehensive missense variant library in TF genes of interest will be constructed such that every amino acid residue at each position in the protein sequence is mutated to each of the other 19 possible alternative amino acids. We will integrate variant libraries and reporters into multiple cell types (N2A, 3T3, and HEK), and assess activity of the mutated TFs in parallel via flow cytometry measurements of fluorescent synthetic reporters (Sort-seq). Preliminary data show appropriate doxycycline dose-responsiveness of GFP fluorescence for transduced HEK cells with two TF constructs. Ongoing work will generate transcriptional activity scores for thousands of missense variants in these genes. By applying DMS variant screening to TFs associated with SUD, we will provide a molecular understanding for the genetic component of SUD and contribute to the future development of rationally designed personalized genetic treatments.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6013W: BMI Interacts with the Genome to Impact Gene Expression Globally, with Emphasis in the Brain and Gut

Authors:

R. Signer[1], C. Seah[2], H. Young[3], K. Townsley[3], A. De Pins[3], J. Johnson[4], J. Lee[5], M. Jia[5], A. Cote[6], K. Johnston[5], J. Xu[5], K. Brennand[5], L. Huckins[5]; [1]Icahn Sch. of Med. at Mount Sinai, Fort Collins, CO, [2]Icahn Sch. of Med. at Mount Sinai, New YorK, NY, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Yale Sch. of Med., New Haven, CT, [6]Icahn Sch. of Med., New York, NY

Abstract:

To fully understand the pathogenesis of complex traits we must jointly model genetics, environment, and gene-by-environment interactions. Although body mass index (BMI) shows associations with mortality, psychiatric disorders, and cancer, the genomics of these traits are modeled without consideration of the impact of BMI on gene expression and regulatory relationships. Here, we test the main effects of SNP and BMI on gene expression and identify SNP*BMI interactions that regulate expression across tissues. We then characterize the cell type and phenotypic associations of SNP-by-BMI interactions.

We use GTEx genome and mRNA sequencing to identify direct effects of BMI (Exp ~ BMI + PC1:20) and SNPs (Exp ~ SNP + PC1:20) on expression. We identify indirect effects of BMI on expression with an interaction (Exp ~ SNP + BMI + SNP*BMI + PC1:20), where genes with significant interaction coefficients are BMI-eGenes. We implement bulk mRNA deconvolution to refine associations in a cell-type specific manner. We build elastic net gene expression prediction models that include BMI*SNP interactions and apply BMI-aware models to depression in the UK Biobank.

We identified 7,563 BMI differentially expressed (BMI-DE) genes (FDR < 0.05) in 18 tissues. Endocrine tissues have the most BMI-DE genes while the brain has no BMI-DE genes. We identified 22,866 BMI-eGenes across tissues (FDR < 0.01). Among large, mid-sized, and small tissues, adipose-subcutaneous (n=6181), Brain-Cortex(n=488), and Brain-Caudate Basal ganglia (n=80) had the most BMI-eGenes. We identified immune enrichments in Colon-Transverse and Minor Salivary gland that implicate IgG immunoglobulin Complex (p=0.026) and type II interferon-mediated signaling (p=0.017). We found correlation of BMI with four immune cell types in salivary gland, microglia in the nucleus accumbens (cor=-0.17,p=6.6x10-3), and myeloid cells in transverse colon (cor=-0.12, p=0.013). The BMI-aware depression TWAS in Brain-Cortex identified genes enriched in the recent MDD GWAS (p=0.01) and novel genes. We identify 42

significant genes unique to depression that are enriched for mitochondrial RNA processing[p=1.1x10-4].
Our study provides evidence of genomic interactions with BMI that impact cellular function in the endocrine system, brain, and gut. We implicate immune and mitochondrial dynamics as processes jointly regulated by the genome and BMI with implications for depression pathogenesis. We show BMI impacts cellular function both through immune cell fraction and dynamic eQTL activity. We create and apply machine learning models to identify novel disease genes in depression and will release these models publicly for use.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6014W: Can biobanks be leveraged to help interpret variants of uncertain significance?

Authors:

C. Avery, L. A. Tang, I. Aka, R. J. Tinker, J. He, R. Hamid, J. D. Cogan, J. A. Philips, 3rd, D. M. Ruderfer, L. Bastarache; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Introduction: Despite significant advancements in the diagnosis of genetic disease through clinical sequencing, a challenge remains in the abundance of variants of uncertain significance (VUS) returned. Analyzing these VUS poses a bottleneck in the clinical genomics pipeline due to insufficient evidence to identify causative variants. Existing *in silico* tools, such as CADD, aggregate population and functional evidence to flag potentially pathogenic variants but do not provide insight into the anticipated phenotype of a variant. Biobanks linked to electronic health records (EHR) provide an opportunity to generate evidence of co-segregation of a variant and phenotype. We developed a pipeline to conduct case/control studies that uses custom phenotypes of interest to interpret VUS. **Methods:** We generated phenotype risk scores (PheRS) based on clinical symptoms of disease as defined in the Online Mendelian disease in Man (OMIM). Human phenotype ontology terms from OMIM were mapped to phecodes from the EHR to generate a PheRS, which represents a weighted sum of OMIM symptoms present for a given disease at the individual level. PheRS were averaged across all individuals with the same variant and compared to patients without that variant. We evaluated the performance of PheRS using area under the receiver operator cuve (AUROC) and the area under the precision recall curve (AUPRC). **Results:** We selected ClinVar benign and pathogenic variants (n=4009) in 14 genes implicated in dominant, multi-system disorders to simulate the use of PheRS on phenotypically heterogenous diseases. Using the eMERGESeq cohort (n = 25000) PheRS

achieved an AUROC of 0.79 and AUPRC of 0.48. PheRS scores had similar positive and negative predictive values to CADD at their respective ROC optimal thresholds (PheRS PPV=0.35, NPV = .98, CADD PPV = .40, NPV = 1), and exhibit higher accuracy in classifying benign versus pathogenic variants. Additionally, the scores were not correlated (Spearman's $\rho = 0.02$, $p = 0.17$), indicating that PheRS contributes independent information to variant interpretation. **Discussion:** This study suggests that EHR-linked biobanks can be used to generate *in silico* predictions of variant pathogenicity. Notably, PheRS can be defined using existing phenotype profiles in resources like OMIM or a personalized phenotype based on a proband's clinical presentation. This pipeline can be applied to challenging cases, such as those in the undiagnosed diseases network, to systematically re-prioritize candidate VUS. This approach can also generate statistical evidence to support VUS resolution, ultimately improving the interpretability of clinical sequencing for all patients.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6015W: CC2D1A causes ciliopathy, intellectual disability, autism, heterotaxy, renal dysplasia and abnormal CSF flow

Authors:

S. Temel[1], A. Kim[2], I. Sakin[3], S. Viviano[4], G. Tuncel[5], S. M. Aguilera[2], G. Goles[2], L. Jeffries[2], W. Ji[2], S. Lakhani[2], C. Kose[6], F. Silan[6], S. Oner[7], O. Kaplan[8], MarmaRare Group, M. Ergoren[9], K. M. Gorur[2], M. Gunel[10], S. Sag[1], E. Deniz[11]; [1]Bursa Uludag Univ., Bursa, Turkey, [2]Yale Univ., New Haven, CT, [3]Evelina London Children's Hosp., King's Coll. London, London, United Kingdom, [4]Yale Sch. of Med., New Haven, CT, [5]Near East Univ., Nicosia, Cyprus, [6]Canakkale Univ., Canakkale, Turkey, [7]Istanbul Medeniyet Univ., Istanbul, Turkey, [8]Abdullah Gul Univ., Kayseri, Turkey, [9]Sch. of Med., Near East Univ., Nicosia, Cyprus, [10]Yale Univ, New Haven, CT, [11]Yale Univ. Sch. of Med., New Haven, CT

Abstract:

Intellectual disability (ID) results from abnormal nervous system development. Over a thousand genes have been associated with ID, driving continued efforts toward dissecting variant functionality to enhance our understanding of the disease mechanism. This report identified two novel variants in *CC2D1A* in a cohort of four patients in two unrelated families. We utilized multiple model systems for functional analysis, including *Xenopus*, drosophila, and patient-derived fibroblasts. Our experiments revealed that *cc2d1a* is expressed explicitly in a spectrum of ciliated tissues, including the left-right organizer, epidermis, pronephric duct, nephrostomes, and ventricular zone of the brain. In line with this expression pattern, loss of *cc2d1a* led to

cardiac heterotaxy, cystic kidneys, and abnormal CSF circulation via defective ciliogenesis. Interestingly, when we analyzed brain development with optical coherence tomography imaging, mutant tadpoles showed abnormal cilia-driven cerebrospinal fluid circulation only in the midbrain region, suggesting abnormal local CSF flow.Furthermore, our analysis of the patient-derived fibroblasts confirmed defectiveciliogenesis, further supporting our observations. In summary, we revealed novel insight into the role of *CC2D1A* in normal development and its associated expanding diseases, including ID, autism, congenital heart disease, and kidney disease, by establishing its new critical role in ciliogenesis and cilia-driven CSF circulation.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6016W: cciQTLs: Linking Human Genetic Variation to Intercellular Communication in the Brain

Authors:

L. Sheu[1,2,3,4], C. Porras[1,2,3,4], P. B. Chandrashekar[5,6], R. Kosoy[1,2,3,4], M. Pjanic[1,2,3,4], J. Bendl[1,2,3,4], P. N.M.[1,2,3,4], K. Therrien[1,2,3,4], D. Mathur[1,2,3,4], S. P. Kleopoulos[1,2,3,4], Z. Shao[1,2,3,4], M. Alvia[1,2,3,4], C. Casey[1,2,3,4], A. Hong[1,2,3,4], S. Argyriou[1,2,3,4], P. Auluck[7], S. Marenco[7], D. Bennett[8], G. Voloudakis[1,2,3,4], V. Haroutunian[2,4,9,10], D. Wang[5,6], G. Hoffman[1,2,3,4], J. F. Fullard[1,2,3,4], D. Lee[1,2,3,4], K. Girdhar[1,2,3,4], P. Roussos[1,2,3,4,10]; [1]Ctr. for Disease Neurogenomics, Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Psychiatry, Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Friedman Brain Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Waisman Ctr., Univ. of Wisconsin-Madison, Madison, WI, [6]Dept. of Biostatistics and Med. Informatics, Univ. of Wisconsin-Madison, Madison, WI, [7]Human Brain Collection Core, Natl. Inst. of Mental Hlth.-Intramural Res. Program, Bethesda, MD, [8]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, [9]Dept. of NeuroSci., Icahn Sch. of Med. at Mount Sinai, New York, New York, NY, [10]Mental Illness Res., Ed. and Clinical Ctr.s, James J. Peters VA Med. Ctr., Bronx, NY

Abstract:

The intricate interaction within and between diverse neuronal and non-neuronal cells in the human brain enables processes like synaptic plasticity, ion/neurotransmitter regulation. Dysregulation of cell-cell interaction (CCI) alters the brain microenvironment that leads to neuropsychiatric (NPD) and neurodegenerative diseases (NDD). With the advent of single-nuclei RNA sequencing (snRNA-seq), it is now possible to quantify these signaling patterns between cells. However, how genetic variation in the human genome impacts these signaling patterns which may lead to debilitating disease remains unknown. To address this, we examined the

impact of genetic variation on cell-cell signaling to identify CCI quantitative trait loci (cciQTL) by leveraging the large-scale snRNA-seq PsychAD cohort that constitutes 6 million nuclei from the Prefrontal Cortex region of 921 Caucasian donors diagnosed with NDD and NPD. Using our in-house pipeline, a total of 26,721 CCIs were identified that showed significant interaction between ligand and receptor genes within and between excitatory neurons (EN), inhibitory neurons (IN), astrocytes (Astro), oligodendrocytes, oligo progenitors cells, Immune, mural, and endothelial cells. Out of 26,721 CCIs, 4429 cciQTLs showed a significant association between CCI scores and the genetic variation in neighboring SNPs of ligand and receptor genes within the CCI. These associations were estimated separately for ligand and receptor genes from 16 sender-receiver cell pairs. A substantial fraction (82.8%) of CCI Genes (cciGenes) showed cell type specificity indicating the presence of unique signaling patterns between specific cell types. Interestingly, receptor cciGenes, not ligand cciGenes, showed significant heritability enrichment of schizophrenia and Alzheimer's disease (AD) in neuron-glial cells and immune cells respectively. Additionally, a comparison of cciQTL with single-gene eQTLs revealed that SNPs neighboring to receptor genes are significantly associated with CCI scores while SNPs proximal to ligand genes are significantly associated with individual ligand gene expression, highlighting the role of receptors-proximal loci in intercellular signaling. Lastly, integration of cciQTL with AD genome-wide association studies (GWAS) revealed colocalization of reactive SNP rs74504435 with receptor gene EGFR in Astro interacting with ligand gene LRIG1 in EN, IN and Astro. In summary, our results show that cciQTLs provide a high-resolution view of human genetic variation that affects intercellular signaling and their relationship with disease-risk genes.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6018W: Comprehensive Mosaic Variant Characterization Across Human Tissues

Authors:

C. Grochowski, M. Gingras, S. Bhamidipati, K. Kottapalli, K. Walker, H. Chao, W. Hale, Z. Khan, Y. Zhang, J. Rogers, E. Atkinson, T. Wu, M. Niu, L. Paulin, H. H. Mehta, C. Zong, F. Sedlazeck, D. Muzny, R. Chen, R. Gibbs, H. Doddapaneni; Baylor Coll. Med., Houston, TX

Abstract:

Genomic mosaic variation represents a phenomenon in which populations of cells with different genetic compositions reside within a single individual. This type of variation can include a single nucleotide variation (SNV) or involve large chromosomal changes. Although mosaicism has been shown to play a critical role in several human disease states; the detection, validation and

interpretation of mosaic variation remains a challenge. The Genome Characterization Center at the Baylor College of Medicine, Human Genome Sequencing Center (BCM-HGSC) is a core member of the Somatic Mosaicism across Human Tissues (SMaHT) consortium whose mission is to develop new tools, methodologies and resources to understand this previously underappreciated type of variation. Pilot benchmarking studies within the consortium include mixtures of sets of: HapMap samples (HG02622, HG002, HG00257, HG02486, HG00438 and HG005); and a cancer and normal cell line (COLO829 and COLO829BL); and control fibroblast. The HapMap cell line mix varies from 0.5% (HG00438) to 83.5% (HG005) and the cancer cell line mix ranged from 1-5% in the normal cell. The composition of the mix concentrations was confirmed via droplet digital PCR (ddPCR) using 5 different genomic positions in the genes *ZNF217, SCN11A, MADD, RELN, MAP4K1.* The accuracy of the mixes ranged from 0.0% to 8.9% (relative difference in ddPCR allelic frequency) when tested at two institutions for the COLO829BLT mix. We then performed deep coverage sequencing for each sample (fibroblast, COLO829BLT and HapMap mix) with Illumina short-read WGS (433-541x coverage), long-read WGS platforms PacBio (108-121x) and ONT (106-115x). Additional tissue sample benchmarking on lung, liver, brain, colon was performed with a short-read WGS coverage of 412-490x. The extremely deep sequence coverage will facilitate discovery of low-allele fraction variants using a variety of bioinformatic tools including the DRAGEN Bio-IT Platform as well as additional custom pipelines which is currently underway. New sequencing platforms and methods focused on increasing the sensitivity of the somatic mutation discovery are also being pursued. This includes development of pipelines and metrics for duplex sequencing methodologies allowing for ultra-low error rates; Single-nucleus RNA and ATAC sequencing; use of Element Aviti instrument as an orthogonal high quality sequencing platform; and evaluation of error rates in sequence data using statistical methods. We expect the SMaHT consortium's efforts to enhance our understanding of somatic variation's role in health and disease.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6019W: Copy Number Variation and Haplotype Analysis of 17q21.31 Reveals Increased Risk Associated with Progressive Supranuclear Palsy and Gene Expression Changes

Authors:

H. Wang[1], T. S. Chang[2], B. A. Dombroski[1], P-L. Cheng[1], V. Patil[2], L. Valiente-Banuet[2], Y. Leung[1], L-S. Wang[1], J-Y. Tzeng[3], D. H. Geschwind[4], G. D. Schellenberg[1], W-P. Lee[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]North Carolina State Univ., Raleigh, NC, [4]Univ California, Los Angeles, Los Angeles, CA

Abstract:

The H1/H2 haplotype on 17q21.31 represent the foremost genetic factor contributing to the risk of progressive supranuclear palsy (PSP). Various structural forms of 17q21.31 characterized by the number of copies of α, β, and γ duplications have been identified. However, the impact of these structural forms and copy number variations (CNVs) on PSP risk has not been systematically assessed. Whole genome sequencing data of 1,684 (1,386 autopsy confirmed) individuals with PSP and 2,392 control subjects were obtained from the Alzheimer's Disease Sequencing Project (ADSP) Umbrella NG00067.v7. Copy numbers of α, β and γ were called by CNVpyter (Version 1.3.1). Expectation-maximization algorithm was used to infer the structural forms of 17q21.31 for each individual. Then, a case-control study was conducted to investigate the association of copy numbers of α, β, and γ duplications and structural forms of 17q21.31 with the risk of PSP. The copy number of γ was independently associated with the increased risk of PSP. Each additional duplication of γ was associated with 1.10 (95% CI, 1.04-1.17; $P = 0.0018$) fold of increased risk of PSP when conditioning H1 and H2. For H1 haplotype, addition γ duplications displayed a higher odds ratio for PSP: the odds ratio increases from 1.21 (95%CI 1.10-1.33, $P = 5.47 \times 10^{-5}$) for H1β1γ1 to 1.29 (95%CI 1.16-1.43, $P = 1.35 \times 10^{-6}$) for H1β1γ2, 1.45 (95%CI 1.27-1.65, $P = 3.94 \times 10^{-8}$) for H1β1γ3, and 1.57 (95%CI 1.10-2.26, $P = 1.35 \times 10^{-2}$) for H1β1γ4. The frequency of the risk sub-haplotype H1c rises from 1% in individuals with two γ copies to 88% in those with eight copies. Additionally, γ duplication up-regulates expression of *ARL17B*, *LRRC37A*, and *NSFP1*, while down-regulating *KANSL1*. Single-nucleus RNA-seq of the dorsolateral prefrontal cortex analysis reveals γ duplication primarily up-regulates *LRRC37A* in neuronal cells. These findings expand our understanding of how the complex structure of H1/H2 affect the risk of PSP.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6020W: Correlation of APOE polymorphism and its expression level with cardiac comorbidity in HIV-1 lipodystrophy patients

Authors:

S. Bharati, G. Namdev, A. Jenitha, H. Singh; ICMR-Natl. Inst. of Translational Virology and AIDS Res., Pune, India

Abstract:

Background: HIV-1 Lipodystrophy is a complex syndrome linked with metabolic disturbances. Protease inhibitors (PIs) are associated with an incidence of dyslipidemia, lipodystrophy, as well as CVDs among people living with HIV. *ApoE* play a role in transport and clearance of

lipoprotein. *ApoE* 112 (rs429358) and 158 (rs7412) polymorphisms were associated with severe dyslipidemia in the general population. **Method:** Hence, we investigated the *ApoE* polymorphism in a total of 200 HIV patients on PI (100 with HIVLD and 100 without HIVLD) and 100 healthy controls using PCR-RFLP, expression using *qRT*-PCR and its level determined using ELISA. **Results:** *APOE* allele 2 was associated with the susceptibility to HIV disease progression and severity of HIVLD (P=0.03, OR=3.02; P=0.05, OR=2.82). *APOE* 4/4 genotype was associated with reduced susceptibility to HIV disease progression (P=0.0001, OR=0.18) and allele 4 was associated with reduced risk for development and severity of HIVLD (P=0.00001, OR=0.39, P=0.0002, 0.40). *APOE* 3/4 genotype along with impaired cholesterol level was associated with development of HIVLD (P=0.02, OR=3.37). *APOE* 4/4 genotype in presence of impaired glucose level was associated with development of HIVLD (P=0.03, OR=8.27). In multivariate analysis, independent impaired cholesterol, LDL and glucose level were associated with severity of HIVLD (P=0.04, OR=2.33; P=0.001, OR=4.055, P=0.05, OR=2.63). qPCR and Elisa test analysis reveals the higher expression of *APOE* gene expression and proteins level in patients with HIVLD as compared to without HIVLD. **In conclusion:** *APOE* allele 2 may facilitate the risk for HIV disease progression and severity of HIVLD. *APOE* 3/4 genotype along impaired cholesterol and *4*/4 genotype along with glucose level may assist the development of HIVLD. *APOE* allele 4 may have a role to reduce the risk for development and severity of HIVLD.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6021W: Critical assessment of missense variant effect predictors on disease-relevant variant data

Authors:

R. Rastogi[1], R. Chung[1], S. Li[1], CAGI Annotate All Missense Participants, P. Radivojac[2], S. Brenner[1], N. Ioannidis[1]; [1]Univ. of California, Berkeley, Berkeley, CA, [2]Northeastern Univ., Boston, MA

Abstract:

Regular, systematic, and independent assessment of computational tools used to predict the pathogenicity of missense variants is necessary to evaluate their clinical and research utility and suggest directions for future improvement. Here, as part of the sixth edition of the Critical Assessment of Genome Interpretation (CAGI) challenge, we assess missense variant effect predictors (or variant impact predictors) on an evaluation dataset of rare missense variants from disease-relevant databases. Our assessment evaluates predictors submitted to the CAGI6 Annotate-All-Missense challenge, predictors commonly used by the clinical genetics community, and recently developed deep learning methods for variant effect prediction. To explore a variety

of settings that are relevant for different clinical and research applications, we assess performance within different subsets of the evaluation data and within high-specificity and high-sensitivity regimes. We find strong performance of many predictors across multiple settings. Meta-predictors tend to outperform their constituent individual predictors; however, several individual predictors have performance similar to that of commonly used meta-predictors. The relative performance of predictors differs in high-specificity and high-sensitivity regimes, suggesting that different methods may be best suited to different use cases. We also characterize two potential sources of bias. Predictors that incorporate allele frequency as a predictive feature tend to have reduced performance when distinguishing pathogenic variants from very rare benign variants, and predictors supervised on pathogenicity labels from curated variant databases often learn label imbalances within genes. Overall, we find notable advances over the oldest and most cited missense variant effect predictors and continued improvements among the most recently developed tools, and the CAGI Annotate-All-Missense challenge (also termed the Missense Marathon) will continue to assess state-of-the-art methods as the field progresses. Together, our results help illuminate the current clinical and research utility of missense variant effect predictors and identify potential areas for future development.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6022W: Critical role of the T-box Glu86 residue in TBX4 in human lung development.

Authors:

P. Szafranski[1], T. Gambin[2], Q. Ain[1], G. Deutsch[3], S. Nassef[1], M. Bailey[1], D. Kearney[4], P. Stankiewicz[5]; [1]Baylor Coll. Med., Houston, TX, [2]Inst. of Computer Sci., Warsaw Univ. of Technology, Warsaw, Poland, [3]Univ. of Washington Sch. of Med., Seattle, WA, [4]Baylor Coll. Med., and Texas Children's Hosp., Houston, TX, [5]Baylor Coll. of Med., Houston, TX

Abstract:

T-box transcription factors (TBX) are a group of evolutionarily conserved regulators of mesoderm specification and development that control gene networks in reciprocal endodermal-mesodermal cross-talks via WNT, BMP, SHH, and FGF signaling. Heterozygous single nucleotide variants (SNVs) or copy-number variant deletions involving *TBX4* have been associated with pulmonary arterial hypertension (PAH), ischiocoxopodopatellar syndrome with or without PAH, and lethal lung developmental disorders (LLDDs), including acinar dysplasia (AcDys), congenital alveolar dysplasia (CAD), and other unspecified primary pulmonary hypoplasias. Most of the 44 *TBX4* missense SNVs identified in PAH patients are scattered across its entire gene coding

region. In contrast, all five pathogenic missense SNVs (Glu86Lys, Glu86Gln, Leu186Arg, Asn243Thr, and Val218Met) found in infants with LLDD map to the DNA-binding T-box domain. Recently, loss- and gain-of-function *TBX4* variants have been proposed to cause pediatric PAH and LLDDs, and adult PAH, respectively. Here, we report another case of Glu86Lys in an unrelated AcDys newborn who died in the second hour of life. On autopsy, the lungs were found to be arrested in the pseudoglandular phase, featuring lack of acinar development, lobules composed virtually exclusively of airways with only a few saccules, all surrounded by poorly vascularized loose mesenchyme. The Glu^{86} residue is located within the *α* strand of the TBX4 T-box β-barrel lid, proposed to act as a scaffold rigidifying DNA-major-groove-binding amino acid residues. Our *in silico* prediction analyses of the TBX4 protein conformation changes, resulting from Glu86Lys and Glu86Gln substitutions, revealed in both cases the loss of hydrogen-bonds (HBs) between Glu^{86} and other highly conserved amino acid residues, including two HBs with Tyr^{230} that directly binds to DNA and six HBs with residues neighboring Tyr^{230}. Using luciferase reporter assay in fetal lung fibroblasts IMR-90, we tested Glu86Lys, Glu86Gln, and Asn243Thr variants generated by site-directed mutagenesis. In addition, RT-qPCR and immunochemistry studies showed significantly decreased expression of the *TBX4* downstream gene, *TMEM100*. Our study implies that Glu^{86} is critically involved in maintaining TBX4 transcription factor activity, essential for airway branching during early stages of human lung development, and is causatively associated with AcDys. We also propose that dominant negative effect of missense SNVs in *TBX4* leads to AcDys and CAD.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6023W: Decreased expression of Presenilin-1 gene in homozygous G206A carrier brain as a potential mechanism for rapid or accelerated cognitive decline

Authors:

K. Celis[1], S. Moura[2], A. Griswold[2], F. Rajabli[2], K. Hamilton-Nelson[3], P. Whitehead[4], J. Arvizu[5], C. Kosanovic[6], R. Castellani[7], A. Frater[8], K. Kotobelli[9], B. Appleby[9], A. Lynn[10], J. Haines[11], J. Vance[2], M. Pericak-Vance[12], J. Young[2]; [1]Univ Miami, Miami, FL, [2]Univ. of Miami, Miami, FL, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]Univ. of miami, Miami, FL, [5]Univ. Of Miami, Miami, FL, [6]The Univ. of Miami Miller Sch. of Med., Miami, FL, [7]Northwestern Univ., Chicago, IL, [8]Northwestern, Chicago, IN, [9]Case western reserve, Cleveland, OH, [10]Case Western Reserve Univ., Cleveland, OH, [11]Case Western Reserve Univ, Cleveland, OH, [12]Univ. of Miami Miller Sch. of Med, Miami, FL

Abstract:

BACKGROUND: Presenilin-1 (*PSEN1*) mutations cause autosomal dominant early-onset Alzheimer disease (AD). The G206A variant in *PSEN1,* identified in AD individuals from Caribbean Hispanic population, is associated with wide age of onset (AOO), making it a good candidate for the identification of modifier effects. Homozygous *PSEN1* mutations E280A and A431E are associated with a more aggressive phenotype than when inherited in heterozygosity. Here, we report prefrontal cortex transcriptomic analysis from homozygous and heterozygous G206A carriers that differ in AOO and rate of progression to identify molecular mechanisms involved in phenotype variability.**METHODS:** We identified two AD G206A mutation carrier brains, both with *APOE33* genotype. The first sample was a homozygous carrier (G206Ahomo), with AOO at 42 years and rapid cognitive decline leading to death at age 43. The second sample was a male heterozygous carrier (G206Ahet), with age at death 62. In addition, we utilized eight AD non-G206A carrier brains (non-G206A), all *APOE33* genotype with an average age of death of 85. We performed Single Nucleus RNA sequencing from prefrontal cortex using the10X Genomics platform. Data was analyzed using Seurat. **RESULTS:** 67,340 total nuclei were sequenced and detected eight different cell type clusters. G206A carriers comparison analysis revealed significantly decreased expression of PSEN1 in glia cells and increased expression of *PSEN1* substrates *APP*, *TMED10* and *BSG* genes in microglia in G206Ahomo vs G206Ahet. In addition, G206Ahomo had decreased expression of AD related genes like *APOE* (microglia, p-value: 1.34E-18, *GPC6* (glial, average p-value: 6.52E-15), *ABCA1* (glial, average p-value: 9.68E-09) and *SORL1* (glial, average p-value: 3.19E-21). Furthermore, expression of PSEN1 was also decreased when comparing G206A carriers vs non-G206A brains in glial cells, suggesting that the G206A variant affect expression of PSNE1. **CONCLUSION:** Our data suggest that PSNE1 G20A carriers have decreased expression of PSEN1 and abnormal protein degradation. Further, decreased expression in PSEN1 levels in G206Ahomo vs G206Ahet postulates regulation of PSNE1 levels as a candidate phenotype modifier mechanism. Identification of molecular changes associated with *PSEN1* mutations that contribute to AOO and disease progression variability will increase our understanding of *PSEN1* role in AD and highlight possible therapeutic avenues.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6024W: Discovering a delayed adipogenesis signature in visceral fat linked to metabolic dysfunction-associated steatohepatitis

Authors:

K. Z. Gelev[1], M. Alvarez[1], R. Mancina[2,3,4], F. Tavaglione[5], S. T. Lee[1], O. Jamialahmadi[2], D. Kaminska[6], V. Männistö[7,8], J. Pihlajamäki[9,10], S. Romeo[2,5,11,12], P. Pajukanta[1,13,14]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]Dept. of Molecular and Clinical Med., Inst. of Med., Univ. of Gothenburg, Gothenburg, Sweden, [3]Dept. of Life Sci., Hlth., and Hlth.Professions, Link Campus Univ., Rome, Italy, [4]Res. Unit of Internal Med., Dept. of Med. and Surgery, Università Campus Bio-Med. di Roma, Rome, Italy, [5]Clinical Med. and Hepatology Unit, Fondazione Policlinico Univ.rio Campus Bio-Med., Rome, Italy, [6]Div. of Cardiology, Dept. of Med., UCLA, Los Angeles, CA, [7]Inst. of Clinical Med., Internal Med., Univ. of Eastern Finland, Kuopio, Finland, [8]Dept. of Internal Med., Kuopio Univ. Hosp., Kuopio, Finland, [9]Inst. of Publ. Hlth.and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, [10]Dept. of Med., Endocrinology and Clinical Nutrition, Kuopio, Finland, [11]Clinical Nutrition Unit, Dept. of Med. and Surgical Sci., Univ. Magna Graecia, Catanzaro, Italy, [12]Dept. of Med., Karolinska Inst., Huddinge, Stockholm, Sweden, [13]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [14]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract:

Metabolic dysfunction-associated steatohepatitis (MASH) is characterized by hepatic steatosis, inflammation, and liver scarring. Previous studies have shown that visceral fat is associated with pathogenesis of MASH; however, the underlying mechanisms have remained elusive. We hypothesize that visceral adipose tissue (VAT) adipogenesis (i.e., adipocyte differentiation) is perturbed in MASH, leading to impaired function and low-grade inflammation. To this end, we performed single nucleus RNA-sequencing on VAT biopsies, an understudied human key metabolic tissue (13 Italians with obesity, 4 with MASH), followed by quality control (QC) to remove contaminated reads and doublets. We then integrated the data with Harmony, subset it to adipocytes (Ad), preadipocytes to adipocytes (PAd-Ad), macrophages (Macr), and monocytes to macrophages (Mon-Macr), and clustered it with Seurat. We obtained spliced and unspliced matrices of the post-QC nuclei and added these to the Seurat objects. To investigate whether differentiation is affected by MASH, we applied the dynamic model of scVelo to calculate velocity for each cell-type object and compared their latent times between the individuals with MASH and healthy liver (HL). We discovered that latent time is lower in MASH than HL for Ad and PAd-Ad, implying disrupted adipogenesis in MASH. However, when we compared Macr and Mon-Macr, we discovered that latent time is higher in MASH than HL, indicating a greater Macr differentiation and a strong invasive inflammatory response in MASH. Following this, we identified 79 genes in Ad, 439 genes in PAd-Ad, 89 genes in Macr, and 190 genes in Mon-Macr correlated with velocity and differentially ranked between MASH and HL. To deepen our insight into these gene sets, we conducted a cell-type level differential expression analysis and pathway enrichment using WebGestalt. In Ad and PAd-Ad, downregulated genes in MASH are enriched for lipid metabolism pathways, while in Macr and Mon-Macr, upregulated genes in MASH are

overrepresented in immune response pathways. We also identified metabolic dysfunction-associated steatosis liver disease (MASLD) GWAS variants within our MASH gene sets using MAGMA. Specifically, the Ad-specific gene set is significantly enriched (p=0.00065) for regional variants of imputed MASLD GWAS that we previously conducted in the UK Biobank (n=28,396 cases and 108,652 controls), thus discovering a potential genetic effect of MASH on VAT adipogenesis. All together, we identified putative genes driving altered differentiation of VAT cell-types in MASH, linking dysregulation of VAT adipogenesis as a potential mechanism contributing to MASH.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6025W: Efficient count-based models improve power and robustness for large-scale single-cell eQTL mapping

Authors:

Z. Zhang, A. Kim, N. Mancuso, S. Gazal; Univ. of Southern California, Los Angeles, CA

Abstract:

Population-scale single-cell transcriptomic technologies (scRNA-seq) enable characterizing variant effects on gene regulation at the cellular level. However, current eQTL mapping approaches (designed for bulk RNA-seq) rely on assumptions that are unrealistic when analyzing sparse counts in scRNA-seq data, which limits the ability to identify cell-type-specific eQTLs. Generalized linear models (GLM) provide a path towards modeling the natural count data. However, their implementations are computationally demanding and fail to scale to the millions of tests required for cis-eQTL testing. To address these limitations, we propose jaxQTL, a flexible framework for single-cell eQTL mapping using highly efficient GLMs given pseudo-bulk data that can run seamlessly on CPUs, GPUs, or TPUs. Using simulations, we observed that jaxQTL with a Negative Binomial (jaxQTL-NegBinom) model outperformed the Poisson and linear models in identifying eQTL effects, while maintaining a calibrated type I error. Importantly, jaxQTL was 700x and 360x times faster in computing cis-eQTL scans compared with the GLM implementation in R and Python statsmodels.We applied jaxQTL-NegBinom across 14 cell types of OneK1K scRNA-seq data (N=982 individuals). We identified 15% more eGenes compared with the linear model at FDR < 0.05 (18,907 vs. 16,390), while maintaining similar replication of lead eSNP-eGenes in eQTLGen (76%). This difference was primarily driven by jaxQTL-NegBinom ability to identify lowly expressed eGenes. Moreover, we identified 12% more eGenes than SAIGEQTL Poisson mixed effect single-cell model, highlighting that pseudo-bulk-approach for single-cell data can be powerful when appropriately modeled.Next, we systematically compared

jaxQTL-NegBinom discovery of cis-eQTLs in each cell type (sc-eQTLs) versus in all cells (bulk-eQTL). After fine-mapping, we observed that sc-eQTLs are more distal from TSS than bulk-eQTLs (P=7.18E-4), and more enriched in cell-type-specific enhancers (P=2.07E-10 when using EpiMap enhancers), suggesting that sc-eQTLs improve our ability to identify enhancers that are missed by bulk-eQTLs. Lastly, we applied S-LDSC on GWAS of 16 blood and immune-related traits and observed annotations built from sc-eQTLs explain 1.6 times more heritability (h2) than bulk-eQTL, even if the fraction of h2 was low (9.8±0.45% vs. 6±0.39%). Taken together, jaxQTL provides a robust, scalable, and powerful approach to identify cis-eQTLs in single-cell data. Our analyses further support that sc-eQTLs capture cell-type-specific regulation that can be missed by traditional bulk analyses and enhance discovery of disease-associated variants.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6026W: Elucidating the ancestry-specific architecture of expression quantitative trait loci in Alzheimer's Disease

Authors:

M. Mews[1], Y. Mustafa[2], N. Wheeler[1], T. Gu[3], L. Gomez[3], L. Adams[3], T. Starks[4], M. Cornejo-Olivas[5], M. Illanes-Manrique[6], C. Silva[7], B. Feliciano-Astacio[7], G. Byrd[8], M. Pericak-Vance[9], J. Haines[10], A. Griswold[3], W. Bush[10]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ., Westlake, OH, [3]Univ. of Miami, Miami, FL, [4]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC, [5]Univ. Cientifica del Sur, Lima, Peru, [6]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [7]Dept. of Internal Med., Univ. Central Del Caribe, Puerto Rico, PR, [8]NC A&T State Univ, Greensboro, NC, [9]Univ. of Miami Miller Sch. of Med, Miami, FL, [10]Case Western Reserve Univ, Cleveland, OH

Abstract:

Alzheimer’s disease (AD) has increased prevalence in African American (AA) and admixed Hispanic (HISP) populations relative to Non-Hispanic Whites (NHW). While it is known that the underlying genetic risk architecture for AD differs between genetic ancestries, we understand less about how these differences in the genetic architecture of AD may contribute to the observed prevalence differences. To explore how the genetic components underlying gene expression may differ among the populations and whether AD status contributes to changes in gene expression, we performed expression quantitative trait loci (eQTL) analysis from whole-blood RNAseq gene expression and genome-wide genotyping data from AA, HISP, and NHW individuals. RNAseq data was processed using the Alzheimer’s Disease Sequencing Project (ADSP) FunGen-xQTL computational protocol. Genotyping array data was imputed using the

TOPMed imputation panel and filtered to include variants with an imputation accuracy prediction of r2 > 0.8 and those directly genotyped by the array. Subsequently, we performed an ancestry-stratified eQTL analysis and included an interaction term for AD disease status. We adjusted for the following covariates in this analysis: sex, age at examination, population substructure (PC 1-3), estimated relative cell-type proportions obtained from CIBERSORTx, and other hidden factors. We identified 447,152 significant eQTLs (qvalue < 0.05) for NHW (N=235) individuals, 248,738 eQTLs for HISP (N=293), and 103,693 eQTLs for AA (N= 224) with a main genetic effect. Interestingly, we identified 54,224 significant interaction eQTLs (G*AD) for NHW, 37,601 eQTLs for HISP, and 31,634 eQTLs for AA. For eQTLs with both a main genetic effect and an interaction effect, we identified that over 99% of these variants had discordant effect size directions. Comparing populations, we identified 46,707 shared genetic effect eQTLs between NHW-HISP, 19,209 NHW-AA, 7,951 HISP-AA, and 6,219 between NHW-HISP-AA. Notably, no eQTLs with a G*AD effect were shared amongst the ancestral groups. However, we identified three genes that harbored G*AD eQTLs that were shared amongst the groups (*HES1*, antisense *RAB8A*, and pseudogene *YBX1P1*). Our results suggest prominent differences in eQTL architecture between populations with AA samples exhibiting at least a 50% reduction in eQTLs relative to NHW and HISP. The presence of exclusively ancestry-specific AD interaction eQTLs highlights the importance of personalized risk assessment and therapeutic targeting informed by genetic ancestry.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6027W: Enhanced Classification of Protein-Truncating Variants Reveals the Need for Improved Predictive Tools

Authors:

Z. Coban Akdemir[1], **J. Xu**[1], J. Klonowski[2], T. Bozkurt-Yozgatli[3], I. Egab[1], J. Posey[4], R. Gibbs[5], E. Boerwinkle[6], C. Lo[2], C. Carvalho[7], J. Lupski[4], C. Shaw[8], L. Mestroni[9], M. Taylor[10], S. Yi[11], S. Jagannathan[12], D. Kostka[2]; [1]UTHlth.Houston, Houston, TX, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Acibadem Mehmet Ali Aydinlar Univ., Houston, TX, [4]Baylor Coll. of Med., Houston, TX, [5]Baylor Coll. Med., Houston, TX, [6]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [7]Pacific Northwest Res. Inst., Seattle, WA, [8]BAYLOR Coll. OF MEDICINE, Houston, TX, [9]Univ Colorado AMC, Aurora, CO, [10]U of Colorado Anschutz Med. Campus, Aurora, CO, [11]The Univ. of Texas at Austin, Austin, TX, [12]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Classifying genetic variants as pathogenic or benign requires diverse information, varying by variant type, and often focuses on those most likely to be pathogenic. In ClinVar, ~42% of pathogenic/likely pathogenic (P/LP) variants are protein-truncating (PTC-variants) with ~16% of these classified as variants of uncertain significance (VUS). The functional impact of PTC-variants depends on whether the mRNA undergoes nonsense-mediated RNA decay (NMD), which can lead to loss-of-function (LoF) (NMD-triggering); or, mRNA can escape NMD (NMD-escape). allowing truncated proteins that may have LoF, gain-of-function (GoF), or neutral effects. Current classification methods inadequately utilize NMD information, despite its importance. NMD targets transcripts with PTC-variants in the final exon or within the last 50-55 base pairs of the penultimate exon (canonical rule), with additional rules indicating that PTC-variants <150 nucleotides from the start codon (start-proximal) or in long exons (long-exon) are likely to escape NMD.Using these rules, we annotated ~42K P/LP PTC-variants from ClinVar, finding ~32.3% predicted as NMD-escapees across 2,459 genes. In contrast, ~47% of ~19K VUS PTC-variants are predicted as NMD escapees (p-value < 2.2e-16), highlighting the classification challenge for these ambiguous variants (LoF, GoF, neutral). This underscores the need to improve classification methods for NMD-escape PTC-variants. We hypothesized a depletion of PTC-variants in gene regions predicted to escape NMD in control databases (i.e., in healthy individuals) could reveal loci associated with disease susceptibility. To test this, we annotated ~752K high-quality PTC-variants from gnomAD v4, finding 45.7% as NMD-escapees (~19.5% canonical, ~7% start-proximal, and ~20.2% long exon, referring to different NMD escape annotation rules). We identified ~18% of 14,650 canonical transcripts depleted for NMD-escape variants in control individuals, with 90% tolerant to NMD-triggering PTC-variants (pLI <0.9 or LOEUF >= 0.6), and ~26.4% associated with disease traits in OMIM. Of these, 402 genes had ≥ 1 NMD-escape P/LP variant, with 70 genes showing clustering in NMD-escape regions, including 19 known disease genes with NMD-escape mechanisms, such as *DVL1* (MIM:601365), *MN1* (MIM:156100), *NOVA2* (MIM:601991), and *PPM1D* (MIM:605100).In sum, there is a need to enhance the classification of NMD-escaping PTC-variants to determine their pathogenicity. This could be achieved by improving prediction tools that analyze the depletion of these variants in regions predicted to escape NMD in control individuals, offering insights into disease susceptibility.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6028W: Enhancer variant screens at major QT interval regulating genes

Authors:

K. Vickery[1], D. Lee[2], A. Kapoor[1]; [1]UTHlth.Houston, Houston, TX, [2]Johns Hopkins Univ Sch. of Med., Baltimore, MD

Abstract:

KCNQ1 (LQT1), *KCNH2* (LQT2) and *SCN5A* (LQT3) are the major genes implicated in long QT syndrome (LQTS), where rare deleterious coding variants together account for ~75% of LQTS cases. Along with the major locus at *NOS1AP*, genome-wide association studies (GWAS) have also identified significant associations at loci encompassing these three genes, highlighting their role in QT interval variation in the general population. However, unlike the rare LQTS, the causal variants underlying these GWAS loci remain largely unknown. Following a *cis*-regulatory hypothesis for GWAS signals in the noncoding genome, our goal is to perform enhancer-based screens for *all* QT interval associated variants at these four major loci. We have previously reported on the identification of five and six enhancer variants at the *SCN5A* and *NOS1AP* QT interval GWAS loci, respectively, from such screens. Here, we report on our enhancer screens at the *KCNQ1* and *KCNH2* based QT interval GWAS loci. At the *KCNQ1* locus, all bi-allelic variants (n=105) observed in 1000 Genomes European ancestry populations within 500kb of the sentinel hit that are common (minor allele frequency >1%) and in high linkage disequilibrium (r^2>0.9) with any of the genome-wide significant GWAS hits were selected. The same criteria led to list of 242 variants at the *KCNH2* locus, which were then overlapped with human cardiac left ventricle DNase-seq based open chromatin regions, as a hallmark of *cis*-regulation, to filter 60 candidate variants. Next, variant-centered PCR amplicons covering 104 and 57 variants at the *KCNQ1* and *KCNH2* loci, respectively, were successfully designed using Primer3 (amplicon size range: 408-1062 bp; median 647 bp). Following amplification from genotyped 1000 Genomes/HapMap reference DNA samples, both alleles for the *KCNQ1* candidate variants have been cloned upstream of a minimal promoter-driven firefly luciferase gene in pGL4.23 vector; the same for the *KCNH2* candidate variants is underway. Luciferase-reporter based enhancer assays will be performed in mouse cardiomyocyte HL1 cells in quadruplicate as before, followed by validation of the identified enhancer variants using variant-centered small deletion constructs. With our plan to also assess their endogenous activities by CRISPR/Cas9-based genomic deletions in human cardiomyocytes, altogether we expect these screens to identify common enhancer variants capable of explaining the observed trait associations at the four major QT interval regulating genes.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6029W: Evaluating the Association of ROCK1P1 Polymorphism in Glaucoma Subtypes in a Nigerian Cohort

Authors:

S. Dede[1], O. P. Elekima[1], S. L. Gillis-Harry[1], J. Okafor[1], I. D. Jaja[2], S. Cookey[3], L. Amachree[4], A. McCartney[1], J. E. Ideozu[1]; [1]MyAfroDNA, Port Harcourt, Nigeria, [2]3Rivers State Hosp. Management Board, Port Harcourt, Nigeria, [3]Rivers State Univ. Teaching Hosp., Port Harcourt, Nigeria, [4]Je Solz Eye Ctr., Port Harcourt, Nigeria

Abstract:

Background: Glaucoma is a leading cause of irreversible blindness, characterized by progressive optic neuropathy and elevated intraocular pressure. The RhoA/ROCK pathway, regulating cytoskeletal dynamics and aqueous humor outflow, is implicated in glaucoma pathogenesis. *ROCK1P1*, a pseudogene related to the *ROCK1* gene, influences this pathway, and its variant (rs34957764) has been identified as a causal variant for glaucoma in a recent African GWAS. Our study aims to replicate previous findings by examining *ROCK1P1* polymorphisms in Nigerian participants diagnosed with various types of glaucoma and related eye conditions. **Methods:** In this study, we analyzed 30 MyAfroDNA biobank participants with various eye conditions, including primary open-angle glaucoma (POAG), normal-tension glaucoma (NTG), acute closed-angle glaucoma (ACG), and related conditions. Genomic DNA was isolated using the ZymoBioMICS 96 MagBead DNA kit, and genotyping of the rs34957764 polymorphism was conducted with 3CR Bioscience PACE genotyping chemistry. We assessed the genotype distribution of the *ROCK1P1* polymorphism and its association with glaucoma. **Results:** The analysis revealed a variable distribution of rs34957764 polymorphism across all indications. The T/T genotype was observed in 10 subjects (33.3%), while the reference (C/C) genotype was found in 11 subjects (36.7%), and the heterozygous C/T genotype in 6 participants (20%). No call results were noted in 4 cases (13%). Notably, the T/T genotype was observed in a higher proportion of participants with NTG (66.7%) compared to POAG (22.2%) and other eye conditions. In POAG patients, the C/C genotype was the most common (44.4%), followed by the T/T and C/T genotypes both at 22.2%. For non-glaucoma conditions, the T/T genotype was also observed frequently (37.5%), indicating a broader distribution across different conditions but with notable prevalence in NTG. **Conclusion:** Our preliminary findings demonstrate a potential association between the T/T genotype of ROCK1P1 rs34957764 and NTG in this Nigerian cohort. These findings reinforce the potential role of the RhoA/ROCK pathway in glaucoma pathogenesis and highlight the importance of further research to explore the functional impact of this variant. Understanding these genetic associations could improve genetic screening and targeted therapies for glaucoma. The limited small sample size limits our

ability to draw definitive conclusions. Ongoing research will validate these observations, increase sample size, and genotype additional variants relevant to glaucoma.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6030W: Examining the Role of *SORL1* in AD-Specific Microglia and Neuron Endolysosomal Dysregulation

Authors:

B. DeRosa[1], Y. Zhang[2], L. Coombs[1], L. Rothschild[1], C. Golightly[1], B. Kunkle[1], M. Cuccaro[1,3], J. Vance[1,3,4], M. Pericak-Vance[1,3,4], P. St. George-Hyslop[2], D. Dykxhoorn[1,3]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Dr. John T. Macdonald Fndn. Dept. of Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [4]Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract:

Endolysosomal dysfunction has emerged as a key pathogenic mechanism in Alzheimer's disease (AD). From the group of AD-associated endolysosomal pathway genes, *SORL1* stands out as one of the strongest risk factors. SORL1 is an endocytic sorting receptor that mediates endosomal trafficking through its interaction with the retromer complex and is involved in the intracellular processing of amyloid precursor protein (APP) and the secretion of amyloid-β (Aβ). We previously identified two siblings affected with early onset AD (EOAD) who were heterozygous carriers of a rare protein-truncating deletion in *SORL1* (rs1343336951; p.C1431fs). Here, we examined the impact of the C1431fs *SORL1* variant on endolysosomal trafficking and APP processing using complementary cell-based model systems. Patient-specific induced pluripotent stem cell (iPSC) lines were derived from the two $SORL1^{+/C1431fs}$ EOAD affected siblings. CRISPR/Cas9 genome-editing was performed on the patient-derived iPSCs to create $SORL1^{+/+}$ isogenic control lines. These isogenic iPSC lines were differentiated into forebrain neurons and analyzed for defects in endosomal trafficking and APP processing. Additional analyses were carried out using HEK293-APPswe cells overexpressing either C1431fs or wild-type (WT) SORL1. Our data show that early endosomes in $SORL1^{+/C1431fs}$ neurons are enlarged ($p = 0.004$) and have increased localization of APP ($p = 0.002$) compared to isogenic control neurons. $SORL1^{+/C1431fs}$ neurons additionally have greater quantities of early endosomes per cell ($p = 0.018$). We also observe $SORL1^{+/C1431fs}$ neurons trending toward secreting higher levels of Aβ42 compared to controls. In HEK293, overexpressing C1431fs leads to increased secretion of Aβ42 ($p < 0.01$) and Aβ40 ($p < 0.01$), as well as APP soluble α-secretase (sAPPα; $p < 0.01$) and β-secretase (sAPPβ; $p < 0.01$) cleavage products. We additionally found C1431fs

increased secretion of soluble SORL1 ($p < 0.01$). Surface biotinylation experiments revealed C1431fs led to lower levels of SORL1 detected at the cell surface ($p < 0.05$). Conversely, C1431fs increased the amount of surface APP ($p < 0.05$). In conclusion, our results indicate that the *SORL1* C1431fs variant is sufficient to induce cellular defects in endosomal trafficking and APP processing. Our findings are consistent with previously reported results from similar studies of SORL1 in cultured neurons. Ongoing studies in *SORL1*$^{+/C1431fs}$ microglia and neurons will provide valuable insight into the role of *SORL1* in endolysosomal impairment in multiple cell types implicated in AD. Studies are ongoing to examine *SORL1*$^{+/C1431fs}$ microglia for defects in endolysosomal function.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6032W: Functional analyses for the late-onset Alzheimer's disease-risk variants of *SHARPIN* using CRISPR/Cas9-mediated knock-in cells

Authors:

Y. Asanomi[1], T. Kimura[1], N. Shimoda[1], D. Shigemizu[1,2], S. Niida[3], K. Ozaki[1,2,4]; [1]Med. Genome Ctr., Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [2]Dept. of Cardiovascular Med., Hiroshima Univ. Graduate Sch. of BioMed. and Hlth.Sci., Hiroshima, Japan, [3]Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [4]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan

Abstract:

Despite the recent large-scale genomic cohort studies identifying many novel genetic variants associated with late-onset Alzheimer's disease (LOAD), the direct biological and physiological insight for these variants remains largely uncovered. We have explored LOAD-risk genetic factors in Japanese patients and identified a rare functional variant of *SHARPIN*, rs572750141 (p.Gly186Arg), significantly associated with LOAD in the Japanese cohort comprising 4563 LOAD patients and 16459 controls (odds ratio = 6.1). Then, our further investigation identified another risk variant of *SHARPIN*, rs77359862 (p.Arg274Trp, odds ratio = 1.43). Additionally, recent large-scale meta-analyses of genome-wide association studies in Caucasians have identified the other common missense variants of *SHARPIN* associated with the risk of LOAD, indicating the significant relevance of *SHARPIN* into LOAD genetics. Although functional analyses of *SHARPIN* variants in our previous studies revealed aberrant cellular localization of this protein and attenuated activation of the NF-κB pathway, a central mediator of inflammatory and immune responses, all these analyses were conducted using exogenous gene transfer. Here, we

performed functional analyses using knock-in cells generated through genome editing techniques to elucidate the role of LOAD-risk *SHARPIN* variants in pathogenesis more physiologically. The CRISPR/Cas9 system enables the knock-in of the target SNVs into HEK293 cells, and we obtained knock-in cells homozygous for the mutation by single-cell cloning. Although aberrant cellular localization of the SHARPIN protein with the exogenous gene transfer did not reproduce in the knock-in cells, activation of the NF-κB pathway by TNF-α was suppressed more strongly. Recent findings have suggested that neuroinflammation in the central nervous system plays a significant role in LOAD pathogenesis. The attenuated NF-κB activity caused by the *SHARPIN* variant may increase the risk of LOAD onset by altering inflammatory and immune responses in the central nervous system. Additionally, the amount of amyloid-β (Aβ), implicated in LOAD pathogenesis, secreted into the culture medium significantly increased in the knock-in cells, although no significant change in the ratio of Aβ40 to Aβ42 was observed. These findings from *SHARPIN* knock-in cells shed light on the important role of *SHARPIN* functions in LOAD. Further investigation of *SHARPIN*-associated pathways may elucidate the mechanisms underlying the onset of LOAD.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6033W: Functional characterisation of genetic variants affecting adipocyte maturation and insulin activity in SGBS cell model

Authors:

P. Toh[1,2], M. García Ureña[1], T. Kilpeläinen[1,2]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]Novo Nordisk Fndn. Ctr. for Genomic Mechanisms of Disease, Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

Insulin resistance (IR) is a pathological condition where cells have a diminished response to insulin, resulting in elevated blood glucose levels. It is becoming increasingly prevalent with the global surge in obesity rates, increasing the incidence of cardiometabolic diseases, especially type 2 diabetes. We aimed to identify novel genetic loci associated with IR by triangulating genome-wide summary data for three hallmark traits of insulin resistance (higher fasting insulin, higher triglycerides, and lower HDL cholesterol) and functionally interpret these loci, focusing on adipose tissue. We identified 282 IR loci with joint genome-wide significance across the three traits. Many of these loci are implicated in impaired peripheral adipose storage capacity and adipocyte function. To identify loci that may increase IR via adipose tissue, we first co-localised variants that drive both IR and fat distribution. We then identified putative causal

genes by querying these loci adipose tissue eQTLs and by applying combined SNP-to-gene mapping, yielding 77 candidate genes. Utilising the adapted Activity-By-Contact model, we further identified 38 variant-gene links involving an enhancer in SBGS cells. In a pilot study, we have performed in vitro validation for the gene TNFAIF8. Its knockdown in SBGS cells significantly reduced the cells' ability to differentiate, demonstrated by an elongated cell morphology and reduced lipid droplet accumulation. Despite TNFAIP8 being part of the TNFα family as an anti-apoptotic negative regulator of the immune response, its knockdown did not significantly decrease cell viability in pre- or mature adipocytes. These findings suggest that TNFAIP8 is required for proper adipocyte differentiation in SGBS cells, independent of its anti-apoptotic function. We are performing characterisation of additional candidate IR genes by tracking changes during differentiation and performing functional assays for insulin sensitivity, glucose uptake, and lipid accumulation and metabolism in mature adipocytes. These studies will provide insights into adipocyte mechanisms that underlie the development of IR, furthering our understanding of this complex condition.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6034W: Functional characterization of Hartnup disease missense variants in *SLC6A19* shows loss of function by disparate mechanisms

Authors:

N. Sallee, H. De Jong, C. Wong, D. Ta, S. Mozaffari, V. Assimon, K. Estrada; Maze Therapeutics, South San Francisco, CA

Abstract:

Introduction: SLC6A19 encodes a sodium-dependent neutral amino acid transporter expressed primarily on the brush border of the small intestine (chaperoned to the cell surface by ACE2) and the proximal tubule of the kidney (chaperoned to the cell surface predominantly by CLTRN). SLC6A19 has emerged as a novel target for the treatment of chronic kidney disease (CKD) based on analyses of large human data sets that have identified a link between loss of function variants with improved renal function and protection from CKD. In particular, the missense variant D173N has a strong association with decreased serum creatinine in UK Biobank (beta = -0.136, p = 1.27e-24). Rare biallelic loss-of-function leads to a largely benign aminoaciduria known as Hartnup disease, with near complete loss of SLC6A19 activity. While previous studies have shown that D173N reduces transporter activity, the structural mechanism by which D173N and other Hartnup mutations confer loss of function remains poorly understood.
Results: We co-expressed SLC6A19 with chaperones CLTRN or ACE2 in a HEK293 cell model to

characterize the impact of D173N and nine other Hartnup-associated variants on the function of SLC6A19. D173N significantly reduces total SLC6A19 protein levels compared to wildtype (3-5 fold, p = 2e-16). Other Hartnup-associated variants also significantly reduce total protein levels (to varying degrees) and these effects were generally more pronounced with co-expression of CLTRN compared to ACE2. We also measured SLC6A19 protein specifically at the cell surface and D173N reduces levels, while some other Hartnup-associated variants show wildtype-like surface levels despite a reduction in total levels. In functional assays measuring the uptake of radiolabeled leucine, SLC6A19 D173N and all other Hartnup-associated variants tested show reduced uptake.
Conclusions: We were able to robustly characterize several Hartnup-associated variants in the context of CLTRN or ACE2 chaperones for their effects on protein levels, trafficking, and amino acid uptake activity. While the mechanism of action for D173N is loss of total SLC6A19 protein levels in the cell, other variants associated with Hartnup disease act by different mechanisms, reducing trafficking of SLC6A19 to the cell surface or disrupting neutral amino acid uptake activity directly. This deeper understanding of variant mechanisms informs our discovery of small molecule inhibitors of SLC6A19 that can elicit similar renal protective effects to the genetic variants.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6035W: Functional characterization of *RECQL4* missense variants: A protein level-based approach

Authors:

S. Brito, A. Monteiro, T. Nepomuceno; Moffitt Cancer Ctr., Tampa, FL

Abstract:

RECQL4 is a gene coding for a member of the RecQ helicases and is found mutated in the germline of approximately two-thirds of Rothmund-Thomson Syndrome (RTS) type 2 patients. RTS type 2 patients are at higher risk of developing osteosarcoma during childhood. Most pathogenic variants associated with RTS lead to premature protein termination, but a subset of variants are missense substitutions, making it difficult to establish their clinical significance. Here we developed an in vitro functional assay to assess the impact of missense variants in *RECQL4*. We selected 16 missense variants from the ClinVar and gnomAD databases. Our dataset consists of three known benign, two pathogenic, one likely pathogenic, and nine variants of uncertain significance (VUS). Variants were generated by site-directed mutagenesis using the pCDNA5/FRT:EGFP_RECQL4 construct. Single nucleotide variants were confirmed by

amplification-free long-read DNA sequencing, and transiently transfected into the osteosarcoma cell line U2OS. Protein expression was assessed relative to the levels of wild-type protein by immunoblotting and normalized by the β-actin loading control. We also evaluated the subcellular localization of the chimeric EGFP_RECQL4 protein in the presence or absence of the DNA-damaging agent Streptonigrin by confocal immunofluorescence. Here, we show that two of the RTS-associated pathogenic variants, p.(P466L) and p.(S1079I), result in significantly lower protein levels relative to the wild-type control. As expected, benign variants p.(E71G), p.(E267D), and p.(R522C) present protein levels comparable to the wild-type protein. The VUS p.(L566P), p.(V768A), and p.(R1058G) exhibit a reduction in protein levels similar to the pathogenic mutations tested, suggesting a possible mechanism of pathogenicity. Interestingly, the p.(P466L) variant, located in the RECQL4 nuclear localization sequence, and the VUS p.(L566P) in the helicase domain resulted in cytoplasmic mislocalization and impaired nucleolar localization upon Streptonigrin treatment. These data suggest that reduced protein expression and/or subcellular mislocalization might represent putative mechanisms of pathogenesis for *RECQL4* missense variants.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6036W: Future-proofing variant classification in *KCNH2*: The industry experience

Authors:

C. Ng[1,2], S. Padigepati[3], Y. Ting[3], F. Facio[3], M. Vatta[3], S. Aradhya[3], B. Johnson[3], S. Poll[3], K. Nykamp[3], B. Kroncke[4], J. Vandenberg[1,2], J. Reuter[3]; [1]Res. Program in Cardiac Electrophysiology, Victor Chang Cardiac Res. Inst., Darlinghurst, Australia, [2]Sch. of Clinical Med., UNSW Sydney, Sydney, Australia, [3]Invitae Corp., San Francisco, CA, [4]Div. of Clinical Pharmacology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

The National Human Genome Research Institute made 10 bold predictions for human genomics by 2030. Prediction number 7 was the clinical relevance of all encountered genomic variants will be readily predictable, rendering the diagnostic designation “variant of uncertain significance (VUS)” obsolete. To investigate whether this bold prediction can be realized for *KCNH2,* which encodes a cardiac potassium channel where loss of function can lead to sudden cardiac death, automated patch clamp functional data and multiplexed assays for variant effects (MAVE) trafficking data were generated. 906 functional data points from patch clamp and 7005 trafficking data points from MAVE were first calibrated against a set of internally curated benign

and pathogenic variant controls. The AUROC of both datasets was ≥0.8. The data were then integrated into a validated, semi-quantitative variant classification framework (Sherloc) based on negative predictive value (NPV) and positive predictive value (PPV). Based on the PPV threshold that discriminates pathogenic variant controls, pathogenic points were assigned to 563 variants from the patch clamp and 2814 variants from the MAVE data. Unfortunately, no benign points can be assigned to either dataset due to insufficient number of benign missense controls for calibrating the NPV. The clinical utility of using these experimental data to reclassify VUS was assessed in the Invitae® dataset, the largest ClinVar submitter of >1,000 *KCNH2* variants to date. We simulated the scenario by removing experimental or functional evidence applied in variant classification of *KCNH2* variants reported by Invitae. Of the 234 missense VUS that have corresponding patch clamp data, 111 were assigned additional pathogenic points, and 42 could be classified as likely pathogenic or pathogenic (LP/P), which affected 151 patients. In comparison, of the 746 missense VUS that have corresponding MAVE data, 279 were assigned additional pathogenic points, and 61 VUS could be upgraded to LP/P, affecting 188 patients. In addition, patch clamp and MAVE data would upgrade 28 and 33 LP variants to P, respectively. Finally, although 69 VUS in 767 patients and 218 VUS in 1,485 patients remained VUS from the patch clamp and MAVE data, respectively, the addition of experimental pathogenicity data moves them closer to future reclassification. In summary, functional data can contribute to VUS upgrades to LP/P and may facilitate positive testing results in the future.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6037W: Genetic mapping in sexual planarians to identify loci involved in reproductive aging rate

Authors:

M. Dawson, L. Guo; Univ. of Michigan, Ann Arbor, MI

Abstract:

Female reproductive system senescence is one of the first age-related declines experienced in many organisms. The relationship between reproductive span and total lifespan remains elusive and the underlying mechanisms behind differences in tissue aging rates are unclear. Planarians are hermaphroditic flatworms and are most well known as a model for regeneration research, as they have the ability to restore any part of their body from just a small tissue fragment. They are also known for their extremely long lifespan and are often considered to be “immortal”. Despite their extreme longevity in somatic tissues, sexual planarians undergo reproductive

senescence in only one to two years. The drastic difference between planarian lifespan and reproductive span presents an opportunity to use the organism as a model that uncouples somatic aging and reproductive decline to understand the mechanisms that drive each. Reproductive senescence rate is a complex phenotype that varies between different strains of sexual planarians. Upon tracking the number of eggs hatched in cohorts of inbred lines, S2Fn and LAF, we found that the LAF line reached reproductive senescence faster than the S2Fn line. By 500 days old, 0% of eggs laid by LAF planarians produced hatchlings. In contrast, ~20% of eggs laid by 500-day old S2Fn planarians produced hatchlings and at 600 days old, ~10% of eggs laid from S2Fn planarians still produced hatchlings. This strain specific difference suggests genetic regulations are involved in the rate of reproductive aging. In recent years, our group and others have worked to assemble and annotate the *Schmidtea mediterranea* reference genome. A chromosome-scale, haplotype-phased genome assembly now exists, containing four chromosomes and a total genome size of ~800 MB. Our highly contiguous, complete genome assembly, a well-established linkage map, and our observation of stain specific phenotypes lead us to propose the use of quantitative trait loci mapping in S2Fn and LAF planarian lines in order to uncover the genetic drivers responsible for reproductive longevity. Genetic loci identified in this “immortal” species not only have implications for the regulation of reproductive longevity, but also highlight the exact nature of reproductive aging.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6038W: Genetic regulation of the targeted plasma proteome in children and adolescents reveals obesity-, puberty-, and age-specific differences

Authors:

R. Thielemann[1], S. Stinson[1], Y. Huang[1], L. Aas Holm[1,2], J-C. Holm[2], S. Rasmussen[1], T. Hansen[1]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]The Children's Obesity Clinic, accredited European Ctr. of Obesity Management, Dept. of Paediatrics, Copenhagen Univ. Hosp. Holbæk, Holbæk, Denmark

Abstract:

Background: Protein quantitative trait loci (pQTLs) have been thoroughly investigated in adults yet studies in pediatric populations are still lacking. PQTL data has proven useful for understanding molecular mechanisms of genetic variants and inferring causal effects of proteins on complex diseases. Here, we aimed to 1) characterize the genetic architecture of the targeted plasma proteome in children and adolescents, 2) explore the genetic interplay with obesity,

puberty, and sex in early life and 3) identify differences in pQTLs across the lifespan by comparing to previous findings in adults. **Methods:** We performed genome-wide association studies on 178 plasma proteins measured with Olink proteomics Inflammation and Cardiovascular II panels in The HOLBAEK Study. We included 3,853 Danish children and adolescents (44.1% boys) with a median age of 11.7 years, recruited from an obesity clinic (n=1,758) and a population-based cohort (n=2,095) in our analysis. We performed obesity-, puberty-, and sex-stratified GWASs and compared *cis*-pQTL loci and effect sizes to identify pQTL differences. We further leveraged pQTL data from the UK Biobank Pharma Proteomics Project to compare differences in pQTL effect sizes between children and adults. **Results:** We identified 1,328 independent, genome-wide associations for 178 proteins in children and adolescents and replicated previously identified pQTLs in adult cohorts based on distance and p-value. Obesity-, puberty-, and sex-stratified analysis identified significant differences in *cis*-pQTL effect sizes (non-overlapping 95% CI) with obesity for five proteins, including TRANCE and IL1ra. We further identified significant *cis*-pQTL differences between prepubertal and pubertal/post-pubertal participants for two proteins: LOX-1 and TM. *Cis*-pQTL effect sizes for LOX1 and TM were also significantly higher in children and adolescents from The HOLBAEK Study compared to adults from the UK Biobank, suggesting age-dependent differences in genetic regulation. **Conclusion:** We performed a large-scale Olink pQTL study on 3,853 children and adolescents. We replicated previously identified pQTLs from adult cohorts but observed significant changes in *cis*-pQTL effect sizes with obesity, puberty, and age. This study addressed the lack of pQTL studies in children and adolescents and provided insights into how the genetic regulation of plasma proteins differs with obesity, puberty and across the lifespan.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6039W: Genetic Variation in *KIF1A* Gene Associated to Neuronal Disorders

Authors:

G. Lacourt Sosa[1], D. Verges Santiago[1], D. P. Torres Rodriguez[1], N. Arciniegas[1], A. Cornier[2], S. Carlo[1]; [1]Ponce Hlth.Sci. Univ., Ponce, PR, [2]San Jorge Children s & Women s Hosp., San Juan, PR

Abstract:

KIF1A-related neuronal disorder is a spectrum of disease caused by a genetic mutation in the *KIF1A* gene. Genetic variants in this gene have been known to show mild to severe symptoms, including cerebellar vermis hyperplasia, ataxia, motor and global developmental delay, as well as intention tremors. Having only 500 known incidences of this pathogenic

variant, ample research is needed to appropriately manage this condition. A 2-year-old female presents with global developmental, motor and speech delay, as well as jerky movements, possible athetosis, hand tremors, ataxia, vertical nystagmus, muscle spasticity in lower limbs and recurrent infections. She was referred to the clinic by her neurologist due to abnormal findings in her MRI, which showed cerebellar vermis hypoplasia. These results, together with her symptoms, suggested a genetical basis, for which a chromosomal microarray (CMA) was ordered. Results for the CMA came back as normal female chromosomal microarray, prompting a whole exome sequencing to be indicated. This test revealed an autosomal dominant heterozygous missense mutation known as c.464G>C (p.R155P). Because there was no significant family history correlating with this disease, it appears the mutation occurred de novo. The gene affected by this mutation is the KIF1A gene, which is located in *2q37.3*.
The *KIF1A* gene is responsible for encoding Kinesin-like protein KIF1A. This protein is responsible for microtubule motor function, as well as ATP binding. Even though the mutation found in this patient was classified as a variant of uncertain significance, her severe symptoms suggest a pathological correlation. Due to the low incidence of this disease and the possibility of neurodegeneration in patients, more research is encouraged to develop effective disease management.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6040W: Genetic variation in the immune response to COVID-19 vaccination reveals vaccine-specific associations for extremes of antibody response in the UK population involving HLA and non-HLA loci

Authors:

E. Karp-Tatham[1], A. Y. Chong[1,2], B. Alamad[1], P. Maclean[1], G. Scozzafava[1], J. Wei[3], COVIG Study Group, A. Walker[4,5], A. Mentzer[1,3], J. C. Knight[1]; [1]Ctr. for Human Genetics, Oxford, United Kingdom, [2]Ludwig Inst. for Cancer Res., Oxford, United Kingdom, [3]Big Data Inst., Oxford, United Kingdom, [4]Nuffield Dept. of Med., Oxford, United Kingdom, [5]NIHR Hlth.Protection Res. Unit in Hlth.care Associated Infections and Antimicrobial Resistance, Oxford, United Kingdom

Abstract:

Vaccination played a pivotal role in controlling the COVID-19 pandemic, helping to significantly reduce cases, hospitalisation and deaths across the world. Vaccines are known to induce variable immunological responses, and thus provide variable protective efficacy. Large studies have shown that older age, obesity, male sex and a range of chronic illnesses are independently associated with poorer COVID-19 vaccine response. Despite this, there is still limited research

into the underlying genetics of poor vaccine response.
The COVID-19 Vaccine Immune Genetics (COVIG) project was set up to advance our ability to identify individuals who have a low response to COVID-19 vaccination and understand the underlying heritable, molecular and immunological mechanisms. Participants were recruited based on antibody response phenotypes identified as part of the COVID-19 Infection Survey (CIS), organised by the UK Office for National Statistics (ONS). Serial antibody titres against COVID-19 were measured from April 2020 to March 2023 with ~535,000 volunteers. Linear mixed models estimated individual vaccine responses from repeated anti-spike Ab titres after the second vaccination with ChAdOx1 or BNT162b2 to estimate the distribution of response, and response was also categorised according to upper limit of detection (>800ng/mL) and levels indicating lack of prior infection (<42 ng/mL). Individuals at the two extremes were approached for individual informed consent for recruitment into COVIG with Axoim™ SARS-CoV-2 array genotyping, targeting recruitment of 2500 "High" and 2500 "Low" responders.
Genome-wide association (GWAS) was carried out on ~4200 individuals using 292,323,504 SNPs imputed using the TOPMed reference panel. Using quantitative anti-spike IgG levels over time (estimated peak, and values 30, 60, 90, 120, 150 days post-second vaccination) we found vaccine-specific differences between participants who received ChAdOx1 nCoV-19 and BNT162b2 vaccines. In a combined GWAS significant genetic associations were seen at Chr1 and Chr6, which were also observed in a binary High vs Low response GWAS. Further work to be presented includes on-going fine mapping of significant associations together with planned flow cytometry, immune repertoire and rare-variant analyses, which will provide further clarification and could impact future vaccine strategy and design.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6041W: Genetic variation influencing brain oxygen stress responses contributes to disease

Authors:

B. Umans, O. Allen, Y. Gilad; Univ. of Chicago, Chicago, IL

Abstract:

Many diseases of the central nervous system (CNS) have been associated with noncoding genetic variants, which are presumed to mediate disease risk through effects on gene regulation. Learning how gene regulatory variants function within different biological contexts is crucial for understanding their effects on complex traits. However, in most cases, it is unclear which genes, cell types, developmental periods, and cellular processes link genetic variants to

disease. Large-scale surveys of gene expression variation in human tissues have begun to fill this gap, but because human brain tissue can be sampled only opportunistically *post-mortem*, we still know relatively little about the function of regulatory variants in brain cell types. In particular, we lack information about how genetic factors interact with external perturbations to affect brain development and function.

To capture gene regulatory effects in contexts likely to be relevant for CNS diseases, we assayed developing cell types in dorsal brain organoids formed from a panel of 21 human iPSC lines using single-cell transcriptional profiling. To further elicit exposure-dependent regulatory effects, we reduced or elevated ambient oxygen for 24 hours before data collection. The brain's metabolic and antioxidant systems are challenged by both oxygen insufficiency (as occurs during sleep apnea or infection) and elevated oxygen (which occurs after resolution of hypoxia and in some therapeutic contexts), and the resulting oxidative stress is a feature of many pathological conditions.

To identify heritable gene regulatory differences, we mapped expression quantitative trait loci (eQTLs) using data from over 170,000 cells, including subtypes of radial glia, intermediate progenitors, excitatory and inhibitory neurons, and glia. Our ongoing analysis has identified thousands of regulatory effects which emerged only after perturbation, including in over 1,600 disease-relevant genes. We have assessed the functional importance of context-specific eQTLs by comparing them to loci nominated by genome-wide association studies (GWAS). The combination of single-cell resolution and diverse sampling conditions allowed us to discover novel associations, including 70 eQTL-GWAS alignments that implicated a different gene than the one initially identified by GWAS. This highlights the value of our experimental system for pinpointing the cell type, condition, and target gene on which regulatory variants act. In future work, we will explore the impact of gene regulatory differences in the context of other perturbations and experimentally characterize disease-associated loci in CNS tissues.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6042W: GWAS for Alzheimer Disease Using Genetic Similarity Clusters

Authors:

J. Farrell[1], C. Zhu[2], N. Kurniansyah[2], Alzheimer Disease Sequencing Project, W. S. Bush[3], E. R. Martin[4], M. A. Pericak-Vance[5], L-S. Wang[6], J. L. Haines[3], X. Zhang[7], L. A. Farrer[8]; [1]Boston Univiersity Med. Sch., Boston, MA, [2]Boston Univ. Med. Sch., Boston, MA, [3]Case Western Reserve Univ, Cleveland, OH, [4]Univ. of Miami Miller Sch. of Med., Miami, FL, [5]Univ. of Miami Miller Sch.

of Med, Miami, FL, [6]Univ. of Pennsylvania, Philadelphia, PA, [7]Boston Univ., Boston, MA, [8]Boston Univ Sch Med, Boston, MA

Abstract:

Purpose
Recently, the National Academies of Sciences recommended that researchers should use measures of genetic similarity rather than reported race and ethnicity in genetic analyses attempting to identify new disease-associated variants. Here we demonstrate bioinformatics approaches to create genetic similarity clusters (CL) within the highly diverse Alzheimer Disease Sequencing Project (ADSP).

Methods
We used GATK best practices for pre-processing and variant calling in 16,361 ADSP whole genome sequencing samples. Sites with significantly different PCR-Plus and PCR-Free allele frequencies were excluded to guard against sequencing artifacts. To account for population substructure, principal components of similarity (PC) were calculated using the GENESIS KING, PC-AIR and PC-Relate programs, and excessively heterozygous variants were filtered using RUTH. Population clusters were discerned by applying a Gaussian Mixture Model to the first 10 PCs. Next, we created an admixture profile for each cluster based on the 1000 genomes five superpopulations using SCOPE. Admixture profiles were confirmed using maternal similarity based on mitochondrial haplogroups called by Haplogrep. Association of AD with variants with AC > 20 was tested in each cluster using a generalized mixed model implemented in SAIGE including covariates for a genetic relationship matrix, age, sex, 10 PCs, and read length. Results from each cluster were combined by meta-analysis using METAL.

Results
We identified 9 clusters (CL) with profiles denoted by percentage of reference panel similarity: CL0 (European-94%); CL1 (European-93%); CL2 (African-83%, European-13%); CL3 (European-40%, African-27%, Amerindian-21%); CL4 (European-47%, Middle Eastern-45%); CL5 (Amerindian-77%, European-0.16%); CL6 (South Asian-95%); CL7 (East Asian-52% Amerindian-29%, European-16%); and CL8 (European-81%, Middle Eastern-9%). CL1 and CL4 were further resolved as Amish and Ashkenazi Jewish, respectively. The meta-analysis of these clusters found significant associations at the following loci: APOE (OR=2.3, P=1.15E-217, AF=0.23); PSEN1 (OR=51.1, P=2.94E-25, AF=0.005); and ZFP37 (OR=0.86, P=2.04E-08, AF=0.26). The rare PSEN1 missense (G206A) association in CL3 was previously established as a founder mutation in the Caribbean region. The ZFP37 gene is highly expressed in oligodendrocytes.

Conclusion
The use of Gaussian Mixture Model on PCs to ascertain genetic similarity clusters is a useful approach to meta-analyze admixed data sets without the use of self-reported race and ethnicity. The admixture profiles enable the replication of results in genetically similar cohorts.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6043W: HHIP protein interactions in lung cells provide insight into COPD pathogenesis

Authors:

D. Deritei[1], H. Inuzuka[2], P. Castaldi[3], J. Yun[4], Z. Xu[1], W. Anamika[1], J. M. Asara[5], F. Guo[6], X. Zhou[1], K. Glass[1], W. Wei[2], E. Silverman[3]; [1]Channing Div. of Network Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [2]Dept. of Pathology, Beth Israel Deaconess Med. Ctr., Harvard Med. Sch., Boston, MA, [3]Channing Div. of Network Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [4]Channing Div. of Network Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, Boston, MA, [5]Div. of Signal Transduction, Dept. of Med., Beth Israel Deaconess Med. Ctr., Harvard Med. Sch., Boston, MA, [6]Jiangsu Key Lab. of Immunity and Metabolism, Jiangsu Intl. Lab. of Immunity and Metabolism, Dept. of Pathogen Biology and Immunology, Xuzhou Med. Univ., Xuzhou, China

Abstract:

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide. The primary causes of COPD are environmental, including cigarette smoking; however, genetic susceptibility also contributes to COPD risk. Genome-Wide Association Studies (GWASes) have revealed more than 80 genetic loci associated with COPD, leading to the identification of multiple COPD GWAS genes. However, the biological relationships between the identified COPD susceptibility genes are largely unknown. Genes associated with a complex disease are often in close network proximity, i.e. their protein products often interact directly with each other and/or similar proteins. In this study, we use affinity purification mass spectrometry (AP-MS) to identify protein interactions with HHIP, a well-established COPD GWAS gene product which is part of the sonic hedgehog pathway, in two disease-relevant lung cell lines (IMR90 and 16HBE). To better understand the network neighborhood of HHIP, its proximity to the protein products of other COPD GWAS genes, and its functional role in COPD pathogenesis, we create HUBRIS, a protein-protein interaction (PPI) network compiled from 8 publicly available databases. Using additional controls from the CRAPome database, we identified 10 and 18 significant interactors of HHIP in IMR90 and 16HBE cells, respectively; 5 of these were shared by the two cell lines: UGGT1, EIF2S2, RPLP0, TFG and CALM1. These interactions do not overlap with the interactions found in public PPI databases. We find that our newly identified interactions shorten the network distance between HHIP and the protein products of several COPD GWAS genes, including DSP, MFAP2, TET2, and FBLN5. These new shorter paths include proteins that are encoded by genes involved in extracellular matrix and tissue organization. Using co-

immunoprecipitation experiments, we validated two interactions with proteins that provide new insights into COPD pathobiology: HHIP-CAVIN1 (IMR90) and HHIP-TP53 (16HBE). CAVIN1 knock-down in IMR90 cells revealed a significant decrease in IL6, an inflammatory cytokine and senescence marker, while Hhip haploinsufficient mice showed enrichment for the p53 pathway in ciliated cells, implying that the newly discovered interactions are involved in oxidative stress response and inflammation.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6044W: High-throughput identification of the impact of genetic variants on splicing in the context of the endogenous genome

Authors:

S. Adamson[1], J. A. Morris[2,1], L. Lu[1], N. E. Sanjana[1], T. Lappalainen[3,1], D. A. Knowles[1,4]; [1]New York Genome Ctr., New York, NY, [2]Univ. of Toronto, Toronto, ON, Canada, [3]SciLifeLab, Stockholm, Sweden, [4]Columbia Univ., New York, NY

Abstract:

Genome wide association studies (GWAS) have identified many trait associated variants, but understanding the molecular underpinnings of these associations is challenging. Determining these molecular mechanisms is key to understanding the genetic component of traits and identifying potential intervention points. We and others have used splicing Quantitative Trait Loci (sQTL) to show that alternative RNA splicing is a key molecular mechanism underlying Mendelian and polygenic traits. To investigate and validate the *cis* impacts of variants on splicing, researchers have developed high throughput splicing massively parallel reporter assays (MPRAs). While scalable, existing MPRAs suffer from limitations including inappropriate epigenetic environment, limited sequence context, and inability to study most intronic variants. To overcome these limitations, we have developed a new high throughput assay to engineer variants into the genome using CRISPR base editors and quantify splicing using single-cell RNA-seq. To establish this framework, we have chosen to study the impact of hematological trait associated variants on splicing. We prioritized candidate variants to study in the pilot assay by colocalizing GTEx splicing quantitative trait loci (sQTLs) with GWAS variants from 29 hematological phenotypes identified by the Blood Cell Consortium. These colocalized variants were filtered for predicted base editing and sQTL effect size. We fine-mapped these candidate sQTLs with SuSiE resulting in 141 variants across 86 genes to study in a pilot assay. We then engineered these variants into K562 cells using CRISPR base editors and assayed splicing with Parse single cell RNA-seq and hybridization capture of putatively altered splicing events. To

support experimental planning, we have developed a single cell-sQTL simulation framework to assess the impact of parameters including number of cells, read depth, sQTL effect size, multiplicity of infection, and editing efficiency on statistical power. These simulations informed experimental design and preliminary studies quantifying the efficiency of base editors. Here we demonstrate the utility of this assay to quantify the impact of splice altering variants on alternative splicing in the context of hematological traits in a pilot assay. Once established, this assay could be applied to study the role of more splice altering variants in diverse trait and cellular contexts.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6045W: Human MASLD associated *GCKR P446L* promotes an increase in fibroblast growth factor 21 via up-regulation of transcription factor ChREBP in Huh-7 cell line

Authors:

Y. Chen[1], R. Hussain[1], J. Chen[2], A. Oliveri[1], B. Halligan[1], E. Speliotes[1], the GOLD consortium; [1]Univ. of Michigan, Ann Arbor, MI, [2]Northwestern Univ., Evanston, IL

Abstract:

Background: Human GWA studies have identified a missense variant *GCKR P446L* (rs1260326) that promotes MASLD as well as hypertriglyceridemia while preventing diabetes. How this variant leads to these effects is not known. *GCKR* gene encodes glucokinase regulatory protein (GKRP) that inhibits glucokinase (GCK) to prevent glucose utilization in liver. Fibroblast growth factor (FGF) 21 is a multifaceted hepatokine that influences hepatic metabolism by promoting fatty acid oxidation, reducing lipid accumulation, and potentially improving insulin sensitivity. In this study, we test whether *P446L* variant affects levels of ChREBP and FGF21 and other key proteins related to metabolic processes in human liver Huh-7 cell line. **Material and Methods:** Using lentiviral integration, we created stable Huh-7 cell lines with WT and *P446L GCKR* overexpressed or *GCKR* knocked down using shRNAs. We measured TG and glycogen using biochemical assays. We stained cells with LipidTOX™ and quantitated lipid using CellProfiler. We measured GYS2, GCK, GKRP, ChREBP, FASN and ACC proteins using Western blot. We knockdown ChREBP in cells using *MLXIPL* siRNAs, then measured ChREBP mRNA and protein levels using qPCR and Western blot analyses respectively. We quantitated extracellular FGF21 levels using ELISA. **Results:** We generated Huh-7 cell lines overexpressing *GCKR P446L* or *WT* variants. We found that *P446L* variant significantly elevated intracellular ChREBP protein levels and FGF21 extracellular levels. In *GCKR P446L* overexpressing cells, knockdown of

CHREBP expression resulted in reduced extracellular FGF21. Overexpression of *GCKR P446L* also increased intracellular TG, glycogen, FASN, ACC, and GYS2 protein levels in Huh-7 cells. **Conclusion:** *GCKR P446L* significantly increased intracellular ChREBP protein levels and extracellular FGF21 protein levels. Knockdown of ChREBP expression reduced extracellular FGF21 protein suggesting that *GCKR P446L* drives FGF 21 secretions via up-regulation of ChREBP. These results help to define the molecular mechanisms of how *GCKR P446L* causes liver damage and other systemic effects.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6046W: IBD-focused eQTL analysis identifies novel GWAS target genes using diseased colon tissue

Authors:

N. Nishiyama[1], S. Silverstein[1], A. Bharadwaj[1], R. Ganesan[2], K. Clough[1], K. Darlington[1], G. Lian[1], M. Schaner[1], W. A. Faubion, Jr.[3], S. Sheikh[1], T. Furey[1]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Carnegie Mellon Univ., Pittsburgh, PA, [3]Mayo Clinic, Phoenix, AZ

Abstract:

While GWAS have identified over 300 loci associated with inflammatory bowel diseases (IBD), the functional mechanism and target gene of most loci remains unknown. Furthermore, eQTL mapped using any normal tissue such as the Genotype-Tissue Expression (GTEx) project only explain about 20% of GWAS loci, including IBD. Context-specific eQTL have identified regulatory mechanisms that may only be detected under specific conditions or stimuli and have provided additional explanation of GWAS loci. Therefore, we have performed the largest single tissue diseased-focused eQTL analysis using colon tissue from IBD patients. We hypothesize that eQTL mapped using diseased tissue will be enriched for disease-relevant molecular effects compared to normal tissue eQTL and will identify novel target genes contributing to IBD through colocalization with GWAS.

We performed paired genotyping and bulk RNA-sequencing on uninflamed colonic mucosa samples from 252 IBD patients (CD n=199; UC n=53). We mapped eQTL using IBD tissue for 30,173 genes using 8.4M genetic variant sites with a MAF >= 0.02 +/- 1 Mb around the gene TSS. eQTL were compared to non-IBD colon eQTL reported by the GTEx Consortium and the BarcUVa-Seq project. We then colocalized eQTL with 320 IBD GWAS loci to identify likely target genes using linkage disequilibrium (LD R2 > 0.2) and a Bayesian colocalization method coloc. We compared eQTL-GWAS colocalizations between IBD and non-IBD tissue to identify novel IBD GWAS colocalizations and target genes.

We have mapped 457,397 significant eQTL in diseased colonic tissue, associated with 4087 genes. Fifty-three lead eQTL signals using IBD tissue colocalized with 41 distinct IBD GWAS loci (coloc H4 Posterior Probability > 0.5). These colocalizations included well-known IBD GWAS candidate genes such as ERAP2, FUT2, and multiple HLA MHC class II genes. eQTL mapped in IBD tissue were 1.5 times more likely to colocalize with IBD GWAS loci compared to non-diseased tissue eQTL (Fisher's exact test p = 0.03302), despite a smaller IBD sample size (GTEx transverse colon n=368, BarcUVa colon n=445). Using diseased tissue, 47% of eQTL-GWAS colocalizations are novel. Furthermore, we detected distinct eQTL signals that colocalized with IBD GWAS loci for reported eGenes, such as the *ABO* gene, that do not colocalize using non-IBD eQTL. Additionally, our results uniquely implicated biologically relevant candidate target genes such as *TNFRSF14*, a TNF receptor and immune response regulator. In contrast, non-IBD eQTL prioritized less relevant genes for this locus. These results highlight the advantages of using diseased tissue eQTL to prioritize trait-relevant candidate genes for GWAS loci.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6047W: Identification & validation of intronic splice mutations in a cohort of inherited retinal diseases

Authors:

H. Hussain[1], A. Katiyar[1], M. Wang[1], J. Lee[1], Y. Li[1], P. Yang[2], M. Marra[2], M. Pennesi[2], R. Clark[2], F. Porto[3], R. Koenekoop[4], I. Lopez[4], R. Sui[5], M. Gorin[6], R. Chen[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Dept. of Ophthalmology, Casey Eye Inst., Oregon Hlth.& Sci. Univ., Portland, OR 97239, Portland, OR, [3]INRET Clínica e Centro de Pesquisa, Rua dos Otoni, 735/507- Santa Efigênia, Belo Horizonte, MG 30150270, Belo Horizonte, Brazil, [4]McGill Ocular Genetics Lab. and Ctr., Dept. of Paediatric Surgery, Human Genetics, and Ophthalmology, McGill Univ. Hlth.Ctr., 5252 Boul de Maisonneuve ouest, Montreal, QC H4A 3S5, Montreal, QC, Canada, [5]Dept. of Ophthalmology, Peking Union Med. Coll. Hosp., Peking Union Med. Coll., Chinese Academy of Med. Sci., WC67+HW Dongcheng, Beijing 100005, Beijing, China, [6]Dept. of Ophthalmology, Univ. of California Los Angeles David Geffen Sch. of Med., 10833 Le Conte Ave, Los Angeles, CA 90095, Los Angeles, CA

Abstract:

Purpose: Advancements in genomics, particularly the widespread adoption of whole-genome & whole-exome sequencing, have led to increased identification of variants of unknown significance (VUS). The challenge lies in the precise interpretation of noncoding variants, especially splicing mutations, which account for a significant portion of disease-causing

mutations. Using the ACMG/AMP framework, the ClinGen SVI Splicing Subgroup recently recommended a predicted score by SpliceAI of 0.5 to 1.0 for spliceogenicity, >0.1 to <0.5 as uninformative, & <0.1 for non-spliceogenicity. A discrepancy is present between the predicted score & observed impact of splicing variants. This gap has prompted the need for the validation & classification of intronic splicing candidate variants found in unsolved patients with inherited retinal disorders (IRDs). **Methods:** To increase the diagnosis yield, we have utilized SpliceAI to analyze whole genome sequencing (WGS) data of unsolved IRD patients & to identify Mendelian disease-causing variants. Starting with intronic variants in unsolved IRD patients captured by WGS, we filtered & annotated genetic alterations with a custom pipeline and SpliceAI predicted effects (with a score range >0.2) in known IRD genes. After visualization to avoid sequencing & variant calling error as well as phenotype/inheritance pattern matching, we selected 25 candidates identified in 34 IRD patients to validate splicing effect by RHC-glo minigene assay. **Results:** Our results showed 18 out of 25 splicing candidate variants can create a new splice site that leads to the inclusion of a portion of intronic sequence in the mature mRNA & causing spliceogenicity. With these experimental validation results, we got a better sense of the SpliceAI score distribution in the variants with true splicing effect. **Conclusions:** Using our optimized *in-silico* score cutoffs, we can now better guide future variant interpretation. However, *in-silico* tools & a minigene system can be utilized to prioritize & validate these splice-disrupting variants. Although these variants can be overlooked causes of human disease, they can be identified using WGS if proper interpretation guidelines are followed.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6048W: Identification of sex-specific Alzheimer's disease causal genes via proteome-wide association studies ⭐

Authors:

D. Reid, M. E. Belloy; Washington Univ. in St. Louis, St. Louis, MO

Abstract:

Women are at higher risk of developing Alzheimer’s disease (AD). Sex differences in AD pathobiology and prevalence appear to be partially attributed to genetics, but sex-stratified genome-wide association studies (GWAS) of AD have demonstrated a relative paucity of associated variants, which may reflect challenges of limited power. Furthermore, these studies lack information regarding the impact on sex-specific proteins and pathways relevant to AD, which limits clinical translation and targeted drug development. Sex-stratified protein quantitative trait locus (pQTL) studies present to address these limitations. We hypothesized

that directly integrating brain pQTL data with the largest to date sex-stratified AD GWAS would increase power to reveal sex-specific AD associated genes. Sex-stratified GWAS of AD included 410,276 female and 385,813 male non-Hispanic White individuals of European ancestry (combining samples from the ADGC, ADSP, and UKB). Two human brain proteome datasets with matching genome-wide data included 722 subjects, and 507 female and 301 male subjects. Briefly, genetic variants were associated with proteins through the FUSION package to estimate effects on protein abundance to produce non-sex-stratified ($N_{Proteins}$=2,934) and sex-stratified protein-specific variant weights ($N_{Proteins}$=1,800 for males; $N_{Proteins}$=2,463 for females). This approach only returned proteins that show significant genetic associations (i.e., protein levels displaying significant heritability). Proteome-wide association studies (PWAS) were performed via FUSION respectively combining sex-stratified AD GWAS with non-sex-stratified and sex-stratified protein weights. Enrichment/pathway analysis for gene and disease ontology was conducted via the DOSE package (R 4.3.3). Sex-matched AD PWAS identified 27 and 2 AD causal genes for females and males, respectively. Sex-stratified AD GWAS integrating non-stratified protein weights identified 30 causal genes in females and 7 in males. Upon filtering results of both discoveries for sex specificity, 18 female and 2 male AD causal genes were prioritized. Enrichment/pathway analysis of female-specific AD causal genes particularly highlighted metabolic-related health outcomes as a putative female-specific AD pathway. These findings increase our understanding of AD pathogenesis and may inform therapeutic target development relevant to sex-specific precision medicine. Supporting analyses in additional proteogenomic resources are underway.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6049W: Identification of *STRC* structural variations in patients with hereditary hearing impairment using long-read sequencing

Authors:

C-Y. Tsai[1,2], Y-S. Lu[1,3], P-H. Lin[1], Y-T. CHIANG[2,4], J. Hsu[2], C-J. Hsu[1,5], C-C. Wu[1], P-L. Chen[2,6]; [1]Dept. of Otolaryngology, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [2]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [3]Inst. of Molecular Med., Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [4]Graduate Inst. of Clinical Med., Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [5]Dept. of Otolaryngology, Buddhist Tzuchi Gen. Hosp., Taichung, Taiwan, [6]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Abstract:

Background: Hereditary hearing impairment (HHI) is the leading cause of permanent hearing loss in childhood and has been linked to over 200 genes. Over the past decade, next-generation sequencing (NGS) has proven effective in uncovering the etiology of HHI. However, the identification of structural variations (SVs) associated with large genetic rearrangements (>50bp), along with the hidden pathogenic single nucleotide variants (SNVs) perturbed from upstream pseudogene *STRCP1* (~97% identity to *STRC*), poses challenges that are not easily addressed by traditional NGS methods, particularly in the *STRC* gene, which harbors the highest prevalence of SVs associated with HHI. This study aims to evaluate the efficacy of an LRS-based approach as an innovative screening tool for detecting SVs and pathogenic single nucleotide variants (SNVs) in the clinical diagnosis of HHI. **Methods:** PacBio Single-Molecule, Real-Time (SMRT) sequencing in combination with the High Fidelity (Hi-Fi) strategy was used for the LRS assays. A total of 100 unrelated hearing-impaired patients of unknown etiology were included in this study. Targeted enrichment was achieved by three 10kbp amplicons covering the entire *STRC* region, followed by PacBio Sequel IIe platform sequencing. The multiplex ligation-dependent probe amplification (MLPA) assays were used to validate the results. **Results:** A total of 11 cases with bi-allelic *STRC* variations were identified, including seven homozygous deletions, two compound heterozygous with deletions and conversions, and two compound heterozygous deletion/conversion with SNV. The allelic ratio of SVs in mild-to-moderate HHI cases (n=73) was 13%, consistent with previously reported results in other East Asian populations. All results were confirmed by MLPA testing. Notably, the LRS assay successfully pinpointed the genetic variant cause in a multiplex family with compound heterozygous *STRC* deletion and SNV, a result not confirmed by previous short-read whole genome NGS assays. **Conclusion:** LRS characterized by direct long-range sequencing, offers advantages in elucidating challenging *STRC*-related etiologies. The development of a targeted LRS strategy represents a cost-effective option suitable for addressing such high-prevalence SVs. To our knowledge, this is the first study applying the LRS technology for genetic screening in a large HHI cohort. The study highlights the efficacy of LRS technology in deciphering underlying pathogenic SVs and SNVs using the high-throughput PacBio platform and demonstrates the potential of this evolutionary approach in clinical diagnostics for HHI.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6050W: Identifying diabetes-associated regulatory variants in HepG2 cells using massively parallel reporter assays ⭐

Authors:

J. Rosen[1], A. Tovar[2], A. S. Etheridge[1], K. Sun[1], S. Vadlamudi[1], C. M. King[1], K. Broadaway[1], K. W. Currin[1], K. Nishino[2], Z. Chen[3], D. A. Dicorpo[4], J. Meigs[5], A. Manning[5], A. Kundaje[3], K. Lorenz[6], B. F. Voight[6], S. Schoenrock[1], M. Stitzel[7], R. Tewhey[8], J. O. Kitzman[2], S. C. J. Parker[2], K. L. Mohlke[1]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Univ. of Michigan, Ann Arbor, MI, [3]Stanford Univ., Palo Alto, CA, [4]Boston Univ., Boston, MA, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Univ. of Pennsylvania, Philadelphia, PA, [7]The Jackson Lab., Farmington, CT, [8]The Jackson Lab., Bar Harbor, ME

Abstract:

While GWAS studies have identified hundreds of signals associated with type 2 diabetes (T2D) risk, most lead associated variants are non-coding, suggesting their mechanism of action is one of gene regulation, and their causal identities are not immediately obvious. To help identify causal regulatory variants for T2D, we performed massively parallel reporter assays (MPRA) in a human liver cell line, HepG2, using a library of >23,000 variants from 667 T2D GWAS signals and ~1,500 rare variants from TOPMed. We designed ~60,000 oligos (ref/alt oligos per variant), including controls and oligos to test the effect of variant positioning within the sequence. Briefly, vectors containing synthesized 200-bp sequences upstream of a promoter and reporter gene were barcoded and cloned into a lentiviral backbone. Cells were transduced and transcriptional activity was quantified by counting sequenced barcodes. We tested each sequence with both a core (SCP1) and liver-specific promoter (*APOA2*), allowing us to detect promoter-dependent activity. Our experimental design further included cell perturbations such as insulin challenge and fatty acid stimulation to identify context-dependent activity. We observed strong correlation of transcriptional activity between biological replicates (mean Pearson correlation = 0.80) and validated an established regulatory variant for LDL cholesterol, rs12740374; we observed higher transcriptional activity for rs12740374-T, consistent with the allele's increased chromatin accessibility in 138 liver samples and prior associations with increased liver *SORT1* expression. We identified additional, potentially novel T2D regulatory variants that showed allelic differences in HepG2 transcriptional activity consistent with allelic differences in liver tissue chromatin accessibility. For example, rs72695652-A is associated with increased risk of T2D (r^2 = 0.99 with the lead GWAS variant rs72695645) and increased liver chromatin accessibility, and showed increased transcriptional activity in our MPRA ($p = 4x10^{-4}$). The variant A-allele strengthens a KLF binding motif, and we hypothesize this mechanism contributes to the observed transcriptional activity. This signal is not a reported liver eQTL; the regulatory element may be linked to a nearby gene by better-powered future eQTL studies, chromatin conformation, or by quantification of the effect on gene expression in a knockdown experiment. We demonstrate how MPRA can identify regulatory variants, providing insight into mechanisms underpinning GWAS associations.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6051W: Identifying genetic variants associated with cell state abundance in T cells ⭐

Authors:

Y. Zhao[1,2,3,4], N. Sugiarto[1,2,3,4], L. Rumker[1,2,3,4,5], S. Raychaudhuri[1,2,3,4,5]; [1]Ctr. for Data Sci., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Div. of Rheumatology, Inflammation, and Immunity, Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [3]Div. of Genetics, Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Dept. of BioMed. Informatics, Boston, MA

Abstract:

T cells play crucial roles in immune mediated diseases and exhibit complex activation and functional states. Advances in single-cell RNA sequencing (scRNA-seq) technology have greatly enhanced our ability to characterize T cell subpopulations and states. Genome-wide association studies (GWAS) have identified hundreds of loci linked to autoimmune diseases and cancers, many of which influence gene expression and protein production in T cells. However, direct links between genetic variation and differential abundance of T cell states have not been characterized for the granular cell states revealed by scRNA-seq. To address this gap, we utilized the Genotype-Neighborhood Associations (GeNA) method to test associations between alleles and T cell state abundance. GeNA leverages single-cell profiling paired with genotyping information and employs principal component analysis to capture cell state variation across individuals while minimizing the testing burden. The outputs of GeNA are cell state abundance quantitative trait loci (csaQTLs). Analyzing 579,240 T cells isolated from the PBMCs of 969 genotyped individuals in the OneK1K cohort, we identified one genome-wide significant csaQTL (lead SNP rs1583810243, $p=8.53\times10^{-10}$) and four suggestive csaQTLs ($p<1\times10^{-6}$). We found that rs1583810243-T on chromosome 7 associates with expansion of a cell state characterized by high expression of a module of mitochondrial genes. This association remained significant ($p=8.57\times10^{-10}$) after all *cis*-genes of the locus were removed from the input scRNA-seq data. Interestingly, targeted pseudobulk expression QTL analyses also detect *trans*-eQTL associations for this locus to the mitochondrial encoded genes, including *MT-CO1* ($p=7.37\times10^{-37}$), *MT-CO3* ($p=5.80\times10^{-56}$), *MT-ND4* ($p=8.49\times10^{-46}$), and *MT-ATP6* ($p=3.70\times10^{-47}$). The csaQTL association was eliminated (p=0.01) after removing these genes from the single-cell dataset. This csaQTL is in strong linkage disequilibrium ($r^2=0.78$) with a missense mutation rs2304694-A in

the *TBRG4* gene, whose protein localizes to mitochondria and modulates the expression of mitochondrial RNAs, thereby regulating cellular energy balance. Our results indicate that genetic variants may affect cell state abundance through *trans*-regulation. Moving forward, we are integrating additional T cell single-cell datasets with genotyping data to further increase power. We anticipate that additional csaQTLs will be identified through meta-analysis of multiple T cell datasets. Through this work, we seek to understand molecular processes that link genetic variation to complex cellular phenotypes in health and disease.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6052W: Identifying suppressors of stress-induced neurodegeneration in a *C. elegans* knock-in *sod-1 G85R* ALS model

Authors:

M. Gallati, K. S. Yanagi, A. T. Lin-Moore, A. C. Hart; Brown Univ., Providence, RI

Abstract:

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease mainly affecting cholinergic and glutamatergic motor neurons. Despite the discovery of several causal mutations in key genes, convergent mechanisms of ALS-associated neurodegeneration have still not been identified. Therefore, identifying genetic modifiers of neurodegeneration in SOD1 ALS models can give insight into the cellular processes that drive ALS-associated neurodegeneration. Previously, the Hart lab generated a single-copy knock-in model of a causative allele of SOD1 ALS in the model nematode *C. elegans*, which leads to specific degeneration of glutamatergic and cholinergic neurons following oxidative stress (PMID: 30296255). Using this model, we performed a forward genetic screen to identify suppressors of neurodegeneration. Lines were screened by assessing the suppression of neurodegeneration in glutamatergic tail neurons (12,586 haploid chromosomes screened), and animals showing enhanced neuronal survival were isolated, ranging from minor improvements in survival to near-complete rescue. I backcrossed the three lines that show the strongest suppression of neurodegeneration to diminish the number of background mutations and to create sibling lines lacking the putative causal mutation. I sent these lines in for whole genome sequencing and, using the sibling subtraction method (PMID: 29237702), generated a list of possible causal mutations by subtracting mutations detected in non-suppressing sibling lines from those in the suppressor lines. I then used recombination mapping and complementation testing to determine the chromosomal location of the mutation and which mutation may be causing the suppression phenotype. This analysis pipeline is currently being implemented for additional suppressor lines

identified from the screen. Identification of these suppressor genes will improve our understanding of the mechanisms underlying ALS and potentially inform the development of future treatments.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6053W: Immunogenetic pleiotropy and context-specific gene regulation in the response to *Yersinia pestis*

Authors:

T. Vilgalys, M. Shiratori, A. Dumaine, L. Barreiro; Univ. of Chicago, Chicago, IL

Abstract:

The human immune system is under strong selective pressure from both infectious and non-communicable diseases, each of which requires a unique protective response. How humans adapt to these diseases remains an open question and competing selective pressures (e.g. antagonistic pleiotropy) may explain why natural selection has maintained risk alleles for common diseases. However, pleiotropy in the immune response is poorly understood. Here, we use *Yersinia pestis*, the causative agent of plague, as a model to investigate how genetic variants that affect the immune response to a deadly, historical pathogen are also active in the immune response to other diseases.

_

We stimulated peripheral blood mononuclear cells (PBMCs) from 90 individuals with live, fully virulent *Y. pestis*, and characterized the response to infection using single-cell RNA sequencing. All major PBMC cell types (monocytes, NK, B, CD4+ and CD8+ T cells) mount a robust response to *Y. pestis* infection (1,752-6,061 genes, 1% LFSR). We mapped expression quantitative trait loci (eQTL) and detect 2,388 genes with at least one eQTL. 32% of eQTL had different effects in the stimulated and unstimulated conditions (i.e., immune response eQTL), including 383 variants that only affect gene expression levels in *Y. pestis* infected cells. Immune response eQTL are systematically different than steady-state eQTL, with greater cell-type specificity and affecting genes under greater selective constraint. Immune response eQTL also have pleiotropic consequences for immune function. They are up to 4x as likely to affect the expression response to other infectious diseases and colocalize with more GWAS associations for autoimmune and inflammatory disorders than steady-state eQTL alone.

_

Together, our results suggest that genetic variation shaping the immune response to *Y. pestis* also impacts the response to other infectious agents and disease risk in modern

individuals. These findings are consistent with widespread pleiotropy in the immune response, shaping the risk for infectious and non-communicable diseases.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6054W: Insights into immunogenetic variability in the IgG heavy chain in COPD

Authors:

K. Hu[1], A. Rocco[1], M. Dransfield[1], M. H. Nahm[1], J. Pandey[2], S. P. Bhatt[1], N. Putcha[3], J. L. Curtis[4], P. Castaldi[5], W. Labaki[4], A. Gaggar[1], M-L. McDonald[6], Z. Chong[1], D. Lafon[1]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Med Univ South Carolina, Charleston, SC, [3]Johns Hopkins Univ., Baltimore, MD, [4]Univ. of Michigan, Ann Arbor, MI, [5]Brigham and Women s Hosp., Cambridge, MA, [6]Univ. of Alabama, Birmingham, Hoover, AL

Abstract:

Exacerbations of chronic obstructive pulmonary disease (ECOPD) are associated with substantial morbidity, mortality, and healthcare costs. We and others have shown that reduced antibody quantity and function are associated with ECOPD risk. It is unclear if frequent ECOPD patients have underlying genetic risk factors for antibody deficiencies. Variants in IgG heavy chain (IGHG) genes on chromosome 14 can result in impaired antibody levels and function. However, this genomic region is not well-characterized due to technical challenges in analyzing whole genome sequencing (WGS) data. To overcome this limitation, we developed a novel pipeline using short-read whole-genome sequencing (WGS) data to characterize single nucleotide variants (SNVs) in the IGHG coding region and assessed their association with COPD status. The pipeline was benchmarked against Genome in a Bottle (GIAB) samples, employing the Genome Analysis Toolkit HaplotypeCaller and a customized hard filter for the IGHG coding region (CDS) and SERPINA1 gene. Long reads from the same sample served as ground-truth for accurate short-read detection. Ancestry-stratified regression analyses were conducted in a cohort of 7,921 COPDGene participants and 2,742 SPIROMICS participants, with meta-analyses performed to assess associations across all cohorts. The pipeline demonstrated 96.7% sensitivity and 97.3% precision based on GIAB samples. Among 23 SNVs identified with COPD cases versus controls, the missense variants rs113678609 on IGHG2 (p-value=0.004076) and rs74093865 on IGHG3 (p-value=0.04409) from the African cohort displayed significant positive associations. For COPD exacerbation, 29 SNVs were identified, with 10 missense variants and 1 coding sequence variant exhibiting significantly positive associations with COPD exacerbation (OR = 1.381~28.16, p-value = 0.001426~0.04912, highest CADD = 23.2). Meta-analysis p-values ranged from 3.16E-83 to

5.87E-08 based on the effect size. Furthermore, 7 missense variants and 1 coding sequence variant were considered to have a positive association with severe COPD (OR = 1.1834~16.39, p-value = 0.006439~0.04409, highest CADD = 17.03). Additionally, the SERPINA gene missense variant rs28929474 and IGHG2 missense variant rs8009156 showed significant positive associations with COPD severity. These two variants were reported in the literature and previous research. In conclusion, we developed a novel pipeline to characterize IGHG CDS region variation, identifying potential significant variants associated with COPD risk. Future research will explore FcGR associations with COPD status, expanding analyses to other COPD cohorts.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6055W: Investigating the bi-directional regulatory effects of rs174557 and rs968567 on *FADS1* and *FADS2* expression

Authors:

S. Yang, S. Choudhary, Y. Sun, H. Zhang, K. Ye; Univ. of Georgia, Athens, GA

Abstract:

Background: Long-chain polyunsaturated fatty acids (LC-PUFAs) play critical roles in human health, and the *FADS1* and *FADS2* genes encode the two rate-limiting enzymes in their biosynthesis. These two genes are located adjacent to each other in a head-to-head orientation, suggesting a shared regulatory region. GTEx data revealed that in this region, two previously experimentally supported causal regulatory variants, rs174557 for *FADS1* and rs968567 for *FADS2*, have opposite effects on the two genes. That is, the allele associated with upregulated *FADS1* is associated with downregulated *FADS2*. However, the exact mechanisms driving these differences remain elusive.**Methods:** We performed uni-directional luciferase reporter assays to assess the effects of the regulatory variants rs174557 and rs968567 on *FADS1* or *FADS2* directions. To further verify the bi-directional regulatory impact on these genes, we performed bi-directional luciferase and fluorescent reporter assays. Additionally, we conducted prime editing in HEK293 and HepG2 cells to specifically replace the rs174557 and rs968567 variants and assessed their effects on *FADS1* and *FADS2* expression using quantitative PCR (qPCR) and RNA sequencing (RNA-Seq). Concurrently, we have implemented shRNA-mediated knockdown to explore the possibility of feedback regulation explaining the reverse effect of eQTLs on *FADS1* and *FADS2*.**Results:** Our reporter assays indicated that rs174557 exclusively affects *FADS1* expression, with the minor allele G significantly reducing the enhancer activity towards *FADS1* compared to the major allele A. No difference was observed in the

enhancer activity towards *FADS2* between these alleles. Additionally, the edited G allele did not affect the expression levels of *FADS1* or *FADS2*. Conversely, the minor allele T of rs968567 notably increased enhancer activity for both *FADS1* and *FADS2* in uni-directional and bi-directional reporter assays. Genome editing showed that rs968567-T significantly upregulated the expression of both *FADS1* and *FADS2* in HepG2 cells, but only increased *FADS2* in HEK293 cells. Knockdown experiments revealed that silencing *FADS1* led to an upregulation of *FADS2*, while silencing *FADS2* resulted in reduced *FADS1* expression.**Conclusion:** Our findings for rs174557 is in contrast with previous studies that identified rs174557 as a causal variant for *FADS1*. We provide novel evidence that rs968567 is a causal variant for *FADS2* and may also affect *FADS1*. Further studies are needed to elucidate the mechanisms by which rs968567 concurrently affects both *FADS1* and *FADS2*.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6056W: Investigating the effect of APOE variants in iPSC-derived astrocytes model of Niemann-Pick disease, type C1

Authors:

F. Khushboo, N. Cawley, F. Porter; NICHD/NIH, Bethesda, MD

Abstract:

Niemann-Pick type C1 (NPC1) is a rare, fatal, neurodegenerative, lysosomal storage disorder due to pathogenic variants in *NPC1*, which causes defects in lysosomal cholesterol transport and the accumulation of unesterified cholesterol in the endo-lysosomal compartments of cells. NPC1 disease has a broad range of visceral and progressive neurodegenerative symptoms including ataxia, cognitive decline, or psychiatric disorders, however, the highly heterogeneous nature of the disease remains poorly understood. Intriguing parallels exist in the cellular pathologies of NPC1 and Alzheimer disease (AD) including lysosomal dysfunction, neuroinflammation, and neurofibrillary tangle formation. Apolipoprotein E4 is one of the three isoforms of APOE serum protein involved in cholesterol transport and has been associated with increased risk and severity of AD. Initial analysis from the genomic sequence data of 121 NPC1 individuals showed that APOE4 is associated with earlier age of neurological onset and death. We hypothesized that APOE4 could act as a genetic modifier for NPC1 disease and used human iPSC-derived astrocytes with different APOE genotypes as our study model. At baseline, APOE4 seems to express higher levels of *APOE* mRNA than APOE3 and APOE2. We treated astrocytes with multiple doses of the U18666A compound which inhibits cholesterol transport by binding to NPC1 and mimics the phenotype of NPC1 disease. We found that 24 hours of treatment with

0.5µg/ml of U18666A could induce the disease phenotype and the intensity of unesterified cholesterol staining positively correlated with increasing doses of U18666A. APOE4 astrocytes showed greater levels of PFO staining corresponding to increased accumulation of unesterified cholesterol in acidic compartments, followed by APOE3 and APOE2. Our findings shed light on the impact of APOE isoform expression on cholesterol storage. Mimicking NPC1 disease via treatment of cultured astrocytes with the U18666A compound supports the hypothesis that expression of the APOE4 isoform in NPC1 individuals results in a more severe disease state. Future investigations are underway to better understand the mechanisms through which APOE isoforms influence the severity of NPC1 disease.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6057W: Investigation of shared and divergent contextual effects on overlapping type 1 and type 2 diabetes variants in steady state, hyperglycemic, and proinflammatory human pancreatic beta cells

Authors:

H. McMillen, A. Tovar, J. Kitzman, S. C. J. Parker; Univ. of Michigan, Ann Arbor, MI

Abstract:

Diabetes encompasses a group of common complex metabolic diseases characterized by persistent hyperglycemia. The two main types, type 1 (T1D) and type 2 (T2D), result from a variety of genetic and environmental risk factors that lead to distinct pathogenesis: autoimmune destruction of insulin-producing beta cells in the pancreas in the case of T1D and overall beta cell dysfunction in T2D. Genome-wide association studies have defined >100 signals associated with T1D and >1000 associated with T2D, most of which are non-coding and likely affect gene regulation. Interestingly, there are overlapping association signals between T1D and T2D, including a subset with opposite allelic directions of effect. While some of these associations are likely driven by distinct signals in close genomic proximity, we hypothesize that some are differentially regulated by cellular and environmental context leading to divergent roles in disease pathogenesis. Here, we used a massively parallel reporter assay (MPRA) to discover the cell-state context of shared T1D and T2D genetic variants. We constructed an MPRA library with 2,958 200-bp oligos spanning 1,453 variants linked to 11 overlapping T1D/T2D signals. We delivered this library to the EndoC-βH1 human pancreatic beta cell line in each of three conditions: (1) basal media, (2) high glucose media to mimic hyperglycemic conditions, or (3) media supplemented with the proinflammatory cytokines interleukin-1β (IL1β), tumor necrosis factor-α (TNFα), and interferon-γ (IFNγ) to model pancreatic islet immune

attack. In total, we identified 1,087 oligos that were active in the basal state, 1,232 oligos that were active in high glucose, and 1,418 oligos active after cytokine stimulation (FDR < 0.05). Further, we observed 122 variants with allelic bias across all three conditions in addition to 38, 62, and 84 variants with basal-specific, high-glucose-specific, and cytokine-specific allelic bias, respectively (FDR < 0.1). We are currently delivering this library to THP-1 macrophages and Jurkat T lymphocytes to investigate cell type contextual effects on regulatory element activity. We plan to further annotate features of these variants by utilizing existing functional genomic data (e.g. ATAC-seq, ChIP-seq, chromatin state annotations).

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6058W: *KANSL1* contributes to the gene regulatory activity of the 17q21.31 locus in the context of Parkinson's disease

Authors:

A. Hicks[1,2], J. Brenton[1,2], A. Fairbrother-Browne[1,2], M. Grant-Peters[1,2], B. O'Callaghan[3,2], R. Reynolds[1,2,4], E. Gustavsson[1,2], K. Montgomery[1,3], R. Real[3,2], H. Morris[3,2], Z. Jaunmuktane[3,2], J. Hardy[2,3], H. Plun-Favreau[3,2], M. Ryten[1,5,2,6,7]; [1]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom, [2]Aligning Sci. Across Parkinson's (ASAP) Collaborative Res. Network, Chevy Chase, MD, [3]UCL Queen Square Inst. of Neurology, London, United Kingdom, [4]Cosyne Therapeutics, London, United Kingdom, [5]UK Dementia Res. Inst. at The Univ. of Cambridge, Cambridge, United Kingdom, [6]UCL NIHR Great Ormond Street Hosp. BioMed. Res. Ctr., London, United Kingdom, [7]Dept. of Clinical NeuroSci.s, Sch. of Clinical Med., The Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

An inversion polymorphism at the 17q21.31 locus creates haplotype variation, with the H1 haplotype linked to the risk of multiple neurodegenerative disorders, including an increased risk of Parkinson's disease (PD). Although the high linkage disequilibrium at this locus has made it difficult to decipher which gene(s) drive the PD association, there is increasing evidence to support the role of *KANSL1*. This gene has been shown to regulate disease-relevant pathways, including autophagy and mitophagy, as well as the expression of other genes associated with PD. At the protein level, KANSL1 is a component of the histone acetylating non-specific lethal (NSL) complex, a primarily nuclear complex involved in transcriptional activation through histone acetylation. The main catalytic driver of this complex, namely lysine acetyltransferase 8 (KAT8), is encoded by an additional gene located at another PD-associated locus, making the regulatory activity of the NSL complex of key interest in this disease. Here, we characterised the

regulatory activity of 17q21.31 haplotype variation by performing bulk and single-nuclear RNA-sequencing on brain samples derived from control and PD patients, and analysing differential gene expression stratified by haplotype. We then assessed the contribution of *KANSL1* to this activity by comparing analysis results to that of bulk RNA-sequencing of neuroblastoma cells following *KANSL1* and *KAT8* short interfering RNA (siRNA) knockdown, as well as published transcriptomic and epigenomic datasets. At the tissue level, we detected differential expression of several genes linked to the immune system, as well as dysregulation of immune-related and lysosomal pathways between haplotypes. Dysregulation of PD-associated genes, including *GBA1* and *PRKN*, was detected in major cell types, namely excitatory neurons and oligodendrocytes respectively. *KANSL1* knockdown altered the expression of 23 genes associated with PD risk, with genes associated with lysosomal pathways also significantly enriched among likely direct targets of the NSL complex. Finally, overlaps between these NSL complex targets and those altered by haplotype variation included pathways relating to lysosomes and autophagy in astrocytes and oligodendrocytes, highlighting the likely contribution of *KANSL1* to pathologically relevant differences observed between haplotypes. Overall, these findings provide a map of the regulatory activity of 17q21.31 haplotype variation in both controls and patients diagnosed with PD, so highlighting the role of *KANSL1* and the NSL complex in the regulation of lysosomal activity and autophagy primarily in glia.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6059W: Large-scale protein variant characterization of the STIM1 calcium binding cEF-hand reveals residues critical for calcium influx

Authors:

N. Kamath[1], K. Matreyek[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Abstract:

For an average individual, there are 40-80 de novo nucleotide substitutions in the genome with 1-2 of those mutations in protein coding regions. Most of this variation is uncharacterized, called ‘variants of uncertain significance (VUS)’. One such protein that remains understudied with 336 VUS currently identified is STIM1, an endoplasmic reticulum (ER) transmembrane protein critical in amplification of intracellular calcium signaling through store operated calcium entry (SOCE). Surface receptor activation causes depletion of calcium in the ER which is sensed by the 36-amino acid canonical EF-hand (cEF-hand) of STIM1, triggering SOCE. A handful of known missense variants in the STIM1 cEF-hand can cause diseases such as Stormorken

Syndrome from high calcium influx, or combined immunodeficiency from low calcium influx. There are 11 VUS present within the cEF-hand and since this domain is critical for the process of SOCE, it is important to understand how protein-coding changes in this domain affect calcium influx and how VUS within the cEF-hand affect physiology so patients can receive the appropriate diagnosis and treatment. To address this, we comprehensively characterized the calcium influx caused by all 720 possible single amino acid variants of the STIM1 cEF-hand domain in the presence and absence of endogenous STIM1 using deep mutational scanning (DMS). DMS couples multiplex genetic assays with high throughput sequencing to assign functional scores to each variant and determine the effect of every residue on a particular phenotype. Our results show that some residues considered critical for cEF-hand structure and calcium coordination, which when mutated were expected to disrupt calcium binding, did not score as predicted and are seemingly dispensable for regulating calcium influx. We also observed a subset of variants that exhibited aberrant calcium influx only in the presence of a WT allele and not in its absence, suggesting an unappreciated dominant interfering activity. Finally, we utilized our DMS scores to characterize the calcium influx caused by the 11 VUS listed in clinical genetics databases. By performing this comprehensive mutational analysis of the STIM1 cEF-hand, we can better distinguish different types of variants from each other and thus aid in improved diagnosis and treatment of patients that present with those changes.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6061W: Loss-of-function variants in *CUL3* cause a syndromic neurodevelopmental disorder ⭐

Authors:

T. Wang[1], P. Blackburn[2], F. Ebstein[3], T-C. Hsieh[4], M. Motta[5], I. Thiffault[6], T. Smol[7], C. Vincent-Delorme[8], B. Cogné[9], P. Joset[10], G. Ferrero[11], A. Ciolfi[5], C. Bacino[12], C. Macmurdo[13], J. Rosenfeld[12], L. Faivre[14], D. Bonneau[15], I. Krapels[16], T. Brunet[17], C. Mignot[18], G. Zampino[19], S. Wortmann[20], R. Feichtinger[21], E. Krüger[22], E. Eichler[23], M. Tartaglia[24], S. Küry[25]; [1]Peking Univ., Beijing, China, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Nantes Université, CHU Nantes, CNRS, INSERM, l'institut du thorax, F-44000, Nantes, France, [4]Univ. Hosp. of Bonn, Bonn, Germany, [5]Molecular Genetics and Functional Genomics, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy, [6]Children s mercy, Kansas City, MO, [7]Univ Lille, CHU Lille, RADEME Team, Inst. de Génétique Médicale, F-59000, Lille, France, [8]CH Arras, Arras, Cedex, France, [9]CHU NANTES, Nantes, France, [10]Med. Genetics, Inst. of Med. Genetics and Pathology, Univ. Hosp. Basel, Basel, Switzerland, [11]Univ. of Torino, Torino, Italy, [12]Baylor Coll. of Med., Houston, TX, [13]Baylor Scott and White Hlth., Temple, TX, [14]CHU Dijon, Dijon, France, [15]CHU Angers Genetique et Biochimie,

Angers, France, [16]Maastricht Univ. Med. Ctr., Maastricht, Limburg, Netherlands, [17]Inst. of Human Genetics, Klinikum rechts der Isar, Sch. of Med., Technical Univ. of Munich, Munich, Germany, [18]APHP.Sorbonne Université, Département de Génétique, Hôpital Trousseau & Groupe Hosp.ier Pitié-Salpêtrière, Paris, France, [19]Ctr. for Rare Diseases and Birth Defects, Dept. of Woman and Child Hlth.and Publ. Hlth., Fondazione Policlinico Univ.rio A. Gemelli IRCCS, Rome, Italy, [20]Amsterdam Univ. Med. Ctr., Univ. of Amsterdam, Dept. of Clinical Genetics, Muenchen, Germany, [21]Univ. Children's Hosp., Paracelsus Med. Univ. (PMU), Salzburg, Austria, [22]Inst. of Med. Biochemistry and Molecular Biology, Univ. Med. Greifswald, Greifswald, Germany, [23]Univ of Washington, Seattle, WA, [24]Ospedale Pediatrico Bambino Gesù IRCCS, Rome, Italy, [25]Nantes Université, CHU de Nantes, CNRS, INSERM, l'institut du thorax, F-44000, Nantes, France

Abstract:

Purpose: *De novo* variants in *CUL3* (Cullin-3 ubiquitin ligase) have been strongly associated with neurodevelopmental disorders (NDDs), but no large case series have been reported so far. Here we aimed to collect sporadic cases carrying rare variants in *CUL3,* describe the genotype-phenotype correlation, and investigate the underlying pathogenic mechanism. **Methods**: Genetic data and detailed clinical records were collected via multi-center collaboration. Dysmorphic facial features were analyzed using GestaltMatcher. Variant effects on CUL3 protein stability were assessed using patient-derived T-cells. **Results**: We assembled a cohort of 35 individuals with heterozygous *CUL3* variants presenting a syndromic NDD characterized by intellectual disability with or without autistic features. Of these, 33 have loss-of-function (LoF) and two have missense variants. *CUL3* LoF variants in patients may affect protein stability leading to perturbations in protein homeostasis, as evidenced by decreased ubiquitin-protein conjugates *in vitro*. Specifically, we show that cyclin E1 (CCNE1) and 4E-BP1 (EIF4EBP1), two prominent substrates of CUL3, fail to be targeted for proteasomal degradation in patient-derived cells. **Conclusion**: Our study further refines the clinical and mutational spectrum of *CUL3*- associated NDDs, expands the spectrum of cullin RING E3 ligase-associated neuropsychiatric disorders, and suggests haploinsufficiency via LoF variants is the predominant pathogenic mechanism.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6062W: MAGPIE accurate pathogenic prediction for multiple variant types using machine learning approach

Authors:

N. Shen, Y. Liu, T. Zhang, N. You, S. Wu; Zhejiang Univ., Hangzhou, China

Abstract:

Identifying pathogenic variants from the vast majority of nucleotide variation remains a challenge. We present a novel method named Multimodal Annotation Generated Pathogenic Impact Evaluator (MAGPIE) that predicts the pathogenicity of multi-type variants using a tree-based model. MAGPIE uses the ClinVar dataset for training and demonstrates superior performance in both independently held-out test set and multiple orthogonal validation datasets, accurately predicting multiple types of variant pathogenicity. Notably, MAGPIE performs best in predicting the pathogenicity of rare variants and highly imbalanced datasets. Furthermore, MAGPIE outperforms other tools, including the state-of-the-art models, in classifying known pathogenic mutations. Overall, results underline the robustness of MAGPIE and highlight its potential as a valuable tool for predicting pathogenicity in various types of human genome variations.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6063W: Metabolomic profiling reveals distinct and common biological pathways in COPD and IPF

Authors:

A. Rosario[1], N. Prince[1], J. Yun[1], C. Clish[2], R. Gerszten[3], E. Silverman[1], R. Kelly[1]; [1]Brigham & Women's Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Beth Israel Deaconess Med. Ctr., Boston, MA

Abstract:

Chronic Obstructive Pulmonary Disease (COPD) and Idiopathic Pulmonary Fibrosis (IPF) are chronic respiratory diseases with high mortality and global disease burden. While some risk factors, such as advanced age and cigarette smoking, are associated with both diseases, they are pathologically quite distinct. COPD is characterized by emphysema and airway disease, while IPF is characterized by fibrosis. Genetic studies have identified five shared genetic loci for COPD and IPF demonstrating opposite directions of effect, but the downstream consequences of these genetic loci remain unexplored. Metabolomic profiling offers insights into shared and unique metabolic pathways influenced by genetics and environment. In this study, we performed untargeted metabolomic profiling of plasma samples from 1507 individuals enrolled in the Trans-Omics for Precision Medicine (TOPMed) initiative as part of the Lung Tissue Research Consortium (LTRC), including 480 COPD, 219 IPF, and 361 control subjects. Logistic regression modeling was applied to identify metabolites significantly associated with COPD and IPF; to account for multiple comparisons, the false discovery rate (FDR) was controlled using the

Benjamini-Hochberg procedure, with an adjusted p-value (Padj) < 0.05 considered statistically significant. Our analysis revealed that 248 (25.03%) of 991 metabolites (ORs = 0.31-2.99, Padj = 1.67x10-8 to 0.05) differed significantly between COPD and control samples. These differences included lower levels of metabolites involved in steroid hormone biosynthesis and tryptophan catabolism, as well as higher levels of metabolites involved in sphingolipid metabolism in COPD patients. Comparison between the IPF and control group revealed 133 (13.42%) significant metabolites (ORs = 0.31-6.5, Padj = 7.43x10-7 to 0.05), including lower levels of steroid hormone biosynthesis as well as higher levels of metabolites of the polyamine metabolism pathway in the IPF group. 39 (3.94%) metabolites differed significantly between the COPD and IPF groups (ORs = 0.21-3.74 Padj = 1.47x10-5 - 0.05), including metabolites of tobacco exposure and tryptophan catabolism. In summary, our metabolomic profiling of COPD and IPF patients revealed distinct metabolomic signatures compared to controls and to each other. While steroid hormone biosynthesis metabolites were commonly found at lower levels in both diseases, dysregulation of other key pathways demonstrated disease specific patterns. These findings deepen our understanding of disease mechanisms and offer potential avenues for targeted therapeutic interventions.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6064W: Mis-ascertainment of human gene regulatory architecture is caused by evolutionary history, but can be improved by comparison with farm species

Authors:

N. Connally, S. Sunyaev; Harvard Med. Sch., Boston, MA

Abstract:

eQTLs are a mechanism for some non-coding GWAS associations, and have been identified for nearly every human gene. Connecting GWAS hits and eQTLs has the potential to improve our understanding of traits and the identification of drug targets. However, ascertained eQTLs fail to explain most GWAS hits. Notably, using colocalization—tests of whether the same variant causally affects both a GWAS trait and gene expression—only a third of GWAS loci are matched to a bulk-tissue eQTL.
Natural selection acts on trait-associated variants, and has been proposed as an explanation for limited power to detect trait-associated eQTLs. But it does not explain the abundance of observed eQTLs that act on constrained, trait-associated genes, but are absent in GWAS. Additionally, our current understanding provides limited guidance for increasing discovery of

trait-associated eQTLs, because there are multiple ways in which selection's effects could limit power. The field would benefit both from a better understanding of which eQTLs to look for (cell type-specific, response, developmental, etc.) and from improvements in the statistical power to detect them.
To better understand the role of evolution in shaping regulation, we compare humans to two species with well-powered eQTL studies, but dissimilar evolutionary histories (cattle and pigs). In both species, GWAS associations are over twice as likely to colocalize with bulk-tissue eQTLs, despite smaller sample sizes. We show that ascertained eQTLs in cattle and pigs have an architecture that comprehensively differs from humans. This regulatory architecture not only matches the architecture of cattle and pig GWAS; it also matches human GWAS more closely than human regulation does. Though selection plays a role in eQTL detection and colocalization, we show that demography—specifically, effective population size over time—is also required to explain species differences. As an example, we identify instances where colocalization fails for a human trait-associated gene, but succeeds with an analogous trait in cattle or pigs. We also propose using variant prioritization based on population and comparative genetics to improve our ability to uncover trait-associated eQTLs.
Our work suggests that difficulties in connecting human GWAS to gene expression are largely due to our evolutionary history, which has led to an architecture of ascertained regulation that mechanically differs from the architecture of trait-associated regulation. By understanding this difference, we can help explain the disconnect between human traits and regulation, and develop strategies for more effectively identifying trait-associated eQTLs.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6065W: Mitochondrial DNA heteroplasmy improves risk prediction for myeloid neoplasms

Authors:

Y. Hong; Johns Hopkins Univ., Baltimore, MD

Abstract:

The most well-known pathogenic risk factor for myeloid neoplasms (MN) is clonal hematopoiesis of indeterminate potential (CHIP). However, MN can develop in CHIP negative individuals, indicating that additional markers of clonal expansion might also be informative. Heteroplasmy, defined as the presence of mitochondrial DNA (mtDNA) mutations in a subset of cellular mtDNA, has been associated with blood cancers. To date, the relationship between heteroplasmy and CHIP, and their association with MN is not known. We explored the

association between somatic mtDNA and nuclear DNA mutations, its impact on MN incidence, and whether its inclusion to the latest CHIP-based MN prediction model improves risk stratification in participants from the UK Biobank (UKB) and Atherosclerosis Risk in Communities (ARIC) studies. We used the MitoHPC pipeline to quantify heteroplasmy from whole genome sequencing data. We defined a heteroplasmy allele frequency of ≥5% as heteroplasmic and identified predicted deleterious mtDNA variants using a constraint-based score, mitochondrial local constraint score sum (MSS). CHIP variants were called from whole exome sequencing data. We evaluated the associations of heteroplasmy, CHIP, and incident MN using Cox proportional hazards models adjusting for age, sex, center, smoking, and a history of cancer diagnosis. Of 441,936 participants from the UKB (N = 434,304) and ARIC (N = 7,632), there were 1,191 and 160 cases of incident MN, respectively. The mean (standard deviation) ages were 56.5 (8.1) and 57.9 (6.0) years. Heteroplasmy was present in 28.3% in both cohorts, and CHIP in 7.7% in UKB and 4.9% in ARIC. We found that MSS was higher in individuals with CHIP, particularly in those with CHIP VAF≥20%, >1 CHIP mutation, and mutations in the spliceosome machinery. In UKB, individuals with both heteroplasmy and CHIP had a higher risk of MN (adjusted hazard ratio [aHR] 12.1; 95% confidence interval [CI] 10.4, 14.2) than participants with either heteroplasmy (aHR 1.7; 1.5, 2.0) or CHIP (aHR 5.3; 4.5, 6.3) alone. Furthermore, even in the absence of CHIP, we found a significant association of predicted pathogenic effect of heteroplasmic variants (MSS) with incident MN, suggesting a role of mtDNA variation in MN pathogenesis. Finally, incorporating heteroplasmy into an existing risk model for MN in individuals with CHIP improved the sensitivity by 13.1% and identified 34.4% more cases in the high-risk group (10-year risk ≥10%). Our findings suggest that heteroplasmy, in addition to being a marker of clonal expansion, may play a role in MN development, with clinical utility in the general population.
Funding: NHLBI, NCI, National Program of Cancer Registries

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6066W: Molecular Characterisation of Filipino patients with Tyrosinemia Type 1 detected via Expanded Newborn Screening

Authors:

E. Maceda[1], B. Cavan[1], C. Abarquez[2], L. Hamoy[1], M. Alcausin[3]; [1]Univ. of the Philippines Manila, Manila, Philippines, [2]NEWBORN SCREENING CTR MINDANAO, Davao City, Davao Del Sur, Philippines, [3]Univ. of the Philippines-Manila, Ermita, Philippines

Abstract:

Background/Objective: Tyrosinemia Type I or Hepatorenal tyrosinemia (HT1), an amino acid disorder which may cause hepatic failure, is one of the conditions included in the Philippine expanded newborn screening. Biochemical tests like analyses of plasma amino acids and urine organic acids are primarily used. Knowledge of molecular characteristics through FAH gene testing can also be done to confirm the diagnosis. This study aims to describe the molecular characteristics of patients with HT1 in the Philippines.***Methodology:*** Patients with biochemically confirmed HT1 detected via expanded newborn screening from December 2014 to December 2022 were included. Analysis of *HT1* gene done in the patients were reviewed.***Results:*** Fourteen patients with HT1 underwent molecular confirmation. Twelve were homozygote and 2 were compound heterozygotes. The most common genotype was homozygous c.122T>C (p.Leu41Pro) mutation in 9 out of 14 patients (64.29%). In addition, both compound heterozygotes have the c.122T>C p.Leu41Pro mutation in of the alleles. There were 5 mutations noted in this cohort. The most common FAH mutations were c.122T>C (p.Leu41Pro) and c.535_536del (p.Gln179Aspfs*4) at 71.43% and 10.71, respectively.***Conclusion:*** This study provides new insight in the HT1 variant spectrum for a better understanding of this condition in the Filipino population. This study may be useful for molecular diagnosis of HT1 in the Philippines and is useful data for future diagnostic availability in the country including premarital screening, prenatal diagnosis and preimplantation genetic diagnosis.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6067W: Multi-ancestry frontal cortex transcriptomic imputation models show variable predictive performance in single ancestry populations

Authors:

H. Young[1], C. Seah[2], R. Signer[1], M. J. Girgenti[3], L. M. Huckins[3]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, New YorK, NY, [3]Yale Univ., New Haven, CT

Abstract:

Genetic and transcriptomic data from different ancestral populations are often analyzed separately in transcriptome-wide association studies to prevent confounding due to issues such as population stratification. However, grouping participants into discrete groups based on genetic ancestry involves arbitrary categorizations based on genetic similarity, often leading to the exclusion of underrepresented groups and potentially reinforcing harmful narratives about human genetic differences. In this study, we compare approaches to creating transcriptomic

imputation (TI) models and ask whether it is better to create population-matched TI models, or whether we should instead maximize sample size by including all samples regardless of genetic ancestry. Using >1800 post-mortem prefrontal cortex samples, we show that transcriptomic imputation models generated using the full multi-ancestry sample perform better or equal to models created using a subset genetically similar to the EUR 1000 Genomes superpopulation (N = 1532) (Spearman rank correlation rho=0.720, rho=0.716 respectively). Additionally, we create the first set of models for genetically regulated gene expression in the frontal cortex derived from individuals genetically similar to the AFR 1000 Genomes superpopulation. These AFR predictor models improved predictive performance in individuals of AFR genetic ancestry, despite a substantially smaller sample size (N=202) (Spearman rank correlation rho=0.257, versus rho=0.186 with EUR and full models respectively). Multi-ancestry models also outperform single population models when validated in a group of samples with low genetic similarity to both the EUR and AFR 1000 Genomes populations (Spearman rank correlation rho= 0.44, rho= 0.40 with full and EUR models respectively). Our results imply that multi-ancestry models may be constructed and applied to EUR samples without loss of predictive accuracy. For AFR samples however, smaller, more precisely matched predictor models yield the highest predictive accuracy; in this instance, precision outweighs sample size. Given the currently available data for the frontal cortex, these results highlight the utility of more generalizable set of predictor models that better capture cross-population genetic regulatory architecture. However, as we find a subset of genes that are only able to be predicted in a single ancestry, there may be a benefit to using single-ancestry models in parallel, depending on the goals of a study.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6068W: Multi-ancestry GWAS and TWAS implicate cell adhesion in anxiety related phenotypes

Authors:

M. Sharma; Loyola Univ. Chicago, Chicago, IL

Abstract:

Anxiety is the most prevalent form of mental illness in the US, its onset and impact vary for every affected individual. Heritability estimates of anxiety range from 0.3-0.6. We aimed to identify potentially causal genetic variation underlying anxiety phenotypes in diverse ancestral populations through integration of genomic and transcriptomic data from the brain. Approximately 80% of genome-wide association studies (GWAS) published use individuals of

primarily European descent. Genetic associations may not replicate across populations due to variation in linkage block patterns and minor allele frequencies because of differing demographic histories. We performed fine-mapping analyses and transcriptome-wide association studies (TWAS) using summary statistics from the Pan-UK Biobank (Pan-UKBB). Our phenotype of interest is the Worrier/Anxious Feelings phenotype (241,833 cases/259,645 controls), it is comprised of individuals of European (231,995 cases/177,677 controls), Central/South Asian (4462 cases/3649 controls), African (2740 cases/3493 controls), East Asian (1214 cases/1261 controls), Admixed American (508 cases/421 controls), and Middle Eastern (914 cases/560 controls) ancestries. Fine-mapping with LocusZoom identified SNPs in 7 loci (P< 5e-8) in the meta-GWAS, one of those being rs62250713, an intronic SNP, found in the *CADM2* gene (beta=0.05, P=5.8e-20). We performed TWAS using Summary-PrediXcan and transcriptome prediction models from the Genotype-Tissue Expression Project V6. We identified two significant gene associations (FDR<0.05) in the African population, namely *LINC02188* (beta=3.8, P=3.0e-06) in spinal cervical tissues and *CTAGE5* (beta=-0.6, P=4.4e-06) in cerebellar hemisphere tissues (FDR<0.05). Among the top 10 African TWAS associations was the *SMAGP* gene. *SMAGP* is a paralog of *CADM2*, these genes demonstrate similar structures and may function similarly in anxiety. Predicted *SMAGP* in substantia nigra tissues had an effect size of -0.65 (P=3.2e-05), meaning that a decreased expression of *SMAGP* is potentially correlated with an increased risk of anxiety, which is supported by knockdown studies in mice. These genes are promising targets for future studies of anxiety related phenotypes and replication studies are ongoing.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6069W: Multi-omics integration deciphers genetic influences of osteoporosis and the role of *AZIN1* in bone homeostasis ⭐

Authors:

J. Tao[1,2], X. Yi[1], Y. Qian[1], M. Miao[1], X. Li[1,2], H-F. Zheng[1,2]; [1]Diseases & Population (DaP) Geninfo Lab, Sch. of Life Sci., Westlake Univ., Hangzhou, China, [2]Sch. of Life Sci., Zhejiang Univ., Hangzhou, China

Abstract:

Osteoporosis, a common disease diagnosed by bone mineral density (BMD), has elusive genetic determinants. Here, we conducted GWAS of BMD measured by dual-energy X-ray absorptiometry at 11 skeletal sites in 41,006 individuals, identifying 18 novel loci, of which 5 annotated genes have not been previously reported in osteoporosis. Among these, the lead SNP

rs2247355, located in the promoter region of *AZIN1* and modified by H3K27Ac, showed a significant association with head BMD, and its correlation with *AZIN1* was demonstrated by eQTL analysis. Gene-based and SMR analyses supported a positive causal link between genetically determined *AZIN1* expression and head BMD.
By integrating head BMD GWAS and scRNA-seq data from bone and brain, we identified MSCs, osteoblasts, monocytes, macrophages, astrocytes, etc., but not osteoclasts, as cell types through which the risk variants impact head BMD. Bulk RNA-seq analysis of BMSCs from young individuals and senile osteoporosis patients indicated that *AZIN1* levels and RNA editing (a post-transcriptional modification) frequency were decreased in aged participants. Utilizing pseudotime analysis of scRNA-seq data from bone, we observed a temporal pattern of *Azin1* expression during osteoblast differentiation, with its levels rising in pre-osteoblasts, peaking in osteoblasts, and declining in terminal osteocytes.
Importantly, *Azin1* knockdown significantly inhibited osteoblast differentiation and mineralization in MC3T3-E1 pre-osteoblast cells while promoting osteoclastogenesis in RAW264.7 cells. µCT data showed decreased bone volume in mice with conditional knockout of *Azin1* in mesenchymal progenitors (Prx1^{cre}) and osteoblast lineage cells (OCN^{cre}). Moreover, compared to overexpression of WT form of AZIN1, overexpression of RNA-edited AZIN1 enhanced osteogenic differentiation and mineralization of MC3T3-E1 cells while suppressing cell proliferation, whereas non-edited AZIN1 had opposite effects.
Mechanistically, CoIP-MS data revealed that RNA-edited AZIN1 exhibited an increased affinity for interacting with osteoblast differentiation-related proteins, including transcriptional factors such as DDX5, suggesting that RNA-edited AZIN1 enters nucleus to promote osteogenic differentiation. Additionally, *AZIN1* knockdown in MC3T3-E1 or knockout in H9 hESCs dramatically reduced PIEZO1 levels, which crosstalks with osteoclasts by regulating the YAP/Col-II/IX axis, thereby promoting osteoclast differentiation.
Collectively, our findings unveil the candidate determinants of BMD and highlight the role of *AZIN1* in regulating bone homeostasis, offering opportunities for drug development.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6070W: Multi-omics integration in Caribbean Hispanics identifies neuronal and cerebrovascular pathways in Alzheimer's disease

Authors:

J. Eissman[1], M. Qiao[1,2], V. Kalia[3], M. Zerlin-Esteves[2], D. Reyes-Dumeyer[1,2], A. Piriz[1,2], S. Dubey[3], R. Nandakumar[3], A. J. Lee[1,2], R. A. Lantigua[4], M. Medrano[5], D. Mejia[6], L. S. Honig[1,2,4], C. L. Dalgard[7,8], G. W. Miller[3], R. Mayeux[1,2,4], B. N. Vardarajan[1,2,4]; [1]Taub Inst. for Res. on Alzheimer's

Disease and the Aging Brain, Columbia Univ., New York, NY, [2]G.H. Sergievsky Ctr., Columbia Univ., New York, NY, [3]Mailman Sch. of Publ. Hlth., Columbia Univ., New York, NY, [4]Vagelos Coll. of Physicians and Surgeons, Columbia Univ., New York, NY, [5]Pontificia Univ. Católica Madre y Maestra, Santiago, Dominican Republic, [6]CEDIMAT, Plaza de la Salud, Santo Domingo, Dominican Republic, [7]Physiology & Genetics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [8]The American Genome Ctr., Ctr. for Military Precision Hlth., Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Abstract:

Multi-omics integration can clarify molecular mechanisms contributing to Alzheimer's disease (AD). We performed a quantitative trait locus (QTL) analysis to identify genetic variants that regulate metabolomics, epigenetics, and gene expression in AD in a cohort of Caribbean Hispanic individuals from the Dominican Republic and New York, with AD or a family history of AD (N=648 with whole genome sequencing [WGS] data and untargeted metabolomics). Liquid chromatography coupled to high-resolution mass spectrometry was used to measure metabolites (N=7,431). Blood collected in PAXgene Blood RNA Tubes (PreAnalytiX) was used for bulk RNA-sequencing, DNA-methylation (NEBNext Enzymatic Methyl-seq) and whole genome sequencing. We normalized, aligned, and filtered WGS data, retaining common variants with >5% minor allele frequency, which resulted in 7,588,678 autosomal variants and 270,531 X-chromosome variants. QTL analyses were performed with the MatrixEQTL package in R, covarying for age, sex, and principal components for population substructure. We adjusted for multiple comparisons with a false-discovery rate (FDR) threshold of $p<0.05$, applying the autosomal FDR threshold to the X-chromosome results. We identified variants in 440 autosomal and 130 X-chromosome genes to be statistically significant (FDR<0.05). Two QTLs in *ANO2* and *LEMD1* co-localized with variants nominally associated ($p<10^{-3}$) with AD at >95% probability from the Bellenguez et al. 2022 AD GWAS. The set of genes passing the FDR threshold were then leveraged in pathway analyses with the clusterProfiler package in R. Significant pathway associations included synaptic regulation and signaling, extracellular matrix pathways, metabolic processes, and cardio/cerebrovascular pathways implicated in AD. In addition, 27 unique metabolomics QTLs (in 27 genes) overlapped with significant trans-eQTLs (autosomal variants, FDR<0.05). Of those overlapping genes, *PLEK* and *CSMD1* are nominated AD gene targets (AGORA). Overall, we identified variants that regulate metabolite levels in a cohort of Caribbean Hispanic individuals, some of which colocalized with known AD variants or genes, and many of which were enriched in biological pathways that are important in AD. We are conducting QTL analysis with epigenetics data to identify variants shared across multiple omics layers that are implicated in AD. Additional future directions include investigating QTLs that are sex-specific in AD.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6071W: Optimizing a primary bronchial epithelial cell culture model to map air pollution exposure response eQTL

Authors:

S. Lester, S. Abdelwahab, S. H. Randell, S. N. P. Kelada; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Exposure to the ambient air pollutant ozone (O_3) triggers airway inflammation and is linked to the development and exacerbation of asthma and COPD. Molecular markers of O_3-induced airway inflammation display notable inter-individual variation in severity and not all exposed go on to develop lung disease; however, the genomic loci contributing to this variation are largely unknown. Mapping O_3 response eQTL will facilitate the identification of these loci and groups at higher risk of air pollution-induced disease but requires an in vitro airway model that recapitulates the in vivo transcriptomic response. Primary human bronchial epithelial cells (HBECs) are a suitable model for this purpose as they can be differentiated into a pseudostratified epithelium with ciliated and secretory cell types and held at an air-liquid interface (ALI) for exposure to airborne pollutants. Multiple protocols exist for culturing HBECs at ALI but how these protocols affect HBEC response to O_3 is not known. We cultured HBECs from 6 healthy donors in two of the most-used protocols involving different media systems, the University of North Carolina's (UNC) system and STEMCELL Technologies' PneumaCult (PC) system, then exposed cultures from each donor to either filtered air (control) or 0.5 parts per million (ppm) O_3 in a paired experimental design. Cultures grown from each protocol differed in their relative proportions of ciliated and secretory cells, with those in UNC having an estimated 24.6% ciliated cells and 2.94% goblet cells, compared to 40.0% ciliated cells ($p = 0.02$) and 6.76% goblet cells ($p = 0.06$) in PC, which better mimics in vivo proportions. RNAseq was used to quantify gene expression as a function of exposure and donor. UNC-grown cultures had 121 DEGs after O_3 exposure ($|\log2FC| \geq 0.58$, adj. $p < 0.05$) while PC-grown cultures had 10 DEGs. Known O_3 response genes *CXCL8* (log2FC = 1.87, adj. p = 2.33e-8) and *HMOX1* (log2FC = 1.99, adj. p = 1.68e-16) were significantly upregulated in UNC-grown cultures but not in PC-grown cultures. With UNC-grown cultures capturing known hallmarks of the transcriptomic response to O_3 and PC-grown cultures displaying more appropriate cell type proportions, we tested a hybrid media system using UNC BEGM and a 4:1 mix of UNC ALI:PC ALI to achieve cultures with both more appropriate cell type proportions (34.9% ciliated cells, 6.96% goblet cells) and pro-inflammatory response to O_3 by qPCR (*CXCL8*: log2FC = 0.57, p = 0.06; *HMOX1*: log2FC = 0.49,

p=0.02). In summary, media choice strongly influences HBEC cell-type proportions and response to O_3, leading us to utilize a hybrid media protocol capturing the best aspects of two systems to enable O_3 response eQTL discovery.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6072W: OSBPL11 is an important regulator of cholesterogenic genes in murine hepatocyte cell line

Authors:

A. Oteng[1], L. Lei[1], R. Mogire[2], A. Doumatey[3], E. Heuston[4], C. Rotimi[5], A. Adeyemo[4]; [1]NHGRI, NIH, Bethesda, MD, [2]NHGRI, Bethesda, Maryland, MD, [3]NIH/NHGRI, Bethesda, MD, [4]NIH, Bethesda, MD, [5]NIH, Bethesda, MD

Abstract:

Introduction: Polymorphisms in Oxysterol-binding protein 11 (*OSBPL11*) have been shown to be associated with cardiometabolic disease risk factors and outcomes (including blood pressure, LDL-cholesterol and hyperglycemia) in humans. OSBPL11 is a member of the oxysterol-binding protein (OSBP) family, and an intracellular lipid receptor localized in the endoplasmic membrane. However, the mechanisms by which it affects these cardiometabolic phenotypes are unknown and its functional role in cellular lipid metabolism remains undefined.**Methods:** Using an *in vitro* approach, we targeted OSBPL11 via siRNA-mediated knockdown in the murine hepatocyte cell line, Hepa1-6. We performed RNAseq analysis, and challenged the cultured cells with different lipids, including cholesterol and specific fatty acids.**Results:** siRNA-mediated targeting resulted in 75% knockdown in mRNA and 60% knockdown in protein levels of OSBPL11. RNA-seq analysis revealed reduction in pathways involved in cholesterol biosynthesis. Notably, we observed a down-expression of Sterol regulatory element binding protein 2 (*Srebp2*), the master transcriptional regulator of de-novo synthesis of intracellular cholesterol, and SREBP cleavage-activating protein (*Scap*). RT-qPCR corroborated the down regulation of *Srebp2*, along with reduced expression of several genes involved in the cholesterol biosynthesis pathway such as HMG-CoA reductase (*Hmgcr*), phosphomevalonate kinase (*Pmvk*), mevalonate diphosphate decarboxylase (*Mvd*). Exogenous cholesterol could not further repress cholesterogenic gene expression in Hepa1-6 cells with a knockdown of OSBPL11.**Conclusion:** At the cellular level, OSBPL11 appears important in regulating the expression of genes that regulate de-novo cholesterogenesis. Further studies are underway to decipher the mechanisms by which OSBPL11 may be involved in regulating intracellular lipid metabolism, and the implications for pathophysiology of cardiometabolic diseases.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6073W: P selectin: A possible biomarker for vaso-occlusive crisis in Sickle Cell Disease.

Authors:

R. Kumar, P. Gupta; ICMR-Natl. Inst. of Res. in Tribal Hlth., Jabalpur, India

Abstract:

Introduction: Sickle cell disease (SCD) is a monogenic disorder characterized by recurring vaso-occlusive crises (VOC) leading to severe pain and organ damage. One of the mechanisms behind VOC involves the adhesion of sickled red blood cells to endothelial cells and white blood cells, facilitated by increased levels of cell adhesion molecules (CAMs) such as P-selectin. Further, the SNPs present in SELP gene may be associated with p-selectin levels. The aim of this study was to identify the possible genetic association of SELP gene SNPs with P selectin levels during crisis and at steady state. **Methods**: A total of 90 SCD patients admitted in a Government Medical College, Jabalpur during September 2022 to February 2024 were recruited in the present study. Serum levels of P-selectin were measured at three different time points (at the time of admission (90), at discharge (81) and during steady state (48) of each patient). Three exons (exon 5, 11 and 13) of SELP gene were sequenced to find out common SNPs. The statistical analysis was performed to see the changes in P selectin levels at different time point. Possible association with SNPs of SELP gene with P selectin levels was also analyzed. **Results**: A total 90 SCD patients were recruited in the present study. The mean age of the patients was 18.2 ± 8.1 years. The serum levels P-selectin were significantly higher during crisis (median =54.38, IQR=38.31 ng/mL) than at the discharge (median=31.29, IQR=29.22 ng/mL) and steady state levels (median=19.83, IQR=11.92 ng/mL) (p value<0.0001). a total of eight SNPs of SELP gene were observed in studied patients. Significant association of NM_003005.4:c.2102-52C>G and NM_003005.4:c.21052T>G with p selectin levels at crisis state only. No significant association of any SNP was observed with p selectin levels at steady state levels. **Conclusions**: P Selectin can be used as a biomarker for predicting VOC in SCD.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6074W: Population-scale non-coding variant effect predictions using an in-silico reporter assay

Authors:

J. Butts[1], S. Rong[2], S. J. Gosai[3], R. I. Castro[4], M. Noon[2], R. Ghosh[2], R. Tewhey[4], S. K. Reilly[2]; [1]The Univ. of Maine, Orono, ME, [2]Yale Univ., New Haven, CT, [3]The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]The Jackson Lab., Bar Harbor, ME

Abstract:

Over 500 million single nucleotide variants (SNVs) have been identified across diverse human populations. The majority of these SNVs are in non-coding cis-regulatory elements (CREs). CREs dictate the precise spatiotemporal expression of genes and variation within CREs has been linked to evolution, traits, and disease. A powerful approach for characterizing non-coding variants is the use of Massively Parallel Reporter Assays (MPRA), which test the regulatory potential of thousands of sequences in a single experiment. However, it is unfeasible to test all human genetic variation by MPRA, both known or yet to be discovered. Thus, there is a need for a sensitive, scalable, and broadly applicable non-coding variant effect predictor with the experimental accuracy of MPRA.

We recently developed a deep learning model that predicts MPRA activity with accuracy approaching experimental replication in three cell types (K562, HepG2, SKNSH, r = .89 All). We applied this model to variant effect prediction, referring to differences in predicted activity between alleles as 'skew' and variants with substantial skew as expression-modulating variants or 'emVars.' We find our model accurately predicts emVar skew (r .71-.77) and that predicted emVars correlate with DNase Allele Specific Effects (K562 r = .7) as well as recover causal variants for complex traits in UKBB/BBJ and eQTLs in GTEx comparably to experimentally defined emVars (UKBB precision .79-.91, recall .25-.02, GTEx precision .78-1, recall .25-.02). Given our performance on direct and orthogonal comparisons we applied our *in-silico* reporter to experimentally intractable repositories, generating variant effect predictions for 575M non-coding SNVs in gnomAD, ClinVar, and COSMIC.

In ClinVar, we call 8,930 emVars, identify enrichments of emVars in pathogenic variants (OR = 2.97, p = 1.87e-226), and accurately discern pathogenic and benign promoter variants (precision .85, recall .42). We nominate 40,460 emVars in COSMIC and identify emVar enrichments in cancer-relevant non-coding features (Recurrence OR = 1.07 p = .015, Cancer Promoter OR = 1.5, p = .011). In gnomAD, we identify 7M emVars across over 500M common, rare, and private SNVs and observe strong evolutionary constraint on loss-of-function (negative skew) variants in promoters (high skew bin, mean phyloP promoter: 2.8, vs. proximal CRE .8 vs. distal CRE .07). These observations demonstrate that our model extends MPRA to the speed and scale required to understand the clinical and evolutionary impacts of human variation genome-wide.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6075W: Premarital screening and genomic counseling in the UAE enabled by the Emirati Genome Program

Authors:

P. Kandavel[1], A. Al Awadhi[1], A. El-Khani[1], A. Attia[1], F. Al Marzooqi[1], F. Aldhuhoori[1], H. Wu[1], H. Al Mabrazi[1], H. Sajad[1], I. Christy[1], J. Quilez[1], J. Mafofo[1], K. Thangarajan[1], O. Soliman[1], S. Zhang[1], S. Pejathaya[1], T. Magalhaes[1], **V. Zvereff**[1], V. Kusuma[1], V. Kumar[1], Y. Jahangir[1], M. Ibrahim[2], I. Eltantawy[2], W. M. Abdulrahman[3], H. H. Mohamed[3], S. M. Al Hosani[3], A. I. Al Mannaei[3], M. S. Al Ameri[3]; [1]M42, Abu Dhabi, United Arab Emirates, [2]IROS, Abu Dhabi, United Arab Emirates, [3]Dept. of Hlth., Abu Dhabi, United Arab Emirates

Abstract:

Premarital screening aims to prevent the transmission of infectious diseases and genetic disorders to offspring. Many countries have premarital screening and genetic counseling (PMSGC) programs to educate couples about potential infections or harmful genetic variants. These programs are more common in populations at higher genetic disease risk due to their history and mating practices. Most PMSGC programs focus on a limited number of genetic disorders.

Here we present an extended PMSGC in the UAE to screen for a comprehensive panel of 840 genetic disorders using whole-genome sequencing (WGS). Such genomics datasets derive from the Emirati Genome Program (EGP), which aims to perform ~30X WGS of ~1 million Emirati nationals. First, we describe the analysis pipeline to support the PMSGC program in the UAE. The pipeline uses genome-wide call sets of small variants, i.e. single nucleotide variants (SNV) and small insertions / deletions (INDEL), called from the ~30X WGS of each individual of the couple. We focus on variants of concern (VoC) within 570 genes associated with over 840 autosomal recessive genetic disorders - defined as Pathogenic or Likely Pathogenic (P/LP) genetic variants in ClinVar. We then combine the genotypes of the VoC in the two genomes and apply pre-defined rules to report cases of high disease risk predicted in the offspring. We next estimated the expected rate of couples carrying VoC in the Emirati population. We used WGS data from ~150,000 EGP participants as well as age and sex information to predict 436 couples with at least one child in common. By applying the analysis pipeline described above, we found that ~9% of those couples carried VoC. This result is like the rate (8.5%) observed by running the same automated pipeline on 142 couples enrolled in the PMSGC so far. Of note, the predicted 9% is significantly higher (empirical p-value $lt 0.05) than the expected in a random mating scenario within the Emirati population - we derived this by pairing each individual of the 436

couples with a random sex-matched individual from the 150,000 and re-calculating the VoC positive rate (we conducted 1,000 permutations to calculate the empirical p-value).
This pilot study demonstrates the feasibility and potential benefits WGS and genome sequencing programs for PMSGC. A key advantage of WGS is the future ability to cover an increasing number of genetic disorders and updated gene panels as scientific knowledge advances. We believe this work paves the way for further research towards a more holistic approach to reproductive health in the UAE.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6076W: Prequalification of genomic newborn screening for severe childhood genetic diseases through federated training based on purifying hyperselection

Authors:

S. Kingsmore, BeginNGS Consortium; Rady Children s Inst. for Genomic Med., San Diego, CA

Abstract:

Genome-based newborn screening (gNBS) has the potential to improve outcomes in hundreds of severe, childhood, single locus genetic disorders (SCGD) with early onset and effective treatments by early identification and therapy. However, a major impediment of gNBS is imprecision due to variants classified as pathogenic (P) that are not SCGD causal. Purifying hyperselection acts to reduce the frequency of SCGD causal diplotypes more than non-causal diplotypes in large cohorts. This phenomenon improved precision of gNBS in federated training with large, diverse cohorts (610,948 exomes from the UK Biobank (UKB470K) and Mexico City Prospective Study, and 7,342 children and parents who received rapid diagnostic genome sequencing (RDGS)). Before training, gNBS with 53,855 P and likely P variants, 342 genes, 412 SCGD, and 1,603 efficacious therapeutic interventions was positive in 73% of UKB470K adults, suggesting 97% false positives. Training of gene-disease-diplotype triads informed by estimated prevalence, penetrance, expressivity, locus heterogeneity, diplotype heterogeneity, and inheritance modes identified 370 variants contributing diplotype frequencies inconsistent with purifying hyperselection. Manual review disclosed little or no evidence of SCGS causality for 280 of these variants. Upon their removal, 1.9% of UKB470K adults were positive. In contrast, trained gNBS was positive in 7.2% of 3,118 critically ill children with suspected SCGD whilst retaining 98.4% recall when compared to RDGS for the same SCGD. Counterfactual analysis of 8 children with positive results by RDGS and gNBS suggested the potential for the latter to decrease time-to-diagnosis by a median of 121 days, avoid life-threatening disease

presentations in 4 children, organ damage in 6, and to save ~$1.25 million in healthcare cost. gNBS by genome sequencing of NBS dried blood spots was positive in 7.9% of 705 unselected infant deaths. Counterfactual analysis suggested that gNBS could have avoided ten (1.4%) infant deaths. Training in large cohorts based on purifying hyperselection provides a general framework for prequalification of gene-disease-diplotype triads for population screening. Federated training provides a privacy-preserving mechanism for analyses across clinical trials which will be critical to qualify gNBS across diverse genetic ancestries.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6077W: Protein-protein interaction explains *trans* regulation of the proteome and complex trait associations ⭐

Authors:

J. Li[1], Y. Li[2], X. Liu[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Chicago, CHICAGO, IL

Abstract:

The number of GWAS variants continues to multiply as sample sizes increase, but the mechanisms underlying most GWAS loci still remain unclear. Recent large-scale proteome studies reported extensive protein QTLs (pQTLs) in plasma (Gudjonsson et al., 2022; Sun et al., 2023), which provide an unprecedented opportunity to investigate the regulatory mechanisms of proteome and their roles in linking genetic variation to complex traits. In this study, we investigate the properties of *trans*-pQTLs and use *trans*-pQTLs to interpret GWAS loci. Comparing *trans*-pQTLs and *cis*-pQTLs, we found that genes with *trans*-pQTLs but no *cis*-pQTLs are under stronger selection constraint than those with *cis*-pQTLs ($p < 0.001$). *Trans*-pQTLs also colocalize at a higher rate with GWAS loci than *cis*-pQTLs (61% vs 18%, $p < 2.2e-16$), indicating an important role of *trans*-regulatory effects in connecting genetic variants to complex traits. Compared with *trans*-eQTLs, *trans*-pQTLs tend to have larger effect sizes, and the proportion of heritability explained by *trans*-SNPs is higher in protein than gene expression (0.90 vs 0.60, $p = 1.4e-37$). Interestingly, while the majority of *cis*-pQTLs were *cis*-eQTLs, only a small proportion of *trans*-pQTLs appeared to be *trans*-eQTLs, suggesting protein-specific trans regulatory mechanisms. Indeed, *trans*-pQTLs are highly enriched in protein-protein interactions (PPI) (5.4 folds, $p < 0.001$) but only mildly enriched in transcription factor activity (1.1 folds, $p < 0.001$). Motivated by the enrichment of *trans*-pQTLs in the PPI mechanism, we applied *trans*-PCO, a multivariate method developed recently by our group (Wang et al., 2024), to identify *trans*-pQTLs of 1,088 PPI clusters using the plasmatic proteome data from UKB-PPP (Sun et al., 2023). We identified 26,028 *trans*-pQTLs for 1,061 PPI clusters, including 6,234 novel *trans*-pQTLs.

The *trans*-pQTLs of PPIs colocalized with on average 66% (27% - 88%) of GWAS loci per trait for 50 complex traits. These colocalized *trans*-pQTLs directly link GWAS loci to PPIs, thus providing biological insights into complex traits. For example, we identified a colocalization between a novel *trans*-pQTL of a PPI cluster (ITGAM, CD177, ITGB2) and a GWAS locus of systemic lupus erythematosus (PP4 = 0.91). The PPI is known to play an essential role in neutrophil activation. Interestingly, the GWAS locus is near ITGAM, which encodes a protein subunit of the PPI and is thus likely the causal gene of the locus. In conclusion, our study supports a prevalent pQTL-specific trans-regulatory mechanism dependent on PPI. Leveraging known PPI, we were able to map novel *trans*-effects of trait-associated loci, and infer likely mechanisms of action.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6078W: SARs-CoV-2 Vaccination Antibody Kinetics: Genetic basis for differences in humoral immune response half-life in the UK-basedCOVIG study cohort

Authors:

B. Alamad[1], E. Karp-Tatham[1], A. Y. Chong[1,2], J. Wei[3,4], COVIG Study Group, A. Walker[4,5], A. J. Mentzer[1,3], J. C. Knight[1]; [1]Ctr. for Human Genetics, Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [2]Ludwig Inst. for Cancer Res., Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [3]Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom, [4]Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [5]NIHR Hlth.Protection Res. Unit in Hlth.care Associated Infections and Antimicrobial Resistance, Univ. of Oxford, Oxford, United Kingdom

Abstract:

During the COVID-19 pandemic, vaccination efforts were crucial in controlling the spread of SARS-CoV-2. While literature reported that factors including old age, immunosuppression, and obesity contributed to a reduced response to vaccination and increased risk of breakthrough infection, the genetic mechanisms underlying these variations, particularly those influencing the waning of vaccine efficacy over time, remain poorly understood. Antibody waning is defined as the decrease in antibody levels over time following natural infection and vaccination. We examined anti-spike IgG waning kinetics in the COvid-19 Vaccine Immune Genetics (COVIG) project, a UK-based case-control study that aims to elucidate the genetic and immunological mechanisms underlying extreme vaccine responses, relating to maximal or minimal antibody responses. We analysed antibody decay kinetics in 3,457 COVIG participants with human genotyping data available (504 samples still to be processed), specifically individual-level

estimates of antibody half-life derived from linear mixed models for anti-spike IgG antibody measurements over time following second vaccination (censoring at third vaccination or infection). We identified significant non-HLA genetic drivers of vaccine waning through genome-wide association (GWAS) mapping in participants receiving two doses of either the ChAdOx1 nCoV-19 (Chr3) or Pfizer-BioNTech (Chr11) vaccines. Fine mapping of these significant associations is currently ongoing, along with validation in a meta-analysis that includes replication datasets. These findings may provide valuable insights into the factors influencing the waning of vaccine efficacy and inform future vaccine design and deployment strategies.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6079W: Sex differences in the genetic architecture of circulating metabolites and disease risks

Authors:

Y. Chen[1], I. Stewart[2], T. Lu[3], S. Yoshiji[4], K. Liang[5], G. Butler-Laporte[5], M. Koprulu[6], J. Carrasco-Zanini[7], T. SASAKO[1], C-Y. Su[5], O. Bezerra[3], J. Luan[8], P. Raina[9], M. Pietzner[10], C. Greenwood[11], S. Zhou[12], N. Wareham[8], C. Langenberg[7], B. Richards[13]; [1]Lady Davis Inst., Montreal, QC, Canada, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Univ. of Toronto, Toronto, ON, Canada, [4]McGill Univ./Broad Inst., Montreal, QC, Canada, [5]McGill Univ., Montreal, QC, Canada, [6]MRC Epidemiology Unit, Univ. of Cambridge, Cambridge, United Kingdom, [7]Queen Mary Univ. of London, London, United Kingdom, [8]Univ Cambridge, Cambridge, United Kingdom, [9]McMaster Univ., Hamilton, ON, Canada, [10]Berlin Inst. of Hlth.at Charité—Univ.smedizin Berlin, Berlin, Germany, [11]Jewish Gen. Hosp., Montreal, QC, Canada, [12]McGill Univ., Montréal, QC, Canada, [13]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada

Abstract:

Sex differences in disease prevalence are well-documented phenomena influenced by a confluence of factors including genetic variation, hormonal, behavioral, and social factors. Whether differential genetic effects on metabolite profiles exist between the sexes, and whether these translate into differences in disease risks, remains largely unexplored. Here, we investigated the sex-specific genetic architecture of the human plasma metabolome, analyzing 1,040 metabolites in 9,318 males and 10,156 females of European ancestry from two large cohorts, the Canadian Longitudinal Study of Aging cohort and the EPIC-Norfolk cohort. Our analysis identified 1,412 significant variant-metabolite associations ($P < 4.8x10-11$) in males and 1,425 in females. Of these associations, 79 (3%) exhibited sex-differential effects, which potentially originate from sex-differential expression trait quantitative loci (eQTLs) and protein

QTLs. Notably, we also found that some of these sex-differential metabolite QTLs (mQTLs) were implicated in sex-specific disease risks by colocalization analyses. For instance, a locus influencing a glutamine degradant appears to impact cholelithiasis exclusively in males. Using 79 diseases with different prevalences in males and females, we evaluated sex-specific causal effects of metabolites using Mendelian randomization. We found that genetically predicted levels of 29 metabolites affected 23 diseases in males and/or females. This includes four metabolites with consistent effects on six diseases in both sexes; interestingly, the levels of these metabolites differed between sexes, in a manner consistent with differences in disease risk by sex. Lastly, through mediation and interaction analysis using individual-level data, we identified 30 variables that mediate the effect of sex on metabolite levels (e.g., metabolite polygenic risk scores) and 17 variables whose effects on metabolites appear to be modified by sex (e.g., alcohol consumption). Together these variables influenced over 300 metabolites in the two cohorts. Overall, this study revealed that most mQTLs showed similar effects in males and females. Our findings also offer insights into the roles of metabolites in the differences in the prevalence of diseases between sexes, thereby highlighting factors that may have utility as biomarkers or targets of intervention for one or both sexes.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6080W: Single-cell analyses with host genetics reveal the cell state-dependent genetic regulation of transcriptional profiles and T and B cell receptors repertoire in immune cells from 234 Japanese including COVID-19 patients and healthy subjects ⭐

Authors:

R. Edahiro[1,2], G. Sato[1,2], Y. Shirai[1], T. Naito[1], K. Sonehara[3], Y. Tomofuji[1], K. Yamamoto[1], S. Namba[3], N. Sasa[1], G. Nagao[4], H. Namkoong[4], K. Fukunaga[4], T. Morita[1], Y. Kato[1], H. Hirata[1], Y. Takeda[1], D. Okuzaki[1], A. Kumanogoh[1], Y. Okada[1,3,2]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [3]The Univ. of Tokyo, Tokyo, Japan, [4]Keio Univ. Graduate Sch. of Med., Tokyo, Japan

Abstract:

Expression quantitative trait locus (eQTL) analyses have answered functional annotation of the variants and filled the path from large-scale human disease genetics (i.e., genome-wide association studies; GWAS) to clinical phenotypes. However, current eQTL catalogues are mostly at bulk resolution and centered on European ancestry, which rationales needs to construct

single-cell eQTL resources in non-European ancestries. Here, we constructed a multi-omics single-cell atlas of > 1.5 million peripheral blood mononuclear cells with whole-genome sequencing data from 234 Japanese, including COVID-19 patients and healthy subjects. This atlas is characterized by the large number of cells per sample and the inclusion of cells under *in-vivo* immunological stimulation by SARS-CoV-2 infection, resulting in covering a variety of cell states. We mapped the germline genetic effects on gene expression within 7 major cell types (Level 1) and 28 fine cell types (Level 2), and identified ~34,000 cis-eQTLs (in Level 2 cells). We further elucidated dynamic genetic regulation of gene expression across cell states by testing an interaction between genotype and cell state, and identified that such eQTLs (i.e., dynamic eQTLs) were more enriched in enhancer regions than cis-eQTLs, implying the distinct epigenetic properties between them. By utilizing single-cell VDJ-seq with genotype data, we explored the HLA and genome-wide associations with T and B cell receptors (TCR/BCR) repertoire. Particularly, TRAV gene usage was associated with HLA class I and II genes exclusively in $CD8^+$ T cells and naive $CD4^+$ T cells and showed significant interactions between COVID-19 status and HLA class I gene variants in $CD8^+$ T cells. These analyses demonstrated the cell type and context-specific HLA regulation of TCR/BCR repertoire. We assessed GWAS signal colocalizations of 13 complex immune traits in East Asian population and our cis-eQTLs, providing evidence that cell types which exhibited colocalization were specific to the traits. In addition, by using dynamic-eQTLs, we showed a value of dynamic genetic regulation of gene expression across continuous cell states to interpret the GWAS signals (e.g., the *PLD4* locus in SLE). Differential gene expression analysis with COVID-19 polygenic risk scores revealed that the polygenic risks affected transcriptional profiles in a cell type and context-specific manner. In conclusion, we demonstrated the cell state-dependent genetic regulation of transcriptional profiles and TCR/BCR repertoire in immune cells, highlighting the value of multi-omics analyses anchored by single-cell data to understand human complex traits at fine resolution.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6081W: Single-Cell Genomic Dissection of Human Genetic Control of Alternative Splicing in Response to Various Influenza Strains

Authors:

L. Wang, G. Connelly, B. Schott, D. Ko; Duke Univ., Durham, NC

Abstract:

Humans demonstrate substantial differences in response to influenza A infection, with symptoms varying from no symptoms at all to death. One contributor to this broad range of

infection outcomes is genetic differences among individuals and their interactions with the environment. However, the exact extent and molecular mechanisms causing diverse responses to influenza are not yet fully understood. In this study, we conducted single-cell RNA-sequencing on pooled lymphoblastoid cell lines (LCL) from 96 persons, infected by six different strains of influenza A virus (IAV). With a customized computational pipeline, we assigned each single cell back to individual LCL based on genomic variation and their overlapping with mRNA reads. Comparing IAV-infected to baseline conditions, we found >20% of transcripts had altered expression after infection. Splicing quantitative trait locus (sQTL) analysis revealed a total of >180,000 sQTL-exon junction pairs. We found ~90% of sQTLs, detected with at least one IAV strain, were also detected under baseline conditions, defining a set of constitutive sQTLs (csQTLs). The percentage of IAV-induced responsive sQTLs (rsQTLs) varied across different IAV strains with about 25% of rsQTLs associated with more than one IAV strain. More than half of sQTLs replicated in GTEx (v8), and the enrichment of csQTLs was significantly higher than rsQTLs. Functional enrichment against genes that carry rsQTLs (sGenes) highlighted the role of pathways related to mRNA splicing, metabolism of mRNA, and influenza viral transcription. We prioritized sQTLs by 1) overlapping them with variants annotated to overlap splice sites and 2) replication in an independent human cohort of IAV challenge volunteers. This yielded a short list of candidates, including sQTLs for MX1, IFNAR2, and PARP2. Notably, we highlighted an sQTL for PARP2 which localizes to a splice donor site, resulting in an isoform with 13 additional amino acids that is associated with increased levels of IAV compared to the shorter protein. This study uncovers a catalog of influenza-induced sQTLs and sGenes and provides insights into viral-host splicing mechanisms in infectious disease, advancing our understanding of the genetic architecture of flu susceptibility and facilitating future diagnostics and treatments.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6082W: Single-cell RNA-seq analysis uncovers transcriptional signatures in pancreatic islet-like cells derived from iPSCs isogenic for the *ABCC8* R1420H loss-of-function variant

Authors:

K. Cheranda, A. K. Nair, M. Traurig, C. Bogardus, L. J. Baier; NIDDK/NIH, Phoenix, AZ

Abstract:

We previously reported an R1420H substitution in SUR1 (encoded by *ABCC8*) causing hyperinsulinemic hypoglycemia in infancy in a homozygous carrier and doubled the risk of type 2 diabetes (T2D) in 246 heterozygous carriers. Modeling this mutation using stem-cell-derived

pancreatic islets (SC-islets) showed that the 1420H allele induced insulin hypersecretion in immature SC-islets (mimics fetal islets) during basal conditions, whereas mature SC-islets (mimics adult islets) with the 1420H allele have a lower insulin secretory response to glucose. This study aims to identify transcriptomic changes in mature SC-islets responsible for the lower insulin secretory response. Single-cell sequencing was used to generate transcriptional profiles for mature SC-islets from 4 independent differentiations using iPSCs edited with CRISPR/Cas9 to create 9 isogenic cell lines carrying either SUR1 1420RR, 1420RH, or 1420HH. Sequencing data for 35,593 cells (10,683 SC-β cells and 10,937 SC-α cells) was analyzed. Unsupervised clustering and differential expression (DE) analysis identified *GCG*$^+$ SC-α cells (30.7%), *INS*$^+$ SC-β cells (30.0%), *TPH1*$^+$ enterochromaffin-like cells (15.7%), *KRT19*$^+$ duct-like cells (4.4%), *SST*$^+$ SC-δ cells (1.3%), *DPP4*$^+$/*ALDH1A1*$^+$ cells (14.5%), and *MKI67*$^+$ proliferating cells (3.2%). Cell type proportions did not differ by genotype. DE analysis by genotype identified dysregulation in <1% of all expressed genes in 1420RH and 1420HH SC-α and SC-β cells. Comparison of fold change differences in 1420RH and 1420HH SC-α and SC-β cells relative to the 1420RR showed most genes to be directionally consistent. KEGG pathway enrichment analysis identified insulin secretion for 1420HH SC-β cells and glycolysis/gluconeogenesis for 1420RH SC-β cells. Glycolytic genes, *GAPDH*, *GPI*, *TPI1*, and *GCK*, were downregulated, while gluconeogenic genes *PDK3* and *G6PC2* were upregulated. These DE results were consistent between the 4 independent differentiations. Lower glycolytic flux due to downregulation of glycolytic genes or upregulation of *G6PC2*, which converts glucose-6-phosphate back to glucose, could lead to lower ATP production in response to glucose and thereby result in lower insulin secretion. Other key observations include dysregulation of ER stress response genes, immediate early response genes, and genes involved in calcium binding/signaling. We also observed significant upregulation of *GCG* expression in both 1420RH and 1420HH SC-β cells, a hallmark of T2D and indicative of either dysfunctional or dedifferentiated β cells. Altogether, our results provide insight into transcriptomic changes in SC-α and SC-β cells due to early hyperinsulinemia.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6083W: Single-nuclei profiling of mRNA alternative polyadenylation in Alzheimer's disease

Authors:

Y. Choi, P. Dong, G. Hoffman, J. Bendl, D. Lee, K. Girdhar, P. N. M., J. Fullard, P. Roussos; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Alternative polyadenylation (APA) generates isoforms of RNA polymerase II transcripts with distinct 3' termini, including those of messenger RNAs (mRNA). APA is a crucial step of gene regulation that not only diversifies the coding sequences (CDS) of multiple mRNAs but also determines the 3' untranslated region (UTR) of most mRNAs. APA affects various downstream gene regulatory layers, including the gain or loss of microRNA binding sites, differential subcellular localizations, and changes in protein sequences. Previous studies have reported that APA is specifically regulated by cell types or disease conditions, suggesting the necessity of single-cell APA analysis in human disease. However, a single-cell view of APA in the human brain and neurodegenerative diseases remains to be determined due to challenges in analyzing isoform-level diversity across cell types in large cohorts. Here, we provide a systematic APA analysis on single-nuclei RNA-seq with the PsychAD cohort, comprising 1,494 unique postmortem brains from individuals with Alzheimer's disease (AD) and controls. We defined cell-type and age-specific polyadenylation sites (PASs) and compared their usage across different cell types, comprising eight classes common in human brains (excitatory neurons, inhibitory neurons, oligodendrocytes, oligodendrocyte progenitor cells, astrocytes, microglia, and vascular cells) and their 28 subclasses. To identify differential APA events from the large cohort and adjust for confounding factors, PAS counts were log-ratio transformed and tested at the pseudobulk level using precision-weighted linear mixed models, with the recently developed pipelines crumblr and dreamlet. We determined the differential APA usage in AD and aging and elucidated their relationship with previously known mRNA 3' end processing mechanisms. Our results suggested that AD related APA can derive the gain or loss of mRNA cis-regulatory elements, potentially targeted by microRNAs or RNA-binding proteins. Lastly, we identified PAS quantitative trait loci (QTLs) and found a substantial fraction of the PAS QTLs colocalized with AD genetic risk variants. Together, our findings highlight the diversity and importance of APA in AD.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6084W: Tissue-specific protein quantitative trait loci: implications for drug discovery in cardiometabolic disease

Authors:

A. Hartley[1], K. Sukhavasi[2,3], S. Hu[1], M. Traylor[1], H. Talukdar[4], E. Bjorkegren[3], A. Ruusalepp[2], J. L. M. Bjorkegren[3], J. M. M. Howson[1], Y. Jamshidi[1]; [1]Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom, [2]Dept. of Cardiac Surgery and The Heart Clinic, Tartu Univ. Hosp. & Dept. of Cardiology, Inst. of Clinical Med., Tartu Univ., Tartu, Estonia, [3]Clinical Gene Networks AB, Stockholm, Sweden, [4]Clinical Gene Networks AB, Stockhom, Sweden

Abstract:

Integrating genetics with plasma proteomics in biobank studies has revealed valuable disease-related insights. The levels of plasma protein levels are largely influenced by that of tissues either through secretion or leakage into the circulation, however the concordance of genetic regulation of protein levels in major tissues and plasma is largely unknown.
We used Olink Explore 3072 for antibody-based proteomic profiling across cardiometabolic disease-relevant tissues (liver [LIV], skeletal muscle [SKM], aorta [AOR], mammary artery [MAM]) and plasma in 260 participants (76.5% male; 76.9% with coronary artery disease (CAD)) from the Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET) study. Genotypes were imputed to the TopMed reference panel, and cis protein quantitative trait loci (pQTLs) identified using REGENIE, adjusting for standard covariates. The cis region was defined as +/-500kb from the transcription start and end sites. We considered $p < 5x10^{-8}$ significant.
We identified pQTLs for 49 proteins in LIV, 34 in SKM, 40 in AOR, 13 in MAM and 175 in plasma. Nine proteins had genome-wide significant pQTLs across all tissues. Proteins with pQTLs in plasma were enriched for immune-related pathways. We observed varying degrees of overlap in genetic regulation (gProteins, i.e. proteins with at least one pQTL) across tissues, with the highest overlap between LIV and plasma and the lowest with MAM and other tissues. Systematic colocalization using HyPrColoc found shared genetic regulation across tissues for 86 proteins (PP>0.7), mainly (72) between plasma and at least one other tissue. 16 proteins had shared genetic regulation across all 5 tissues. Fourteen additional colocalizations were between at least two tissues excluding plasma, including CEACAM21 which had distinct pQTLs in plasma compared to tissue (r2=0.1). We also assessed tissue-specific effects on CAD. We found that nine proteins, including SORT1 in the liver only (PP=0.98), colocalized with CAD; increasing SORT1 gene and protein expression in the liver is suggested to reduce CAD based on Mendelian randomization.
Our findings underscore the significance of studying tissue-specific pQTLs to enhance our mechanistic understanding of protein function in disease, paving the way for more precise drug and biomarker discovery efforts.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6085W: TLR4 gene Polymorphism and biochemical markers as a tool to identify risk of Osteoporosis in women from Karachi

Authors:

R. Baig[1], R. Rehman, Professor[2], R. Ahmed, AP[1]; [1]Univ. of Karachi, Karachi, Pakistan, [2]Aga Khan Univ. Pakistan, Karachi, Pakistan

Abstract:

Background: Osteoporosis, characterized by low bone mineral density, poses a global health concern. Diagnosis increases the likelihood of developing osteoporosis, a multifactorial disorder marked by low bone mass, elevating the risk of fractures in the lumbar spine, femoral neck, hip, vertebras, and distal forearm, particularly in postmenopausal women due to bone loss influenced by various pathophysiological factors.
Objectives: To investigate the association of serum cytokine, bone turnover marker, bone mineral density and TLR4 gene polymorphism in pre and post-menopausal women and to find if any of these can be the potential predictor of osteoporosis in postmenopausal women. **Material and methods:** Study participants were consisting of Group A (n=91) healthy pre-menopausal women and Group B (n= 102) healthy postmenopausal women having ≥ 5 years' history of menopause. ELISA was performed for cytokine (TNFα) and bone turnover marker (carboxytelopeptides), respectively. Bone Mineral Density (BMD)was measured through dual X-ray absorptiometry (DEXA) scan. Toll-like Receptors 4 (TLR4) gene polymorphisms (A896G; Asp299Gly) and (C1196T; Thr399Ile) were investigated by PCR and Sanger sequencing.**Results:** Statistical analysis reveals positive correlation of age and BMI with T scores in premenopausal group whereas in post-menopausal group found a significant negative correlation between age and T-score at hip (r = - 0.352**), spine (r = - .306**), and femoral neck (r = - 0.344**) and a significant negative correlation of BMI with TNF-α (- 0.316**).No association and significant differences were observed for TLR4 genotype and allele frequencies among studied groups However, both SNPs exhibited significant association with each other. **Conclusions:** This study concludes that BMI, BMD and TNF-α are the potential predictors of osteoporosis in post-menopausal women. However, CTX and TLR4 gene polymorphism did not appear as potential predictors of bone loss in this study and apparently cannot help in predicting bone loss in post-menopausal women.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6086W: Transcriptome-wide characterization of genetic perturbation atlases across cell types and dosage levels

Authors:

A. Nadig[1], J. M. Replogle[2], A. N. Pogson[2], S. A. McCarroll[3], J. S. Weissman[2], E. Robinson[4], L. J. O'Connor[3]; [1]Broad Inst., Allston, MA, [2]Whitehead Inst. for BioMed. Res., Cambridge, MA, [3]Harvard Med. Sch., Boston, MA, [4]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Genetic perturbation effects may vary across cell types and gene dosage; such variability provides insight into genome regulation and complicates prospects for clinical translation. Recent methodological advances have enabled the creation of massive genetic perturbation atlases. However, the data produced by these screens are noisy, and traditional statistical approaches may miss the majority of differential expression effects. This statistical challenge parallels the problem of "missing heritability" in GWAS, which was addressed by quantifying the SNP-heritability due to all measured variation, significant and otherwise.
We developed Transcriptome-wide Analysis of Differential Expression (TRADE), a statistical method that explicitly models the distribution of true differential expression effects across genes and experiments. TRADE fits a flexible mixture model to observed differential expression point estimates and standard errors, avoiding the use of arbitrary significance thresholds. TRADE can be used to analyze perturbations in one setting (the "transcriptome-wide impact," loosely parallel to the SNP-heritability) as well as across settings (estimating the "transcriptome-wide impact correlation", loosely parallel to the genetic correlation). We analyzed data from two new essential gene perturbation atlases (N = 2393 genetic perturbations in HepG2 and Jurkat cell lines), in addition to four published datasets.
Across all datasets, our analyses revealed that significant genes capture only 36% of the transcriptome-wide impact. Comparing essential gene perturbations in K562, RPE1, Jurkat, and HepG2 cell lines, we find that transcriptome-wide impact is moderately correlated across cell types (correlation ranging from 0.4 to 0.75), but larger in RPE1, suggesting that this cell line is more generally sensitive to perturbations. Across cell lines, perturbation effects were more correlated between K562/Jurkat and RPE1/HepG2 than between these pairs, possibly reflecting similarity in developmental origin and p53 status. A small number of perturbations (241) showed evidence of strongly cell-type-specific effects. In datasets titrating essential gene dosage, perturbations at different dosage levels varied not only in magnitude of effects, but also in what genes are affected, with only moderate transcriptomic effect correlations (correlation = 0.60) between strong and weak perturbations of the same gene.
Our findings highlight the insight that can be drawn from principled integration of perturbation results. More broadly, our approach lays an analytic foundation for the systematic characterization of perturbation atlases.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6087W: Transgenic denAsCas12a-CRISPRa/i mouse models for combinatorial in vivo perturbations

Authors:

N. Nerurkar[1,2], K. Mouri[1], R. Tewhey[1,2,3]; [1]The Jackson Lab., Bar Harbor, ME, [2]Graduate Sch. of BioMed. Sci., Tufts Univ. Sch. of Med., Boston, MA, [3]Graduate Sch. of BioMed. Sci. and Engineering, Univ. of Maine, Orono, ME

Abstract:

Over the last decade, genome-wide association studies (GWAS) have been instrumental in discovering over ten thousand variants associated with complex traits and risk for disease in the human population. It is estimated that ~90% of susceptible variants reside within non-coding regions of the genome, including cis-regulatory elements (CREs) such as promoters and distal regulatory elements, suggesting that the disruption of gene regulatory function plays a key role in complex disease progression and risk. Most non-coding trait-associated variants have small effect sizes on phenotypic output, complicating our ability to study and leverage variant catalogs. To overcome this obstacle, multiplex modulation of CREs in which variants reside can lead to a more significant (and measurable) phenotypic effect. Cas12a, a type V CRISPR endonuclease, has intrinsic RNase activity that cleaves precursor crRNA itself without host RNA polymerase III or a tracrRNA scaffold. Cas12a's self-cleaving activity can be uniquely leveraged to process many crRNAs in parallel from a single construct to target multiple CREs in a single cell. Given the polygenic nature of complex disease, assessing the compounded effect of disease risk variants in a complete physiological system is critical. Combining DNase-inactive Cas12a (dCas12a) with an epigenetic repressor or activator allows for multiplex modulation of targeted CREs across multiple loci without editing target DNA. Currently, mouse models have been developed to express dCas12a transiently, resulting in the potential for variable expression levels that ultimately lead to the uncertain measurement of small effect phenotypes. To this end, we have engineered transgenic denAsCas12a activator (VP64) and repressor (KRAB) mouse models generated by inserting a ubiquitous Cre recombinase-inducible construct into the *Rosa26* safe harbor locus on chromosome 6 of C57BL/6J mice. These new models will allow for stable and constitutive gene modulation across multiple tissues and loci, providing greater elucidation of the underlying genetic mechanisms and regulatory networks of heritable disease risk.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6088W: UBE2A mutations linked to X-linked intellectual disability type Nascimento produce variable effects on protein stability and stress resistance

Authors:

R. Field, G. Barros, G. Silva; Duke Univ., Durham, NC

Abstract:

Mutations in the ubiquitin-conjugating enzyme UBE2A are linked to X-linked intellectual disability type Nascimento, which manifests as moderate to severe intellectual disability, speech impairment, craniofacial dysmorphisms, urogenital abnormalities, and additional symptoms. Although a range of mutations with diverse clinical presentations have been identified, very little is known about how the mutations affect protein functionality and thus lead to disease. Therefore, we developed a yeast model system to systematically compare disease-associated UBE2A variants and examine their effects on gene expression, protein stability, protein function, and impact in cellular physiology. Our findings demonstrate that mutations such as Q128X reduce UBE2A protein levels, whereas the Q93E mutation increases protein abundance under the same promoter. Analysis of gene expression and protein half-life suggest that different mechanisms are responsible for defining this aberrant abundance of individual UBE2A mutants. We next showed that this altered protein abundance disrupts selective UBE2A-mediated pathways. Among these pathways, our group has shown that the yeast homolog of UBE2A, Rad6, mediates global repression of translation under oxidative stress by conjugating K63-linked polyubiquitin to ribosomes. By measuring translation rates and using immunoblotting assays, we determined that UBE2A complements Rad6 translation repression function in yeast, and mutants like Q128X and Q93E disrupt this activity. Surprisingly, while mutants like Q128X exhibit decreased accumulation of K63-linked polyubiquitin, which is aligned with a loss of function phenotype, the Q93E variant shows increased accumulation. Furthermore, consistent with the idea of impaired translation control, UBE2A mutant variants show substantial sensitivity to hydrogen peroxide, similar to null *RAD6*-deleted strains. Together, our results reveal that selective patient-derived UBE2A mutations might lead to Nascimento Syndrome through distinct molecular and functional mechanisms. Our work further illustrates that dissecting the impact of each mutation on UBE2A expression and function will be critical to develop selective therapies and better understand the varied clinical manifestation observed in patients.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6089W: Uncovering Genetic Factors Influencing Gut Microbiome Composition in the Korean Population: A Genome-Wide Association Study

Authors:

S. Lee[1], Y. Jung[1], S. Lee[1], Y. Chang[2], S. Ryu[3], H-L. Kim[4], H. KIM[1]; [1]Samsung Advanced Inst. for Hlth.Sci. & Technology ,Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Kangbuk Samsung Hosp., Seoul, Korea, Republic of, [3]Kangbuk Samsung Hosp., Sungkyunkwan Univ., Seoul, Korea, Republic of, [4]Ewha Womans Univ, Seoul, Korea, Republic of

Abstract:

The gut microbiome plays a key role in human health, influencing various physiological processes and disease status. While recent genome-wide association studies (GWAS) have provided insights into the genetic determinants of gut microbiome composition, most studies have focused on European populations and utilized 16S rRNA sequencing, which has limitations in taxonomic resolution. This study aims to address these gaps by conducting a comprehensive GWAS to identify host genetic factors that influence gut microbiome diversity and composition in a Korean cohort. A GWAS was conducted on 1,179 Korean individuals, utilizing whole-genome shotgun sequencing to profile the gut microbiota with high taxonomic resolution. SNP genotyping data were imputed using the 1000 genomes reference database. We used PLINK to perform a GWAS of host genetic variations for 174 bacterial species. Several loci were identified with significant associations to 64 bacterial species, including *Prevotella copri*, with p-values ranging from 2.87×10^{-10} to 5×10^{-8}. Additionally, we found an association between the *KALRN* gene locus and *Phocaeicola plebeius* species. The *KALRN* gene encodes for the Kalirin protein, which is crucial in neuronal development and synaptic plasticity, and its variants have been linked to neuropsychiatric disorders. Certain genetic variants identified in previous studies of European populations were found to be relevant in the Korean population, although there were some population-specific differences. This GWAS using whole-genome shotgun sequencing represents a significant advancement in elucidating the genetic determinants of gut microbiome composition in the Korean population. These findings provide valuable insights for personalized medicine and underscore the necessity of including diverse populations in microbiome research. Our results can also be used to explore causal relationships between genetic variants and microbiome-associated diseases.
This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2023R1A2C2006416).

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6090W: Using bulk RNA-sequencing analysis to study the effects of rs174557 and rs968567 for gene expression

Authors:

S. Choudhary[1], S. Yang[2], Y. Sun[3], H. Zhang[4], K. Ye[2]; [1]Inst. of Bioinformatics, Univ. of Georgia, Athens, GA, [2]Univ. of Georgia, Athens, GA, [3]Univ. of Goergia, Athens, GA, [4]Inst. of Bioinformatics, Athens, GA

Abstract:

Abstract Background: Long-chain polyunsaturated fatty acids (LC-PUFAs) are essential for human health, with the *FADS1* and *FADS2* genes encoding the two key enzymes responsible for their biosynthesis. Numerous studies have associated genetic variants in *FADS* cluster with LC-PUFA levels and various diseases. However, the underlying mechanism are still largely unknown.**Methods:** In this study, we analyzed the transcriptome profile from bulk mRNA sequencing of four different haplotypes for rs174557-rs968567 in HEK293T and HepG2 cells, generated by prime editing. We compared the expression levels of our target genes, *FADS1* and *FADS2,* as well as some transcription factors that have been predicted to interact with rs174557 and rs968567. In addition, we performed differential gene expression analysis and GO/KEGG pathway analysis to examine overall effects of the rs174557 and rs968567 and analyzed how they might influence the risks of different diseases. Furthermore, we compared the pattern of the effects of rs174557 and rs968567 in HEK293T and HepG2 cells to investigate on the tissue-specific regulation mechanisms. Lastly, we also compared the transcript-specific expression levels for *FADS1* and *FADS2.***Results:** Bulk RNA sequencing data showed that in HEK293T cells, both rs174557-G and rs968567-T significantly upregulates the expression of both *FADS1* and *FADS2*. In HepG2, rs968567-T enhances both *FADS1* and *FADS2* expressions as well. However, rs174557-G decreases *FADS1* and *FADS2* when rs968567 is T allele while increases *FADS1* and *FADS2* when rs968567 is C allele.**Conclusion:** The transcriptome data of different haplotypes in HEK293T and HepG2 cells revealed that rs968567 is a causal variant for both *FADS1* and *FADS2,* with T allele upregulating *FADS1* and *FADS2.* The effect of rs174557 on *FADS1* and *FADS2* depends on the genotype of rs968567, indicating the interaction of two causal variants. **Keywords:** Bulk RNA Sequencing; Prime editing; *FADS1*; *FADS2*; Causal variants; rs174557; rs968567;

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6091W: Using co-expression networks of transcriptome, translatome, and proteome to dissect the risk of schizophrenia

Authors:

Q. Liang[1,2], Y. Wei[2], C. Liu[2,1], C. Chen[1]; [1]Central South Univ., Changsha, China, [2]SUNY Upstate Med. Univ., Syracuse, NY

Abstract:

Introduction: Genes cannot function independently; they co-work on molecular functions, and co-regulate phenotypes. Multiple omics approaches, including transcriptomics via RNA-seq, translatomics via ribo-seq, and proteomics via mass spectrometry, quantify genes in their corresponding forms at various steps of the information flow in the Central Dogma. Co-expression modules of each omics can capture the organized regulatory relationships among genes at each level and help us uncover their genetic regulation. **Results**: A total of 44, 37 and 22 co-expression modules were detected for transcriptome, translatome and proteome, respectively based on RNA-seq, ribo-seq and mass spectrometry data of 185 prefrontal cortex samples. Through this genome-wide search, 206 mod-QTLs were identified for transcriptome, 459 for translatome and 193 for proteome at P value < $5e^{-8}$. Only 10 mod-QTLs were found by either cis-QTLs or trans-QTLs of module genes, emphasizing the importance of mod-QTLs. Notably, it was observed that the transcriptomic M29 module exhibited the most of genetic regulations (136 M29 mod-QTLs out of 206 all mod-QTLs, 136/206=66%), as did translatomic M2 module (268/459=58%) and proteomic M13 module (142/193=74%). Furthermore, 33 out of 44 transcriptomic modules were not preserved in either translatome or proteome, implying the dynamic changes of regulatory relationships through the Central Dogma. Six transcriptomic modules were preserved in the other two omics. Two of the transcriptomic modules preserved only in translatome and three only in proteome. These two scenarios may involve more complicated mechanisms, such as the combined effect of post-transcriptional and post-translational regulations. Finally, using two-sample mendelian randomization, transcriptomic M29, M30 and M43 modules were found to contribute to schizophrenia risk. Similarly, the translatomic M2, M35 and M36 modules, as well as proteomic M2 and M21 modules, were also identified to contribute to schizophrenia risk. **Conclusion**: Co-expression network analyses indicate that genes are clustered differently in transcription, translation and protein levels. Genetic regulations of several modules were identified. Genetic regulators of distinct modules at different omics contribute to schizophrenia risk.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6092W: Validation and development of functional assays for *RYR1* variants

Authors:

E. Ying[1], J. J. Johnston[1], A. Douglass[1], R. T. Dirksen[2], K. Stowell[3], L. Biesecker[1]; [1]NIH NHGRI, Bethesda, MD, [2]Univ. of Rochester Med. Sch., Rochester, NY, [3]Massey Univ, Palmerston North, New Zealand

Abstract:

The *RYR1* gene encodes a sarcoplasmic reticulum (SR) calcium channel involved in excitation-contraction coupling in skeletal muscle. *RYR1* gain of function variants cause Malignant Hyperthermia Susceptibility (MHS) and are characterized by increased sensitivity to *RYR1* agonists leading to increased calcium release from the SR. Genetic testing for *RYR1*-related MHS relies on accurate *RYR1* variant classification. The ClinGen *RYR1* Variant Curation Expert Panel (VCEP) refined the ACMG/AMP criteria for variant pathogenicity using a Bayesian model. However, variants of uncertain significance (VUS) still comprise a majority of classified *RYR1* variants. We are developing a high-throughput assay to expand the availability of functional data and reduce the large number of VUS. We have developed a truth set of 6 pathogenic and 17 benign *RYR1* variants using a constitutive expression vector, pcDNA5/FRT, expressed in HEK293 Flp-In cells. A fura-2 calcium indicator was used to measure caffeine-induced calcium flux in the HEK293 Flp-In system, showing poor EC_{50} separation between pathogenic and benign variants. Across 12 benign variants, the mean EC_{50} was 0.4862 mM with a 95% confidence interval of 0.3626 to 0.6099. Across 5 pathogenic variants, the mean EC_{50} was 0.2432 mM with a 95% confidence interval of 0.1420 to 0.3445. We re-engineered variants in pcDNA5/FRT/TO, an inducible expression vector expressed in HEK293 Flp-In T-REx cells, and preliminary data suggest better EC_{50} separation. The VCEP determined that the *in vitro* calcium flux assay reached a pathogenic evidence strength of moderate (odds of pathogenicity >4.3), based on published data of 35 pathogenic and 10 benign variants. We will recalculate the evidence strength with data from our inducible clones. To identify targets for a high-throughput functional assay, we analyzed RNA-Seq data from constitutive cell lines, identifying 5 candidate transcripts that were differentially expressed between pathogenic and benign variants. We have used TaqMan assays to test these candidate transcripts in inducible cell lines but were not able to replicate the gene expression changes. We are collecting additional RNA-Seq data using the inducible cell lines with the goal of identifying a single transcript that can be selected in a large-scale assay either through direct detection or reporter gene fusion. Our results show differences in EC_{50} separation in constitutive compared to inducible expression systems, highlighting the need to independently validate *in vitro* models. These results expand understanding of *RYR1* functional assays, refine their implementation, and show potential for an improved approach.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6093W: Variant-aware training of machine learning models for predicting the effects of genetic variants on chromatin accessibility

Authors:

J. Jaureguy[1], D. Laub[2], A. Pekis[2], B. Balderson[1], A. J. Ho[1], S. Huang[3], U. Jain[2], T. Arthur[4], J. Nguyen[5], F. Monroe[1], S. Gabriel[6], K. Brown[1], H. Carter[7], M. D'Antonio[1], K. Frazer[8]; [1]Salk Inst. for Biological Studies, San Diego, CA, [2]UCSD, San Diego, CA, [3]UCLA, Los Angeles, CA, [4]Univ. of California, San Diego, La Jolla, CA, [5]Univ. of California - San Diego, La Jolla, CA, [6]California Polytechnic State Univ., Pomona, CA, [7]UCSD, La Jolla, CA, [8]UC San Diego, San Diego, CA

Abstract:

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with human traits and diseases. However, most of these variants lie in the non-coding genome, and their molecular function and causality are unclear. Molecular quantitative trait locus (QTL) analysis can help determine variant function by identifying variants associated with molecular traits such as chromatin accessibility (caQTLs) and gene expression (eQTLs). However, due to linkage disequilibrium, the identity of causal variants remains ambiguous. Furthermore, QTL studies lack power to detect associations with rare genetic variants. Machine learning (ML) has emerged as an alternative method for determining variant function. ML models predict gene expression, histone modifications, and chromatin accessibility (CA) from DNA sequence alone. However, current ML models are trained on the reference (ref) genome and do not consider genetic variation in training. We hypothesized that including genetic variation and allele-specific chromatin accessibility (ASCA) in training would improve ML predictions of variant function. Here, we introduce a "variant aware" (VA) modeling paradigm for predicting the impact of genetic variants on CA. We utilized transfer learning to incorporate genetic variation and allele-specific information into the training of Basset, a convolutional neural network designed to predict CA. For training and evaluation, we selected a human induced pluripotent stem cell dataset containing ATAC-seq data and genotypes for 65 unrelated donors. We evaluated the ability of the Basset VA and ref models to predict ASCA at caQTLs. The VA model outperformed the ref model with a Pearson correlation of r=0.40 vs. r=0.33. Additionally, the VA model outperformed the ref model in predicting CA across individuals at caQTLs (median r=0.21 vs. r=0.15). For some caQTLs, the VA predictions were substantially better than the ref predictions. Furthermore, we will evaluate the ability of VA large language models (LLMs) to predict CA. We pretrained genomic large language models (LLMs), including HyenaDNA and Mamba on the human ref genome and vertebrate genomes from a 447-way mammalian alignment. We trained 1.4M and 7M parameter models at 2kb and 65kb sequence length, and will fine-tune them to predict CA using our VA approach. In summary, VA ML models leverage genetic variation and allele-specific information to more accurately predict the effects

of genetic variants on CA demonstrating their potential for interpreting trait-associated genetic variation.

Session Title: Molecular Effects of Genetic Variation Poster Wednesday Session

Board 6094W: Whole exome sequencing of siblings identifies a novel variant p.A399fs in *ANOS1* gene associated with Kallman syndrome

Authors:

S. Mathur[1], H. Singh[1], A. Singh[1], P. Tiwari[1], N. Mathur[1], V. Scaria[2], S. Sivasubbu[2]; [1]Dept. of Endocrinology, Sawai Man Singh Med. Coll. and Hosp., Jaipur 302004, India, India, [2]CSIR Inst. of Genomics and Integrative Biology, Mathura Road, Delhi 110025, India, India

Abstract:

Abstract: Kallman Syndrome is an uncommon condition characterised by absent pubertal development and abnormalities of smell, secondary to GnRH deficiency and aplasia or hypoplasia of olfactory bulb. Kallman Syndrome is the most common cause of isolated hypogonadotropic hypogonadism, but in majority of cases genetic diagnosis remains elusive. The arrival of Next generation Sequencing (NGS) in clinical landscape has revolutionized genomic research and helped in expansion of genetic repertoire of Kallman Syndrome. The whole exome analysis was performed in two siblings with Kallman syndrome to identify the causative genetic defect. The variants were prioritized using standard open-source computational pipelines. A novel frameshift mutation (c.1195_1196insG) in the exon 8 of ANOS1 (KAL1) gene was identified. The absence of dental agenesis, synkinesis and renal agenesis in one sibling otherwise commonly seen with ANOS1 variants suggest the p.A399fs variant may associated with a milder phenotype. To the best of our knowledge, this is the first report describing Kallman syndrome due to this novel frameshift ANOS1 gene mutations. This report also exemplifies the utility of whole exome sequencing in cases of atypical presentation of rare diseases.

Session Title: Omics Technologies Poster Wednesday Session

Board 1100W: A genome-wide association study of mass spectrometry proteomics using the Seer Proteograph platform

Authors:

S. Batzoglou[1], K. Suhre[2], Q. Chen[3], A. Halama[4], K. Mendez[4], A. Dahlin[4], S. Nisha[4], G. Theraja[4], H. Sarwath[1], H. Guturu[1], V. Dwaraka[5], F. Schmidt[4], J. Lasky-Su[6]; [1]Seer Inc, Redwood City, CA, [2]Weill Cornell Med. - Qatar, Doha, Doha, Qatar, [3]Proteomics Core, Weill Cornell Med.-Qatar, Education City, Qatar, [4]Weill Cornell Med.-Qatar, Education City, Qatar, [5]TruDiagnostic Inc., Lexington, KY, [6]Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA

Abstract:

Genome-wide association studies (GWAS) with proteomics are essential tools for drug discovery. To date, most studies have used affinity proteomics platforms, which have limited discovery to protein panels covered by the available affinity binders. Furthermore, the extent to which protein epitope changing variants interfere with the detection of protein quantitative trait loci (pQTLs) has not been quantified to date. Mass spectrometry-based (MS) proteomics can overcome some of these limitations. Here we report a GWAS using the MS-based Seer Proteograph™ platform with blood samples from a discovery cohort of 1,260 American participants and a replication in 325 individuals from Asia, with diverse ethnic backgrounds. We analyzed 1,980 proteins quantified in at least 80% of the samples, out of 5,753 proteins quantified across the discovery cohort. We identified 252 and replicated 90 pQTLs, where 30 of the replicated pQTLs have not been reported before. We further investigated a combined 200 of the strongest associated cis-pQTLs previously identified using the SOMAscan and the Olink platforms and found that up to one third of the affinity proteomics pQTLs may be affected by epitope effects, while another third were confirmed by MS proteomics to be consistent with the hypothesis that genetic variants induce changes in protein expression. The present study demonstrates the utility of using mass spectrometry for proteomics research and reports potential pQTLs not accessible to affinity proteomics, suggesting that many more pQTLs remain to be discovered using MS-based platforms.

Session Title: Omics Technologies Poster Wednesday Session

Board 1101W: A statistical framework to model and infer gene networks from genome-scale genetic perturbation data

Authors:

J. Choi; Inst. for Hlth.Care Delivery Sci., Dept. of Population Hlth.Sci.and Policy, and Tisch Cancer Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Identifying gene networks responsible for various biological processes is important for elucidating molecular and cellular underpinnings of human health and disease. Recent advances in CRISPR-based genome editing science and technology have enabled systematic investigation of genetic perturbations and their corresponding molecular and cellular phenotypes. This progress motivates the development of analytical methods to map perturbed gene networks underlying phenotypic outcomes at the genomic scale. In this work, I present a statistical framework applicable for modeling and inferring gene networks that regulate the expression of a particular gene of interest using genome-scale genetic perturbation data. The key idea is that the change in the expression level of a particular gene is influenced by a group of directly and indirectly interacting genes subjected to systematic genetic perturbation, such as genome-scale CRISPR interference. The co-regulatory relationship between a gene of interest and other genes can be represented and inferred using a regularized Gaussian graphical model, as the number of interacting genes for a particular gene of interest is expected to be sparse at the genomic scale. Particularly, I investigate the feasibility of a regularized Gaussian graphical model inference method and existing implementation, which allows testing the presence of co-regulatory relationships in a specific graph model based on confidence interval estimates for the inverse covariance matrix elements. To investigate the feasibility of the proposed method, I simulated the expression changes of a gene of interest influenced by other genes on a path graph of size 10, 15, and 20, obtained from the inactivation of individual genes in 100 hypothetically identical cells, each of which consists of a total of 100 genes. The study shows that the power of the proposed method to detect the edges of the ground truth path graphs depends on (i) path graph sizes and (ii) correlation coefficients between the expression levels of two adjacent genes. For example, the method can detect approximately 85-90% of edges of path graphs of sizes 10, 15, and 20 with a false discovery rate controlled at 10%, when the correlation coefficients between adjacent genes' expression levels range from 0.55 to 0.65. When the correlation coefficients range from 0.45 to 0.55, the fraction of edges detected reduces to 60-74% for the path graphs of sizes 10, 15, and 20. I am currently extending the simulation study to examine the feasibility of the proposed method in broader and more realistic settings.

Session Title: Omics Technologies Poster Wednesday Session

Board 1102W: Accelerating interpretation of rapid whole genomes and diagnosis in the intensive care unit

Authors:

S. Levesque[1], J. Cayouette[1], P. Thibault[1], E. Gravel[1], S. Gravel[1], N. Anastasio[1], F. Hamdan[2], J. Gauthier[2], C. Varin-Tremblay[2], J. L. Michaud[2]; [1]CHUS, Sherbrooke, QC, Canada, [2]CHU Sainte-Justine, Montreal, QC, Canada

Abstract:

Background: Analysis of rapid whole genomes for patients in intensive care units requires a fast turnaround time, while interpretation of variants can be time-consuming. Prioritization of variants for review is thus of utmost importance. To enhance rapid interpretation of whole genomes and facilitate quick review by qualified laboratory personnel, we implemented a series of improvements in our published prioritization tool, PhenoVar, which uses patient's genotypic and phenotypic data. We validated its performance in a group of patients with known diagnoses. **Methods**: We adapted the point system of the ACMG v4 and ClinGen SVI recommendations for variant classification, integrated bioinformatics predictions, supported UPD, CNV, and parental genotypes, and enhanced phenotype comparisons to minimize the impact of the non-specific clinical presentations in neonates. All the evidence of variant classification and prioritization can be visualized on the web interface, enabling rapid independent assessment. We tested these improvements on a set of 29 rapid whole genome trios with known diagnoses, and default parameters were adjusted. Performances were then assessed on a second set of 131 patients: 62 rapid whole genomes and 69 exomes from outpatient clinics. SNVs and indels represented 90.6% of all causal variations, while CNVs, SVs, and UPD accounted for 8.1%, 0.6%, and 0.6%, respectively. **Results:** Overall, PhenoVar identified the diagnosis in 93.8% of the 160 cases, with the correct diagnosis found among the top 5 predictions 93.3% of the time. This decreased to 92.4% and 92.6% if we excluded the 29 rapid genomes used to adjust the default parameters. Although not significant, trio/duo (n=105) performed better than singleton (n=55) in terms of diagnoses among the top 5 ranks with 94.9% and 90.4%, respectively. Missed diagnoses were mostly secondary to limited phenotype descriptions of causal disorders in databases, delay in the inclusion of causal genes in the OMIM database, structural variation (unsupported), or poor variant genotype quality. Diagnoses not ranking in the top 5 were most likely to come from singleton analyses because of ambiguous phase or inheritance of variants. Limitations of databases could be overcome by a second-step analysis, which unmasks predicted pathogenic variants in genes considered as incidental findings or of unknown clinical significance, thus increasing identification of diagnoses to 96.9%. **Conclusion:** PhenoVar effectively prioritizes variants in whole genomes and exomes, identifies the correct diagnosis in a high proportion of patients, and enables rapid assessment of pathogenic evidence by laboratory personnel.

Session Title: Omics Technologies Poster Wednesday Session

Board 1103W: An automated variant verification system to improve reporting efficiency

Authors:

S. Katragadda, S. Ghosh, S. NSN, A. Das Mahapatra, S. Aliya Afreen, A. Singh, A. Janakiraman, V. Veeramachaneni; Strand Life Sci. Private Limited, Bengaluru, India

Abstract:

Background
Workflows for NGS-based rare disease diagnosis result in a candidate set of variants in genes strongly correlated with the clinical phenotype. Before the variants can be included in a report, a final check on variants with borderline quality indicators is carried out by experienced bioinformaticians to ensure the variants are not artifacts arising from genome assembly differences, homology, sequencing errors, or pipeline parameters. This variant verification (VV) process can be time consuming since it often involves the use of external tools and databases. It can also be subjective since it depends on experts reviewing the reads in a genome browser. As NGS laboratories move to whole exome sequencing, a larger number of variants are being shortlisted in each case - and VV is increasingly becoming a significant bottleneck in clinical reporting.
Our system aims to significantly reduce the VV time, by computationally determining if the variant in question is an artifact or not.
Methods
Our system consists of three tools which check for specific types of artifacts. When a VV request is made for a specific variant in a sample, each tool assesses if the variant could be a specific type of artifact and returns a verdict (Yes/No/Maybe) with supporting information such as blast alignments. The three tools are - a) an assembly artifact checker which extracts a representative read with the variant allele, aligns it against multiple genome builds, and analyzes the alignments to determine if the variant call is the result of build differences b) a homology artifact checker which uses a database of pre-calculated differences between homologous regions to assess if the variant read originates from a homologous region c) a homopolymer checker which uses an error profile computed from 202 samples to assess if an indel in a homopolymer stretch is likely to be genuine. The system invokes and summarizes information from the three tools. A variant is considered to be a false positive if any of the 3 tools assesses it to be an artifact.
Results
109 VV requests from 55 WES cases were handled by the standard manual process and were also passed through the system. In 31 cases (28%) all checkers indicated it was not an artifact and all these variants were independently confirmed true positives. In 3 cases an artifact was

detected confidently. In 75 cases (68%) with at least one “Maybe” verdict, the variant was confirmed as a false positive in 70 cases. The adoption of this system is expected to result in a 84% reduction in VV time.

Session Title: Omics Technologies Poster Wednesday Session

Board 1104W: An Integrated RNA-Seq Pipeline for Concomitant Human Genetic and Microbial Analyses in Persistent Cutaneous Infections

Authors:

H. Vahidnezhad[1], F. Palizban[2], S. Biglari[3], L. Youssefian[4], A. Saeidian[5], F. Vahidnezhad[6], H. Hakonarson[7], J-L. Casanova[8]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Lausanne, Lausanne, Switzerland, [3]Univ. of Pennsylvanaia, Philadelphia, PA, [4]Univ. of California Los Angeles, Los Angeles, CA, [5]Baylor Coll. of Med. - - Houston, TX, Houston, TX, [6]Univ. of California Santa Cruz, Santa Clara, CA, [7]Children's Hosp. of Philadelphia, Philadelphia, PA, [8]The Rockefeller Univ., New York, NY

Abstract:

Enhancing the management of patients dealing with persistent cutaneous infections (PCI), especially those affected by inborn errors of immunity (IEI) and monogenic skin barrier defects, can be vastly improved through genetic diagnosis. Previously, we introduced the RNA-Seq method as a primary diagnostic approach for single-gene disorders, enabling comprehensive screening for genomic mutations without reliance on DNA analysis. In this study, we expand the utility of RNA-Seq by developing HumMicroScan, a computational pipeline for exploring both host omics and microbial species from cutaneous biopsies of PCI patients. This innovative approach enables the simultaneous identification of genomic variants linked to human diseases and the characterization of microbial species. In this regard, skin biopsies from both unaffected and affected areas of PCI patients underwent RNA-Seq analysis. Reads aligning with the human genome facilitated unbiased detection of rare human genomic variants alongside somatic mutation signature analysis. Unaligned reads were exploited to identify viral, bacterial, and fungal species. HumMicroScan pipeline was tested on a series of real samples and successfully detected human variants and associated microbial species, consistent with validated experimental and clinical evidence: (1) a *CORO1A* mutation alongside Mycobacterium tuberculosis, epidermodysplasia verruciformis (EV)-related HPV5, and Staphylococcus aureus; (2) a *TP63* mutation with Candida auris; (3) an *ITK* mutation with HPV3; and (4) an *RTEL1* mutation with HPV2 in a patient exhibiting multiple cutaneous horns. Further validation was conducted through gene expression profiling and splicing analysis. Overall,

HumMicroScan pipeline expands the clinical application of RNA-Seq, furnishing clinicians with an additional tool for effectively managing patients with persistent infections. Availability: https://github.com/FPalizban/HumMicroScan

Session Title: Omics Technologies Poster Wednesday Session

Board 1105W: An interactive web application for exploring human plasma and fibroblast metabolomics data from patients with inborn errors of metabolism

Authors:

L. Cai[1], H. Vu[1], W. Gu[1], H. Chen[1], J. Franklin[1], L. A. Haidar[1], Z. Wu[1], C. Pan[1], F. Cai[1], P. Nguyen[1], B. Ko[1], C. Yang[1], L. Zacharias[1], J. Sudderth[1], S. Montgomery[1], C. Uhles[1], H. Fisher[1], J. Hudnall[1], C. Hornbuckle[1], C. Quinn[1], D. Michel[1], L. Umana[1], A. Scheuerle[1], M. Mcnutt[1], G. Gotway[1], B. Afroze[2], M. Ni[1], R. Deberardinis[1]; [1]UT Southwestern Med. Ctr., Dallas, TX, [2]Aga Khan Univ., Karachi, Pakistan

Abstract:

Metabolomic profiling is instrumental in understanding the systemic and cellular impact of inborn errors of metabolism (IEMs), monogenic disorders caused by pathogenic genomic variants in genes involved in metabolism. This study encompasses untargeted metabolomics analysis of plasma from 474 individuals and fibroblasts from 67 subjects, incorporating healthy controls, patients with 65 different monogenic diseases, and numerous undiagnosed cases. We introduce a web application designed for the in-depth exploration of this extensive metabolomics database. The application offers a user-friendly interface for data review, download, and detailed analysis of metabolic deviations linked to IEMs at the level of individual patients or groups of patients with the same diagnosis. It also provides interactive tools for investigating metabolic relationships and offers comparative analyses of plasma and fibroblast profiles. This tool emphasizes the metabolic interplay within and across biological matrices, enriching our understanding of metabolic regulation in health and disease. As a resource, the application provides broad utility in research, offering novel insights into metabolic pathways and their alterations in various disorders.

Session Title: Omics Technologies Poster Wednesday Session

Board 1106W: Assessing the readiness of Oxford Nanopore sequencing for clinical genomics applications

Authors:

J. Arres[1], A. Saad[1], A. El-Khani[1], A. Attia[1], C. Minas[1], D. Sanchez[1], F. Aldhuhoori[1], G. Katagi[1], H. Wu[1], J. Quilez[1], J. Mafofo[1], O. Soliman[1], S. Elavalli[1], S. Behl[1], S. Purohit[1], T. Cardoso[1], **T. Magalhaes**[1], V. Zvereff[1], F. Sedlazeck[2]; [1]M42, Abu Dhabi, United Arab Emirates, [2]Baylor Coll. of Med., Houston, TX

Abstract:

Long-read sequencing (LRS) technologies, e.g. Oxford Nanopore Technologies (ONT), have emerged as promising solutions to overcome the limitations of short-read sequencing (SRS). However, the integration of ONT into clinical practice is challenged by the need for rapidly changing and high-maintenance analysis pipelines and its perceived lower accuracy compared to established high-throughput sequencing (HTS) standards.

Here, we optimized analysis pipeline for whole-genome sequencing (WGS) ONT data. The pipeline employs cutting-edge software designed to detect genetic variations used for clinical diagnostics and is scalable, ensuring readiness for both clinical scenarios requiring a rapid response and extensive population genome projects. To validate its effectiveness, we sequenced 17 well-characterized Coriell reference samples at ~30X genome coverage using ONT (both R9 and R10 chemistries) as well as SRS Illumina and MGI. We evaluated our integrated ONT WGS and analysis pipelines for precision in identifying genetic variations within individual WGS datasets. First, we benchmarked variant calling performance of single nucleotide variant (SNV) and small insertion-deletion (INDEL) in a parent-offspring trio. We found exceptional consistency and robustness in SNV detection across platforms, achieving an F-score >0.975 - ONT (both R9 and R10) performed as well as Illumina (~0.980) and better than MGI (0.976). ONT's weak point continues to be INDEL detection (F-score<0.80), especially compared with SRS (>0.930). We highlight though the notable improvement of INDEL calling in R10 relative to R9 (0.749 and 0.674, respectively). Second, we assessed the accuracy of each platform in detecting known disease-causing mutations in the remaining 14 Coriell samples (9 different diseases linked to SNV, INDEL, CNV and STR mutations). We detected the expected mutation in most cases and consistently across platforms - exceptions were balanced between ONT and SRS (ONT imprecisely detected a CNV while Illumina and MGI underestimated an STR).

Overall, our work supports the increasing readiness of ONT for clinical applications and provides key information to ONT users on analysis pipeline considerations as well as persistent limitations of this long-read technology. To our knowledge, this is the first multi-platform comparison including both ONT chemistries (R9 and R10) and an SRS technology other than Illumina.

Combined with the comprehensive set of reference samples, we believe this multi-platform dataset will be valuable for the genomics community.

Session Title: Omics Technologies Poster Wednesday Session

Board 1107W: Association of sitting time and physical activity with gut microbiome

Authors:

S. Lee[1,2], Y. Jung[1,2], S. Lee[1,2], H-L. Kim[3], Y. Chang[4,5,6], S. Ryu[4,5,6], H-N. Kim[1,2]; [1]Dept. of Clinical Res. Design and Evaluation, Samsung Advanced Inst. for Hlth.Sci. & Technology, Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]BioMed. Statistics Ctr., Res. Inst. for Future Med., Samsung Med. Ctr., Seoul, Korea, Republic of, [3]Dept. of Biochemistry, Coll. of Med., Ewha Womans Univ., Seoul, Korea, Republic of, [4]Dept. of Occupational and Environmental Med., Kangbuk Samsung Hosp., Sungkyunkwan Univ. Sch. of Med., Seoul, Korea, Republic of, [5]Ctr. for Cohort Studies, Total Hlth.care Ctr., Kangbuk Samsung Hosp., Sungkyunkwan Univ. Sch. of Med., Seoul, Korea, Republic of, [6]Dept. of Clinical Res. Design & Evaluation, SAIHST, Sungkyunkwan Univ., Seoul, Korea, Republic of

Abstract:

The risks attributed to sedentary activity (SA) has been reported to be substantially attenuated by physical activity (PA) at high intensities. We hypothesized that the benefits of PA could be due to changes in gut microbiota. This cross-sectional study aimed to reveal the association of SA and PA with gut microbiome in 1328 participants, who underwent a health examination at the Kangbuk Samsung Hospital Healthcare Screening Center in South Korea between 2014-2021. The PA (Vigorous/Moderate to Inactive) and SA (Sedentary/Non sedentary) were assessed using the validated Korean version of the international Physical Activity Questionnaire Short Form. According to the combination of PA and SA, participants were categorized into four groups; Sedentary-Vigorous PA (SED-V), Sedentary-Moderate to Inactive PA (SED-MI), Non sedentary-Vigorous PA (noSED-V), and Non sedentary-Moderate to Inactive PA (noSED-MI). We investigated the gut microbiome, metabolic pathways, and fecal metabolites using shotgun metagenomic sequencing and analyzed them in linear regression model with pairwise comparisons. The sex, age, and BMI were included as covariates. In diversity analysis, no significant α-diversity differences were identified between groups. The significant dissimilarity was identified for the Jaccard β index between SED-MI and noSED-MI groups ($R^2 = 0.002$, PERMANOVA, $p < 0.05$), and between SED-MI and noSED-V groups ($R^2 = 0.003$, $p < 0.05$). Compared to SED-MI group, *Phocaeicola coprophilus* (coef: 1.35, MaAsLin2, $p < 0.05$) was

enriched in noSED-MI group. Compared to SED-MI group, *Megamonas funiformis* (coef: 2.06, $p < 0.05$) and *Parabacteroides faecavium* (coef: 0.75, $p < 0.05$) was enriched, while *Blautia A sp900066355* (coef: -0.73, $p < 0.05$) was depleted in noSED-V group. The CMP pseudaminate biosynthesis pathway (coef: 0.56, $p < 0.05$) and L-glutamate and L-glutamine biosynthesis pathway (coef: 0.18, $p < 0.05$) were enriched in noSED-MI group than SED-MI group. The 6-hydroxymethyl dihydropterin diphosphate biosynthesis pathway (coef: -0.34, $p < 0.05$) was depleted in SED-V group compared to SED-MI group. Stearoyl ethanolamide (predicted metabolite, coef: -0.04, $p = 0.056$) was marginally depleted in noSED-MI group compared to SED-MI group. Bilirubin (predicted metabolite, coef: 0.05, $p = 0.058$) was marginally depleted in SED-V group compared to SED-MI group. Our results suggest that the detrimental effects of sedentary activity could be alleviated by vigorous PA, which could be mediated by the changes in gut microbiome. This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2023R1A2C2006416).

Session Title: Omics Technologies Poster Wednesday Session

Board 1108W: B cell immune repertoire sequencing in smoking, vaping, and chronic obstructive pulmonary disease in the COPDGene cohort

Authors:

M. Moll[1], Z. Xu[2], M. Ryu[1], E. Silverman[1], M. Cho[3], C. Hersh[4], M. Sauler[5], F. Polverino[6], C. Vollmers[7], P. Castaldi[8]; [1]Brigham & Women s Hosp., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Brigham and Women s Hosp., Duxbury, MA, [4]Brigham and Women's Hosp., Canton, MA, [5]Yale, New Haven, CT, [6]Baylor Coll. of Med., Houston, TX, [7]UC Santa Cruz, Sanata Cruz, CA, [8]Brigham and Women s Hosp., Cambridge, MA

Abstract:

Rationale: Cigarette smoking (CS) adversely affects B-cell function and antibody production, heightening susceptibility to infections. The effects of e-cigarette use (‘vaping’) and combined CS and vaping use (‘dual use’) on B-cell activity are unclear.
Objective: To examine alterations in B-cell receptor sequencing (BCR-seq) profiles associated with smoking, vaping, dual use, COPD-related outcomes, and demographic characteristics including sex and race.
Methods: We performed BCR-seq in blood RNA samples from 240 subjects in the Genetic Epidemiology of COPD (COPDGene) study. We tested the multivariable associations of measures of B-cell function (immunoglobulin heavy chain (IGH) subclass expression and usage, class-switching, somatic hypermutation (SMH), and clonal expansion) with current cigarette smoking,

vaping, or dual use, measures of COPD severity, age, sex, and race. Multiple comparisons were adjusted using the Bejamini-Hochberg method, with significant associations identified at 5% false discovery rate (FDR) and suggestive associations at 10% FDR.
Measurements and Main Results: Among 240 COPDGene non-Hispanic white (NHW) and African American (AA) participants with smoking/vaping and BCR-seq data, we observed significant positive associations between smoking and dual use with the proportion of secretory IgA production with the strongest associations observed for dual use. Dual use was also significantly positively associated with class switching and SHM, indicating an overall pattern of increased B-cell activation that was detectable in dual users but not as evident in smoking or vaping subjects. We did observe a significant negative association between vaping and SMH in IgM/IgD isotypes as well as a suggestive association to secretory IgA production. We also observed significant associations between self-identified race and IgG antibody proportion, with self-identified African American participants having a higher proportion of IgG subclasses and lower IgM production relative to White participants.
Conclusions: We observed synergistic effects of smoking and vaping on B cell activation, which was more clearly detectable in dual users compared to those who smoked or vaped alone. Self-reported race was strongly associated to IgG isotype usage. These results identify significant associations between B-cell activation and antibody transcription that may contribute to differences in immune responses influenced by smoking and vaping behaviors as well as self-identified race.

Session Title: Omics Technologies Poster Wednesday Session

Board 1109W: Best practices Terra workspace for processing PacBio long reads sequenced on the Revio system: Single sample, trio, and cohort modes

Authors:

M. Fleharty; Broad Inst., Cambridge, MA

Abstract:

The latest generation of long-read sequencing technologies bring the promise of more comprehensive variation characterization in every sample. However, most secondary analysis infrastructure has been built to process short reads and cannot handle long read data. We present a best practices pipeline (hosted in a Terra Workspace) optimized for processing long reads human whole-genome sequencing (WGS) from the Pacific Biosciences Revio platform. This workspace makes use of a WDL-based pipeline developed by PacBio and available from

GitHub. Our workspace supports single sample, trio, and cohort modes. Key steps include read preprocessing, error correction, alignment, and variant calling.
An optional tertiary analysis step includes advanced variant filtering by joint calling samples in either trio or cohort mode. Outputs are generated at both the sample level (e.g., BAM statistics, small variant VCFs, ROH outputs, phased variant VCFs, HiFiCNV outputs, CpG methylation data) and the cohort level (e.g., jointly called SV and small variant VCFs, haplotype phasing statistics). Tertiary analysis outputs include filtered and compound heterozygous variants, detailed TSV files, and filtered structural variant data from svpack. The workspace features conditional execution based on available data and parameters, along with customizable options for backend environments and computational resources, ensuring robust performance for diverse genomic studies.
We validated the performance of this pipeline using the Ashkenazim trio sample HG002, HG003, and HG004. These samples were run in both single sample and trio modes, leveraging extensive truth data available for HG002 for SNP and indels. Additionally, we used T2T-ACE, a tool we have created for the purpose of evaluating the positive predictive value (PPV) for copy number variant (CNV) detection.
This pipeline provides a robust, scalable platform to perform complete human LR-WGS for research, translational, and ultimately clinical intended uses.

Session Title: Omics Technologies Poster Wednesday Session

Board 1110W: Brain RNA degradation is highly regulated in Alzheimer's disease

Authors:

J. Sanford, D. Dikec, C. Wang, D. Western, C. Cruchaga; Washington Univ. in St. Louis, Saint Louis, MO

Abstract:

Introduction: Alzheimer's disease (AD) is the most common form of dementia and is characterized by neuronal death and microglia activation. These processes are thought to be highly regulated. Therefore, we hypothesize that RNA stability, including degradation and protection, is regulated and associated with disease. To study RNA stability in human brains, we calculated Transcript Integrity Number (TIN) in RNA-seq datasets. TIN is calculated at the transcript level which allows us to identify transcripts that have reproducible changes in stability in AD. **Methods:** We performed a three-stage design: 1) discovery in post-mortem parietal samples from 344 Knight-ADRC individuals, 2) replication in ROSMAP (n=530; 864 samples) and

MSBB (n=207; 698 samples), 3) meta-analysis. In addition, we analyzed blood RNA-seq from Knight-ADRC individuals (n= 1161). We analyzed three AD phenotypes: Clinical Dementia Rating (CDR), Braak stage, and AD-control status. TIN was calculated in the bulk RNA-seq data with RSeQC. Linear regression and mixed models were used to test the association of transcript TIN with AD. Dependence on age or transcript counts was assessed. Pathway and cell type enrichment analyses were performed on the significant transcripts. **Results:** After excluding age-associated transcripts, we identified 1,736 transcripts with TIN associated with CDR, 539 with Braak, and 291 with status (meta-FDR $p < 0.05$). 502 CDR, 65 Braak, and 42 status associated transcripts were found to be independent of the expression level of the transcript. Enrichment analyses indicate that the transcripts are part of the neurogenesis (adjusted p-value = 0.0119) and extracellular matrix organization (adjusted p-value = 0.0120) pathways and enriched in neuron (CDR p-value = 1.42×10^{-5}, Braak p-value = 0.0197) and astrocyte cells (CDR p-value = 5.09×10^{-5}, Braak p-value = 7.90×10^{-4}). In the blood analysis, 32 CDR and 8 status hits ($p < 0.05$) overlapped with the brain results. The transcripts overlapped with AD-GWAS genes: 109 from the CDR, 25 from the Braak, and 14 from the status analysis. **Conclusion:** We identified transcripts whose integrity was associated with AD phenotypes. These transcripts are part of AD-relevant pathways and overlap with previously identified AD-risk genes and appear to be mostly tissue-specific. The changes in integrity across disease status could help pinpoint dysregulated transcripts, indicating that some transcripts are actively and specifically degraded or protected in the course of disease. This method of studying RNA regulation could lead to new insights into AD pathology, allowing for identification of targets for intervention in AD.

Session Title: Omics Technologies Poster Wednesday Session

Board 1111W: Building large-scale population databases of structural variant calls from low- and high-coverage long-read sequencing data

Authors:

D. Miller[1], J. Lin[1], J. Gustafson[1], E. Eichler[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA

Abstract:

Our understanding of the normal patterns of structural variant (SV) across human populations is limited because prior methods, such as short-read sequencing identified fewer than half of the approximately 25,000 SVs present in any one person. New technologies, such as long-read sequencing (LRS) have been shown to both identify the expected number of SVs per genome, and also resolve their complex structure. Thus, there is broad interest in using LRS to 1) better

understand normal patterns of SVs across populations, 2) evaluate challenging unsolved cases in the research setting, and 3) to serve as a first-line test for Mendelian conditions in the clinical setting. A major barrier to these aims is the absence of large-scale SV control databases that can be used for variant prioritization and filtering.
We have consolidated SV calls from several projects that generate both low- and high-coverage LRS datasets to generate a harmonized call set that can be used to better understand SV structure and allele frequency in population controls. Our current high-confidence call set includes data from nearly 350 individuals from the HGSVC, HPRC, and 1000 Genomes ONT Consortium. These data, which are of both high-coverage (> 30x depth of coverage) and high average read lengths (read N50 > 40 kbp for Nanopore-only data), were used to identify more than 175,000 SVs across all individuals, with approximately 50,000 of these SVs observed in only one individual. By combining only high-coverage, high read N50 Nanopore data for 100 samples with nearly 1,000 low-coverage, low read N50 data recently generated by Ingelheim-Boerhinger we identified approximately 625,000 SVs, but observed nearly 3x as many insertion as deletion events, and a significant reduction in the number of calls per sample. These results suggest challenges with accurate SV identification and merging when lower coverage data is used. These results demonstrate the need for more robust SV calling and integration tools, as well as an evaluation of how data quality (depth of coverage and read lengths) affects SV calling, in order to facilitate uptake of LRS in the clinical setting.
Moving forward, our goal is to generate the equivalent of a gnomAD database of high-quality SV calls using LRS data from more than 50,000 individuals from diverse populations for use by the genomics community to accurately annotate SVs identified using LRS approaches in the research and clinical settings.

Session Title: Omics Technologies Poster Wednesday Session

Board 1112W: Clinical and protein impact of human loss-of function mutations in British Bangladeshi and Pakistani participants of the Genes and Health Study

Authors:

C. Langenberg[1,2], J. Carrasco Zanini[1], d. van heel[3], Genes and Health Investigators and Industry Consortium; [1]Queen Mary Univ. of London, London, United Kingdom, [2]Berlin Inst. of Hlth.at Charité Univ.smedizin, Berlin, Germany, [3]Barts and The London, London, United Kingdom

Abstract:

Loss-of-function (LOF) mutations affecting both copies of a gene in humans ("human knockouts") provide a natural experiment to better understand gene function. Here, we performed plasma proteomic profiling (of 5416 unique proteins) and Whole Exome Sequencing (WES) on a subset of 1071 British Bangladeshi or Pakistani descent adults from Genes & Health, a population enriched for biallelic (homozygous) human knockouts due to high autozygosity. This, together with NHS health record linkage, provides a unique opportunity to assess proteomic and phenotypic effects of specific human knockouts. We identified 241 predicted LOF (pLOF: stop, frameshift, splice) homozygous rare variants in 223 genes in 188 individuals. Target cis-proteins were covered by the Olink Explore HT platform for 71 genes with 78 unique pLOF genotypes in 76 individuals. We observed strong under-expression of 15 target proteins in cis with 18 unique pLOF genotypes (N individuals = 19; median z-score = -13.89 (IQR = -10.67 to -7.92)), and strong over-expression of 1 target protein (N individuals = 1; z-score = 5.06) of the 34 which were highly detectable in plasma (above the limit of detection in > 95% samples). We observed no evidence of a wider effect of pLOF variants on deregulation of global proteomic profiles or proteins associated to the knockout target gene. We also observed 47 missense rare homozygous genotypes (at 40 genes, 45 variants) with strong underexpression of the cis protein. Phenotypic consequences matching OMIM and PanelApp gene annotations were observed in 4 individuals with strong pLOF genotype effects on plasma protein levels. One such example included a human knockout for PLA2G7 (protein z-score = -8.96), who had allergic chronic rhinitis and asthma consistent with this genes' association to atopy susceptibility. As a previously considered drug target, this example points to potential on-target side effects of pharmaceutical inhibition of PLA2G7. We demonstrate the value of proteogenomics in diverse populations to 1) understand phenotypic consequences and drug development opportunities through gene disruption in human knockouts, and 2) characterize the effect of rare variation on plasma protein levels individuals of South Asian ancestry, which are systematically underrepresented in research.

Session Title: Omics Technologies Poster Wednesday Session

Board 1113W: Clinical Validation of a Haplotyped Whole Genome Sequencing Assay Using PacBio HiFi Sequencing

Authors:

D. Corney, **C. Ziemniak**, W. Wei, Y. Qiu, C. Mozdzierz, J. Otto, H. Latif; GENEWIZ from Azenta Life Sci., South Plainfield, NJ

Abstract:

Next-generation sequencing (NGS) has revolutionized genomics and enabled unprecedented insights into the molecular basis of health and disease. Short-read NGS technologies have become integral to the diagnosis of genetic diseases, and in supporting clinical trials. Unfortunately, while current short-read technologies are more cost-effective and scalable than ever before, short read-lengths are limited in resolving complex genomic regions, detecting structural variations, and covering repetitive sequences. Additionally, short-reads may not be able to accurately differentiate between trans and cis compound heterozygosity. This is critical in human germline diploid studies, since whether variants are present on the same or different chromosomes may confer vastly different phenotypes. While such challenges can be mitigated by long-read sequencing approaches, widespread clinical adoption is limited due challenges in sample handling, scalability, and cost.

To address these limitations, we report the CLIA/CAP validation of a clinical long-read WGS sequencing assay for the comprehensive genome-wide detection and phasing of SNVs, INDELs, CNVs, and CpG methylation. High molecular weight DNA was extracted from frozen EDTA whole blood, followed by DNA fragmentation to ~15-20 kb by automated pipette-mediated mechanical shearing and SMRTbell library preparation. After sequencing each sample on a PacBio Revio SMRTcell, ~20-30x average genome-wide coverage was obtained. Variant calling accuracy was assessed by comparing variant calls in healthy donors to those obtained by Illumina PCR-free WGS assay, as well as analysis of Genome In A Bottle (GIAB) well-characterized reference samples. Overall, the variant calling concordance F1 score was 0.99 and 0.98 for SNVs and INDELs, respectively. Phasing accuracy was performed by trio analysis on specific genes of interest.

This validated long-read WGS assay overcomes key limitations of short-read technologies, enabling comprehensive genome-wide detection and phasing of all variant classes. By leveraging automation and high-throughput workflows, this assay provides a powerful tool for precision genomic medicine, with the potential to improve diagnostic accuracy, treatment selection, and patient outcomes.

Session Title: Omics Technologies Poster Wednesday Session

Board 1114W: ClioMD: an artificial intelligence model for ciliopathies

Authors:

M. Ergoren[1], N. Senturk[2], M. S. B. Ali[1], &. O. Ozcelik[1], K. D. Erol[1], S. Temel[3], M. Dundar[4]; [1]Faculty of Med., Near East Univ., Nicosia, Cyprus, [2]Faculty of Engineering, Near East Univ., Nicosia, Cyprus, [3]Bursa Uludag Univ., Bursa, Turkey, [4]Faculty of Med., Erciyes Univ., Kayseri, Turkey

Abstract:

Cilia are highly specialized cellular organelles that serve multiple functions in human development and health. Their central importance in the body is demonstrated by the emergence of various developmental disorders resulting from defects in cilia structure and function caused by several different inherited mutations in more than 150 different genes. Genomic analyzes have rapidly improved our understanding of the intracellular molecular biological basis of ciliopathies over the past two decades, and new technological advances have accelerated this progress. However, most of the time, correlation of phenotypic results with genetic variation and environmental factors, patient phenotypes do not match with the thought disease despite being a basic search in genomic medicine, candidate variants are in genes not characterized by disease, model organisms are insufficient to explain the disease, etc. Many obstacles continue to hinder rapid and accurate diagnosis. Artificial intelligence models using advanced computing tools can phenotypically identify overlapping disease models such as ciliopathies in research and diagnostic contexts. Large-scale integration of model organism and clinical trial data can provide a wealth of knowledge not available in individual sources and contextualize data back to these sources. In this context, with the machine learning platform we designed, ClioMD, we designed a program that is compatible with the HPO guideline and OMIM, which provides treatment and genetic counseling recommendations online in English, enabling individuals affected by ciliopathies such as Joubert syndrome, Cornelia de Lange, Bardet Biedl, etc. to get a fast and accurate diagnosis. In conclusion, the ClioMD platform enables you to explore the relationship between phenotype and genotype for disease and as a tool to help you make an accurate diagnosis.

Session Title: Omics Technologies Poster Wednesday Session

Board 1115W: Cohort-level differential distributional analysis for studying microglia in Alzheimer’s disease via single-cell RNA-sequencing

Authors:

K. Lin, **W. Zhang**, K. E. Prater; Univ. of Washington, Seattle, WA

Abstract:

Microglia are the resident immune cells in the brain, and their cellular changes in Alzheimer's Disease (AD) play a critical role in neuroinflammation and AD pathogenesis. Our study targets donor-level differences in their microglia to study their cell gene expression based on single-cell RNA sequencing (scRNA-seq). Current scRNA-seq methodologies primarily focus on the mean differential expression (DE) at the single-cell level, thus neglecting significant within-individual variability and broader distributional changes. Our research aims to provide a more accurate

and comprehensive representation of the microglia changes across donors based on their AD pathology beyond only changes in mean expression. Our differential test method is inspired by IDEAS (Zhang et al., Genome Biology, 2022), which tests for differential distribution via the Wasserstein distance. We additionally use an exact decomposition of this distance to investigate if the differences in gene expression distributions among donors are driven by changes in mean, variance, or shape. Differences in mean expression suggest repression or activated pathways and are the focus of most studies. However, changes in variance or shape suggest changes in gene coordination or cell state composition. We analyze data from the Seattle Alzheimer's Disease Brain Cell Atlas (Gabitto et al., Preprint, 2023) and other sources (Prater et al., Nature Aging, 2023). Our preliminary results highlight specific pathways differential in distribution based on AD burden that other methods overlook. Specifically, we find that most of the differential genes originate from changes in variance, not changes in mean. This finding suggests many undiscovered molecular mechanisms associated with increasing AD burden in microglia. Moving forward, we plan to corroborate our results by analyzing the microglia cells in the ROS-MAP consortium. Altogether, Our method enables a comprehensive picture of microglia changes during AD progression, and is applicable for studying population dynamics in other systems as well.

Session Title: Omics Technologies Poster Wednesday Session

Board 1116W: Community engagement efforts to decrease barriers to access to the undiagnosed diseases network for populations facing health disparities

Authors:

T. Cassini; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

The Undiagnosed Diseases Network (UDN) is a multicenter program that aims to provide answers to participants and further medical science. Given its limited capacity, the program has traditionally allocated acceptance to those who have undergone extensive evaluation prior to application. Often, this has meant that those facing barriers to access to care may not have met the threshold for acceptance. To explore methods to increase community engagement, the Vanderbilt Center for Undiagnosed Diseases (VCUD) was awarded supplemental funding for six months. This provided for 11 subject evaluations and additional resources to investigate methods to address barriers to access. We used a multifaceted approach to increasing awareness, recruitment, retention, and improvement for populations facing health disparities.

This included forming relationships with strategic advisory groups, including the Recruitment Innovation Center (RIC) at Vanderbilt and the Culture Shift Team (CST). The RIC assisted in designing the Clinician Study App, which is available on smartphones and allows potential referring providers to easily access the UDN eligibility criteria and referral page. They also facilitated the survey distribution to learn more about barriers precluding evaluation. Significant conclusions were that more non-white individuals, compared to white individuals, reported being denied a test and that significant barriers to access included time for visits and travel. The CST coordinated outreach events targeting the Latino and Black communities. These included in-person and media events estimated to reach 500 individuals directly and over 200,000 through media. We also formed bidirectional relationships with clinical partners and social organizations. The clinical partners include Shade Tree Clinic and Siloam Health, which care for uninsured or underinsured individuals, and WeCare Clinic, which cares for the Plain community. We have also partnered with Homeland Heart, a center for birthing and holistic wellness professionals of color. They have provided valuable feedback and served as a trusted messenger to patients. The VCUD has, in turn, provided educational opportunities, volunteer efforts, and additional resources for diagnostic evaluation.

Session Title: Omics Technologies Poster Wednesday Session

Board 1117W: Comparative analysis of cell type annotation methods for single-cell RNA-seq

Authors:

D. Chilin-Fuentes[1], G. Ildefonso[1], S. Rosenthal[2]; [1]UCSD Ctr. for Computation Biology and Bioinformatics, La Jolla, CA, [2]UCSD, La Jolla, CA

Abstract:

In the rapidly evolving world of single cell analysis, there currently exist many automated methods and tools for labeling clusters with an appropriate cell type annotation. These methods each have advantages and limitations, depending on factors such as tissue type, and level of granularity desired for the research project at hand. While it is known that using a larger reference of marker genes improves cell type annotation accuracy, many researchers currently rely on small lists of 'canonical', manually curated marker genes. In order to provide a resource to researchers looking to move to an automated method for annotating single cell clusters, we undertook a quantitative comparison of five commonly used cell type annotation tools (Cell Asign, ScType, SCINA, decoupleR, and CellTypist). We evaluated the performance of each tool on published single cell RNAseq datasets, and compared the cell type annotations to the cell type

in the published work, as well as how much overlap there is among the 5 annotation tools. We further conducted a simulation study where we injected random noise to the cluster marker genes, to evaluate how well each tool performs in environments with noisy signal. Of the cell type annotation tools tested, no single tool performed ideally, but certain tools were better suited to certain scenarios. For example, scType enabled the most flexibility of any tool tested, as it allows for user input of cell-type marker genes. However, this also required that the user has a list of high quality marker genes for their cell types of interest, which may not always be the case. Furthermore, we provide a summary of strengths and weaknesses of each tool, which may prove useful to researchers who are unsure of which cell type annotation tool would be best suited to their project.

Session Title: Omics Technologies Poster Wednesday Session

Board 1118W: Comparison of Five different library prep kits for mRNA sequencing optimization from low input and degraded FFPE samples

Authors:

N. Prasad, S. Singh, b. Bowman, C. Dixon, R. Beatty, R. Melanie, W. Dineen, U. Boris, H. Thomas; Discovery Life Sci., Huntsville, AL

Abstract:

Formalin-fixed paraffin-embedded (FFPE) techniques have emerged as cost-effective and long-term preservation methods for clinical samples, facilitating the utilization of next-generation sequencing (NGS) for expanded discovery and diagnostic purposes. Among the powerful NGS methodologies, mRNA sequencing stands out for its ability to delve into the transcriptome dynamics of genes, revealing insights into transcription levels and splicing variants. In this study, we assessed five commercially available RNA-Seq library preparation kits tailored for low-input, degraded FFPE samples: Illumina Stranded Total RNA Prep with Ribo-Zero Plus, KAPA RNA HyperPrep Kit with RiboErase (HMR) Globin, NEBNext Ultra II Directional RNA libraries, SMARTer Stranded Total RNA-Seq Kit v2, and Watchmaker RNA Library Prep Kits with Polaris Depletion. Each kit was evaluated using FFPE samples in triplicate at varying input amounts of 100ng, 50ng, 25ng, and 10ng, respectively. Sequencing was performed on NovaSeq 6000 as 100bp paired-end reads, generating approximately 50-100M reads per sample. All tested samples yielded sequencing-ready libraries successfully, boasting an average Q30 exceeding 90% across all preparations. Remarkably, the distribution of mRNA content exhibited consistency across all kits and sample types, with the KAPA kit demonstrating superior performance compared to others in this comparative analysis. Notably, intergenic reads from FFPE samples and input amounts

displayed variability, with the pico v2 kit exhibiting intergenic reads surpassing 20%, followed by Illumina at close to 15%, and the KAPA kit averaging around 10%. The KAPA HyperPrep and the Watchmaker RNA library prep methods are fully automated end-to-end on our Beckman i7 robots. The KAPA HyperPrep provides options to incorporate UMIs in the library prep. In conclusion, our comprehensive assessment underscores the KAPA RNA HyperPrep Kit with RiboErase (HMR) Globin as the preferred choice for processing highly degraded FFPE samples, particularly those with limited RNA input material.

Session Title: Omics Technologies Poster Wednesday Session

Board 1119W: Contamination detection in NGS libraries with mitochondrial DNA: Limits of detection and effects on library performance

Authors:

B. Lohman[1], A. Gerasimova[1], D. Goos-Root[1], R. Gutierrez[1], C. Ma[1], C. Elzinga[2]; [1]Quest Diagnostics, San Juan Capistrano, CA, [2]Quest Diagnostics, Marlborough, MA

Abstract:

Sample cross contamination in NGS libraries is an important quality control issue in both research and clinical laboratories. Existing methods to detect such contamination require high coverage because they rely on expected allele frequencies at predetermined genomic sites. One alternative is to use mitochondrial DNA. Because of the nature of inheritance of mitochondrial DNA, prior knowledge of mitochondrial phylogenies can be used to detect sample contamination; phylogenetically incompatible mitochondrial haplotypes present in the same sample is strongly suggestive of contamination. Here, we extend this approach by constructing standard curves of sample cross contamination in NGS library types which are of interest to basic researchers and clinicians. We test values of contamination not included in previous studies to establish the lower limit of contamination detection. We used well-characterized control samples from the Genome In A Bottle (GIAB) collection with high-quality variant calls. We built paired tumor-normal samples with a whole exome strategy and low-coverage whole-genome (lcWGS) libraries from these samples. After sequencing, fastq files from pure samples were mixed in known ratios to form in silico contaminated samples, using custom software, to produce samples with 0.5%, 0.75%, 1%, 2.5%, 5%, and 10% contamination. We processed all fastq files with their respective DRAGEN pipeline: paired tumor-normal or DNA with imputation. We reliably detected contamination in paired tumor-normal libraries at 0.5% for tumors with 300x coverage contamination and in lcWGS as 2.5% for 0.75x genome-wide coverage and using

Haplocheck, software that utilizes mitochondrial phylogenies to estimate contamination. We compared estimates of contamination from Haplocheck to those from DRAGEN, which uses allele frequency distributions, and found that while both approaches over-estimate contamination, DRAGEN is more accurate and has a lower standard deviation around its estimate. Finally, using the GIAB platinum truth set to estimate precision and recall, we established how the performance of these libraries is affected across a range of contamination. We found that lcWGS libraries can tolerate up to 2.5% contamination and paired-tumor normal samples can tolerate up to 1% contamination before performance begins to degrade. These limits provide guidance for sample quality control in laboratories of all sizes. The use of mitochondrial DNA for detection of sample contamination in NGS libraries is a sensitive screen that could be implemented in both small- and large-scale operations.

Session Title: Omics Technologies Poster Wednesday Session

Board 1121W: Cross-modal machine learning enables researchers to easily leverage public transcriptomics samples for data-driven discovery of biomedical contexts associated with their genes of interest

Authors:

P. Hicks, A. Krishnan; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Millions of publicly available transcriptomics samples are available via resources like the Gene Expression Omnibus. These data can be leveraged to discover associations between a gene set of interest and biological contexts such as tissues, phenotypes, diseases, and treatments. However, current methods/tools available for leveraging these data rely on tedious manual curation which doesn't scale to the millions of available samples, and precalculated differentially expressed gene signatures which limits discovered associations to predetermined biomedical factors. To address the limitations of current methods, we have developed *Exp2Meta*, a novel cross-modal machine learning framework for self-supervised data-driven discovery of biomedical contexts associated with any gene list of interest based on >200k public human transcriptomics samples. *Exp2Meta* leverages the paired relationship between genome-wide expression profiles and their unstructured free-text metadata to identify the biological contexts associated with an input gene set. It does so by building a machine learning model that identifies text features in metadata that are predictive of the input gene signatures. *Exp2Meta* is extremely flexible in accepting a variety of gene-based inputs including ranked or unranked gene sets, expression profiles from an entire sample or study, or logical combinations of gene

sets. In addition to discovering query-related contexts, *Exp2Meta* can also use its semantic modeling of unstructured text to prioritize samples and studies most relevant to the query, including those from other omics technologies (e.g., proteomics, metabolomics, etc.).

Session Title: Omics Technologies Poster Wednesday Session

Board 1122W: Cross-validation of next-generation sequencing for genetic stability assessment in cell therapy products

Authors:

J. Yun; Natl. Inst. of Food and Drug Safety Evaluation, Cheongju-si, Korea, Republic of

Abstract:

Cell therapy products are subject to genetic alterations as a result of in vitro culturing. This risk is particularly pronounced in stem cells, where genetic changes can lead to the development of tumors. The exact cause of these genetic alterations remains unknown, underscoring the need for thorough investigation. To ensure the genetic stability of stem cell-based products, regulatory guidelines from the FDA, EMA and MFDS strongly recommend appropriate testing methods; however, commonly used methods such as karyotyping, fluorescence in situ hybridization (FISH), and comparative genomic hybridization (CGH) arrays have limitations, low resolution, which hinder the detection of small structural changes or subtle chromosomal abnormalities and reduce the reliability of data related to precise genetic loci or targeted genomic regions. In this study, we assessed the genetic stability of hiPSCs and hiPSC-derived cardiomyocytes (CMs) during the proliferation and differentiation process using next generation sequencing (NGS)-based analysis, and cross validated it using two NGS platforms: synthesis by sequencing and nanoball sequencing. In addition, the analyzed genetic mutations data were compared with a high risk of cancer released on the COSMIC (Catalogue of Somatic Mutation in Cancer). We detected four mutations that were verified as high risk for tumorigenicity due to nonsense or frameshift mutations. *MUC4* c.8032_8033insA; p.(Pro2678fs) and *KMT2C* c.2263C>T; p.(Gln755Ter), *BCOR* c.1487_1500del; p. (Ile496AsnfsTer17) were analyzed to be associated with potential tumorigenicity.
Additionally, *AKAP9* c.4004_4006dupAA; p.(Lys1335_Leu1336insGln) could be associated with potential abnormal myocardium. Additional NGS platforms were used to validate the genetic mutations found in the stem cell-based products. The results confirmed the reproducible detection of these mutations. This approach increased the accuracy and reliability of the detected genetic variants. In conclusion, our study utilized NGS to uncover genetic variants that may emerge during the development of stem cell-based cell product, circumventing the

limitations of conventional techniques such as karyotyping and CGH array. This comprehensive approach offers valuable insights into potential genetic alterations that could impact the safety of stem cell-based product, thus contributing to enhanced quality control and regulatory compliance in the field of regenerative medicine

Session Title: Omics Technologies Poster Wednesday Session

Board 1123W: Curated loci prime editing (cliPE) for accessible multiplexed assays of variant effect

Authors:

J. Calhoun, C. Biar, C. Pfeifer, G. Carvill; Northwestern Univ., Chicago, IL

Abstract:

The number of variants of uncertain significance (VUS) reported is consistently increasing; it is paramount that we develop strategies to resolve VUS in an expedient manner. Multiplexed assays of variant effect (MAVEs) are a new approach which enables simultaneous characterization of many coding variants. Prime editing, a gene editing technology combining a nicking Cas9 with a reverse transcriptase to precisely edit endogenous loci, has been recently adopted for producing MAVE data. Our goal herein was to develop an accessible, low-cost experimental pipeline to enable many labs to perform subsaturation MAVEs on a target gene. Curated loci prime editing (cliPE) consists of the following steps: (1) prioritization of coding regions, (2) pre-screening of spacers for activity, (3) design and cloning of prime editing libraries, (4) transfection of haploid cells, (5) selection, (6) targeted amplicon sequencing, and (7) generation of functional scores (Enrich2). We developed the cliPE method using the gene *TSC2*. Loss-of-function variants in *TSC2* are a major cause of tuberous sclerosis resulting from mTOR pathway hyperactivity. Prime edited cells were selected by sorting for high or low levels of phosphorylated S6, a biomarker of mTOR activity. Our assay validated well on internal assay controls and a truth set of 25 missense variants, which enabled us to utilize functional scores at a moderate level of evidence for either benignity or pathogenicity. We generated functional scores for 106 *TSC2* VUS, which enabled resolution of 92 VUS; 86 variants were resolved to likely benign and 6 variants were resolved to likely pathogenic.

Session Title: Omics Technologies Poster Wednesday Session

Board 1124W: Delineating intercellular and intracellular gene regulation from spatial genomics data with SLEGRN

Authors:

Q. Liang, K. Chen; The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract:

Gene regulatory networks (GRNs) are dynamically shaped by ongoing biological processes of cells during cellular development or disease progression, or by various spatial contexts of cells that confer different intercellular signaling and microenvironmental cues. Although single-cell RNA-seq (scRNA-seq) and spatial transcriptomics (ST) technologies can uncover cell-to-cell transcriptome variations, most of the current GRN inference methods construct one global regulator-target network and thus fail to account for GRN dynamics. Cell-specific network approaches have been proposed and applied to scRNA-seq data. However, they all model regulator-target relationships using linear models (e.g., Pearson's Correlation), which is inaccurate with regards to the nonlinear nature of biological networks. To overcome these limitations, we introduce SLEGRN (spatial-local-explainable GRN), a novel cell-specific GRN inference method that employs Local Interpretable Model-Agnostic Explanations (LIME) techniques to derive cell-specific GRNs. SLEGRN first models target gene expression using intercellular signaling activity and transcription factor expression with an ensemble, tree-based model. Global regulators will be inferred based on the feature importance and local, cell-specific regulators will further be inferred through a LASSO regression model that approximates the global model at each locality. Thus, we capture nonlinear higher-order interaction of genes using tree-based methods, construct the GRNs for every single cell, and still retain good scalability. We demonstrate the utility of SLEGRN in the delineation of the contribution of intercellular and intracellular signaling molecules to gene expressions in human prefrontal cortex ST data. For instance, we found that some broadly expressed genes, such as NFIA, were dynamically regulated by different upstream signals, which also displayed layer-specificity in the cortex structure. Given the trend that an increasing number of ST studies are being conducted, we expect SLEGRN will substantially improve our understanding in the dynamics of gene regulation in spatial contexts.

Session Title: Omics Technologies Poster Wednesday Session

Board 1125W: Delineating the spatial molecular landscape of human heart transplant rejection using longitudinal cardiac biopsies

Authors:

A. Oill[1], A. L. Williams[1], R. Shah[2], K. Schlendorf, H[2], N. Chow[2], H. K. Siddiqi[2], J. N. Menachem[2], R. Hoffman, D[2], A. S. Rali[2], D. M. Brinkley[2], K. V. Amancherla[2], N. Banovich[1]; [1]Translational Genomics Res. Inst., Phoenix, AZ, [2]Vanderbilt Univ., Nashville, TN

Abstract:

Heart transplantation (HT) is the definitive therapy for end-stage heart failure. However, roughly 20% of patients experience at least one episode of acute rejection (AR) within the first-year post-HT. AR increases the risk of graft failure and is diagnosed by endomyocardial biopsy with H&E staining; however, AR grading shows substantial inter-pathologist variation. Similarly, treatment for AR is inconsistent within and between transplant centers, but is largely based on invasive hemodynamics and immunopathological findings. Improving understanding of the molecular underpinnings of AR, may inform novel diagnostic and therapeutic approaches and improve graft survival post-HT. Here, we leverage a unique longitudinal human cardiac tissue biorepository to identify cell-specific transcriptomic programs in a spatially-aware context during AR and following treatment with a wide range of anti-rejection therapies. Using the 10X Genomics Xenium platform we profiled expression of 477 genes at sub-cellular resolution from serial cardiac biopsies from heart transplant recipients. We analyzed 164,165 cells from a total of 199 biopsies across 63 patients. Patients were grouped into three broad categories of rejection - acute cellular, antibody-mediated, or mixed rejection - and analyses were anchored on treatment time point as well as outcome data. We identified 23 cell types and characterized cellular and molecular changes within canonical pathologies and treatment approaches. As heart transplantation is a last resort treatment in advanced heart failure, a better understanding of the molecular and cellular changes in human rejection may identify novel methods to accurately diagnose rejection while simultaneously offering the opportunity to provide personalized therapy, leading to longer survival and better outcomes.

Session Title: Omics Technologies Poster Wednesday Session

Board 1126W: Differential depth targeted gene panel sequencing to simultaneously detect somatic and germline variants in saliva-derived DNA

Authors:

R. O'Reilly[1], P. Harraka[1], D. Belluoccio[2], P. Yeh[3,4], K. Howlett[1,4], K. Behrouzfar[4], A. Rewse[1], H. Tsimiklis[1], The Archimedes Investigators, The ABC Study Investigators, G. G. Giles[1,5,6], J. L. Hopper[6], K. Bubb[7,8], S. J. Nicholls[7,9], R. L. Milne[5,1,6], M. C. Southey[1,5,10]; [1]Precision Med., Sch. of Clinical Sci. at Monash Hlth.,Monash Univ., Monash Univ., Melbourne, Australia, [2]Diagnostics and Genomics Group, Agilent Technologies, Melbourne, Australia, [3]Monash Haematology, Monash Hlth., Melbourne, Australia, [4]Dept. of Med., Sch. of Clinical Sci. at Monash Hlth., Monash Univ., Melbourne, Australia, [5]Cancer Epidemiology Div., Cancer Council Victoria, East Melbourne, Australia, [6]Ctr. for Epidemiology and Biostatistics, Melbourne Sch. of Population and Global Hlth., The Univ. of Melbourne, Melbourne, Australia, [7]Victorian Heart Inst., Monash Univ., Melbourne, Australia, [8]Biomedicine Discovery Inst., Monash Univ., Melbourne, Australia, [9]Victorian Heart Hosp., Melbourne, Australia, [10]Dept. of Clinical Pathology, The Univ. of Melbourne, Melbourne, Australia

Abstract:

Targeted gene panel (TGP) sequencing has been optimised to deliver consistent sequencing depth that has simplified data interpretation and facilitated lower cost & faster turnaround times for diagnostic and commercial applications. However, in some settings technologies that can measure genomic variation at different depths, in one assay are attractive as they reduce time, costs, and conserve biological resources. In our population-based epidemiological study setting that collects saliva as a source of DNA we developed an 80 gene custom TGP sequencing approach to detect germline variation (requiring sequencing depth >50x) and detect clonal haematopoiesis of indeterminant potential (CHIP; requiring sequencing depth >500x) in one workflow. To achieve differential depth in sequencing, probe ratios were adjusted for genomic regions relevant to identifying CHIP in comparison to those relevant to germline variation (**panel A**). An additional custom panel containing only the genomic regions relevant to the identification of CHIP was also manufactured (**panel B**).We assessed the capacity of **panel A** to detect both CHIP and germline variants in 156 blood-derived DNAs; 94 DNAs from research participants aged 65-75; 16 DNAs with known germline variants; 16 DNAs with known germline variants (titrated 0-100%); 24 DNAs from individuals aged <40 years; 3 commercial CHIP controls and 3 HMW DNA controls. The sequencing depth ratio between the CHIP and germline relevant genomic regions was 4.7:1. All fourteen CHIP associated variants were called in data from both panels (*DNMT3A* (10), *TP53* (2) and *TET2* (2)) at a median variant depth of 1382X. CHIP was not identified in the 24 DNAs from individuals <40 years. All known germline variants were identified in all test samples (251X median variant depth).We also considered if saliva derived DNA was a suitable template for detection of CHIP by using paired saliva and blood-derived DNA samples from the 94 research participants described above (188 test samples) and 4 controls using **panel B**. The 14 CHIP associated variants identified in blood derived DNA were also detected in the saliva DNA pair, demonstrating that saliva derived DNA is suitable for CHIP

analysis. The demonstration of targeted, differential depth sequencing using a single panel design and saliva-derived DNA that confidently detects variants with allele fractions ranging from >0.01-1 in one workflow offers a cost-effective solution for many situations including combining the detection of CHIP and germline variation. We have successfully applied our assay to 2,500 prospective cohort study participants to characterise germline and somatic variation in saliva derived DNA.

Session Title: Omics Technologies Poster Wednesday Session

Board 1127W: Dynamics of age-dependent alterations in the human cardiac transcriptome at single-nucleus resolution

Authors:

M. Kim, S-J. Paek, K-T. Kim; Ajou Univ. Sch. of Med., Suwon, Korea, Republic of

Abstract:

The increasing prevalence and fatality rates of heart failure are largely attributed to age-related alterations in cardiac structure and function. Recent efforts to construct a comprehensive atlas of heart cells have enhanced our understanding of cardiac cell-by-cell functional topologies. However, the dynamics of cardiac cell-to-cell signaling transduction across different stages of the lifespan and the molecular markers responsible for aging-related heart failure remain challenging. In this study, we characterized the heterogeneous cardiac cellular states in an age-dependent manner by integrating 1.5 million single-nucleus RNA sequencing profiles from four independent cohorts, comprising 52 healthy donors and 33 heart failure patients aged 22 to 72 years. To ensure accurate cell identification, we employed variation inference-based batch correction to meticulously remove batch effects and address variations arising from different data sources and technical discrepancies. Using a non-linear modeling approach as a function of aging, we identified cell type-specific aging-associated changes and quantified differentially expressed genes that vary with age, thereby leading to the development of a cardiac aging clock model. This model predicted biological age and highlighted cell type-specific changes, demonstrating that age-related alterations in healthy hearts also manifest in disease states. The signatures identified by the aging clock model showed significant enrichment with the gene set of the Aging Atlas, that was consistent with our findings. Furthermore, we observed that intercellular communication between cardiomyocytes, endothelial cells, and fibroblasts, that are essential for maintaining cardiac function, diminished with aging, similar to the reduction observed in heart failure conditions. Importantly, we also detected an increase in inflammation-related signatures within the cardiac immune system as a function of aging. Our findings

underscore the significant impact of aging on heart failure development. By leveraging large-scale data sources, we identified major cardiac cell types and explored cellular diversity in the context of aging, providing novel insights into the molecular perturbations associated with heart failure.

Session Title: Omics Technologies Poster Wednesday Session

Board 1128W: Empowering Biological Research: An Integrated Ecosystem of Tools in the Reactome Pathway Knowledgebase

Authors:

N. Li, Reactome Team; Ontario Inst. for Cancer Res., Toronto, ON, Canada

Abstract:

The Reactome Pathway Knowledgebase, supported by NIH NHGRI, is recognized as a gold-standard data resource due to its meticulous human curation practices. However, the complexities inherent in handling vast biological data challenge traditional curation methods. To address this, we're developing an integrated ecosystem of tools that address the needs of users across biological domains, biological research areas, and various technical backgrounds. Our current developments are toward an AI-augmented curation system that more accurately surveys the literature for new curation opportunities, extending our coverage of known human biology. Our web-based curation platform, empowered by Large Language Models (LLMs) is guided by Reactome's data schema and curation requirements, this platform leverages RAG (retrieval-augmented generation) technology to efficiently associate unannotated genes with Reactome pathways and extract functional relationships from literature evidence. Reactome's rigorous curation guidelines require manual curation and peer-review prior to incorporation into the database. Simultaneously, Reactome is pioneering a conversational chatbot, providing users with intuitive access to complex pathway information and encouraging deeper engagement. Future integration of LLMs into the chatbot will enable sophisticated analyses within Reactome, catering to users with diverse technical backgrounds.Complementing this, ReactomeGSA facilitates multi-omic pathway analysis by integrating diverse datasets onto a common pathway framework. Accessible via web interface, Bioconductor R package, and API, ReactomeGSA optimizes analyses by incorporating public datasets and supports pathway-level differential expression analysis, thus serving as a crucial tool for omics data interpretation.Furthermore, Reactome's developmental lineage path project aims to provide a continuously updated repository of cell lineage paths, aiding researchers in single-cell omics data analysis for cell lineage tracing studies. Collaborations with biological experts enhance Reactome's annotations,

while user-friendly tools cater to researchers across various bioinformatics expertise levels. Together, these initiatives form an integrated ecosystem aimed at streamlining data curation, analysis, and interpretation, thereby advancing biological research and fostering deeper understanding of complex biological systems.

Session Title: Omics Technologies Poster Wednesday Session

Board 1129W: Enabling spatial transcriptomics to reveal gene expression landscapes with near single-cell resolution in large-sized tissues with iStar-XL

Authors:

A. Schroeder[1], M. Loth[1], C. Luo[1], D. Zhang[1], S. Piya[2], E. Plowey[2], W. Hu[2], S. Chan[2], T. Reynolds[2], T. Carlile[2], P. Cullen[2], H-H. Tsai[2], B. Zhang[2], M. Li[1]; [1]Dept. of Biostatistics, Epidemiology and Informatics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]Res. Dept., Biogen Inc., Cambridge, MA

Abstract:

Recent developments in spatial transcriptomics (ST) technologies have enabled comprehensive gene expression profiling across tissues while preserving crucial spatial context. However, existing ST platforms often suffer from high cost, low resolution, limited genes, and small tissue capture areas, restricting their use for large-scale investigations, especially when studying sizable tissues common in human studies. To address these significant challenges, we propose iStar-XL (**I**nferring **S**uper-resolution **T**issue **AR**chitecture in e**X**tra **L**arge-sized tissues), a novel framework for predicting gene expression across large tissues at near single-cell resolution. The model extracts both local and global tissue structures from large-sized histology images. By leveraging information from a small set of standard-sized ST captures, iStar-XL utilizes a deep learning neural network framework to learn the relationships between the histological features and gene expression patterns. The model provides gene expression estimates for each 8 µm × 8 µm superpixel across the entire large-sized tissue section, including regions where gene expression was not previously measured. Through the application of iStar-XL to multiple sclerosis (MS) brain tissue samples with size ~2 cm × 3 cm, we demonstrate its utility in assisting scientists with automatic high throughput and unbiased tissue annotation, inferring cell type composition, and pinpointing regions of interest for features not clearly discernible through human visual assessment alone. iStar-XL also revealed MS lesion associated cell characteristics and immune responses not captured by conventional ST experiments. The accuracy and robustness of iStar-XL's predictions were confirmed through rigorous evaluations, involving IHC

staining of crucial markers, manual annotations by pathologists, and assessment of prediction accuracy and reliability using a large hold-out test set. Overall, iStar-XL has broad applicability for studying in-depth tissue architectures and cellular-level spatial relationships in large-sized tissues across diverse tissue types and diseases. The iStar-XL framework is versatile and can be applied to a wide-range of ST platforms enabling the prediction of single-cell resolution gene expression in tissue regions outside the directly measured field of views (FOVs), expanding gene expression analysis across entire tissue sections. By creating high-resolution spatial gene expression maps for large-sized tissues, iStar-XL provides crucial information on disease development and therapeutic response, allowing for significant advancements in the field of spatial omics.

Session Title: Omics Technologies Poster Wednesday Session

Board 1131W: Expression, splicing or protein QTLs have substantial contributions to SNP-based heritability and polygenic prediction in complex traits ⭐

Authors:

S. Liu[1], J. Yang[2,3], P. Visscher[4], J. Zeng[5]; [1]Inst. for Molecular BioSci., Univ. of Queensland, Brisbane, Australia, [2]Sch. of Life Sci., Westlake Univ., Hangzhou, China, [3]Westlake Lab. of Life Sci. and Biomedicine, Hangzhou, China, [4]Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [5]The Univ. of Queensland, Brisbane, Australia

Abstract:

Genetic variants can affect trait phenotype by regulating gene expression and translation. However, recent studies have onlyfound limited overlap between GWAS hits and expression quantitative trait loci (eQTL). In this study, we systematicallyevaluated the contributions of expression (eQTL), splicing (sQTL) or protein QTL (pQTL) on the SNP-based heritability andpolygenic prediction in complex traits. Using a state-of-the-art Bayesian mixture model (SBayesRC), we integrated GWAS summarystatistics of 39 complex traits/diseases from the UK Biobank (UKB) with functional annotations defined by the xQTL from blood(eQTL data from eQTLGen, sQTL from GTEx, and pQTL from the UKB). In the analysis, we focused on the cis-region of eachgene/protein with a 1Mb flanking window on either side. To account for LD between SNPs, we conducted conditional and jointassociation (COJO) analysis to identify conditionally independent xQTL signals (P-value &lt 5×10^(-6)) for each gene/protein. Byassigning a value of one to these COJO significant xQTL and zero to the remaining SNPs, we

created a binary genomic annotationfor the eQTL, sQTL and pQTL, respectively, as well as a union of these annotations, referred to as the espQTL. Of the 7.4million SNPs used in the analysis, 1.1%, 0.1%, 0.2%, and 1.3% were annotated as eQTL, sQTL, pQTL and espQTL, respectively.Incorporating these xQTL annotations in SBayesRC found that, on average, 20% of the SNP-based heritability were explained byespQTL, 14% by eQTL, 3% by sQTL, and 4% by pQTL, with a substantial variability across traits. Of note, blood markers had thelargest proportions of SNP-based heritability explained (41% by espQTL) and the largest folds of per-SNP heritabilityenrichment (27x in eQTL, 75x in sQTL and 149x in pQTL, respectively). Polygenic prediction in a hold-out sample showed that 34%of the prediction R2 were attributed to the espQTL, comparable to the contribution of coding or evolutionarily constrainedregions with a similar number of SNPs in the annotation. Consistent with the SNP-based heritability estimation, these xQTL (allassayed in blood) had a major impact on the polygenic prediction in blood markers, contributing to 44% of the predictionaccuracy by espQTL and 17% on average across the three types of xQTLs (36x enrichment in per-SNP prediction accuracy by espQTLand 65x on average across xQTLs). Our results showed substantial contributions of expression, splicing or protein QTL to thegenetic variation and phenotype prediction in complex traits, highlighting the importance of analysing xQTL data from thetrait-relevant tissues.

Session Title: Omics Technologies Poster Wednesday Session

Board 1133W: Gene expression profiling of peripheral blood mononuclear cells from women with cervical lesions reveals new markers of cancer.

Authors:

M. Ndiaye[1], G. Diop[1], C. Derbois[2], J. Spadoni[3], J. Noirel[3], R. Medina-Santos[3], C. Coulognes[3], M. Torres[4], A. Dieye[1], M. Sembene[1], J. Deleuze[5], A. Toledano[6], A. Dem[1], J. Zagury[3], S. Leclerc[3]; [1]Univ. Cheikh Anta Diop of Dakar, Dakar, Senegal, [2]Genopole CGNRH, Paris, France, [3]CNAM, Paris, France, [4]Univ. marseille, Marseille, France, [5]CEA, EVRY, France, [6]HESAM Univ., Paris, France

Abstract:

Cervical cancer (CC) is a multifactorial disease of which human papillomavirus (HPV) is the main etiological agent. Despite cervical Pap smear screening and anti-HPV vaccination, CC remains a major public health issue. Identification of specific gene expression signatures in the blood could allow better insight into the immune response of CC and could provide valuable information for the development of novel biomarkers. The present study performed a transcriptomic analysis of peripheral blood mononuclear cells (PBMCs) from Senegalese

patients with CC (n=31), low-grade cervical intraepithelial neoplasia (CIN1; n=27) and from healthy control (CTR) subjects (n=29). Individuals in the CIN1 and CTR groups exhibited similar patterns in gene expression. A total of 182 genes were revealed to be differentially expressed in patients with CC compared with individuals in the CIN1 and CTR groups. The IL1R2, IL18R1, MMP9 and FKBP5 genes were the most upregulated, whereas the T-cell receptor α gene TRA was the most downregulated in the CC group compared with in the CIN1 and CTR groups. The pathway enrichment analysis of the differentially expressed genes revealed pathways directly and indirectly linked to inflammation. To the best of our knowledge, the present study is the first large transcriptomic study on CC performed using PBMCs from African women; the results revealed the involvement of genes and pathways related to inflammation, most notably the IL-1 pathway, and the involvement of downregulation of the T-cell receptor α, a key component of the immune response. Several of the stated genes have already been reported in other cancer studies as putative blood biomarkers, thus reinforcing the requirement for deeper investigation. These findings may aid in the development of innovative clinical biomarkers for CC prevention and should be further replicated in other populations.

Session Title: Omics Technologies Poster Wednesday Session

Board 1134W: Genetic Architecture of plasma metabolites in the Southeast Asian population ⭐

Authors:

P. Jain[1], N. Sadhu[1], D. Low[1], T. Mina[1], D. Tay[1], H. Ng[1], X. Wang[1], M. Hebrard[2], P-A. Goy[2], K. Wong[3], R. Sarangarajan[3], N. Bertin[2], M. Lam[1], P. Eliott[4], E. Riboli[4], J. Ngeow[5], P. Tan[2], G. Michelotti[3], P. Sheridan[3], M. Loh[1], J. Chambers[1]; [1]LKC Med., Nanyang Technological Univ., Singapore, Singapore, [2]Genome Inst. of Singapore, A-STAR, Singapore, Singapore, [3]Metabolon Inc., Durham, NC, [4]Imperial Coll. London, London, United Kingdom, [5]Natl. Cancer Ctr. Singpore, Singapore, Singapore

Abstract:

Metabolites and variation in their levels provide key information about the wellbeing of humans and the mechanisms that control key physiological processes involved in health and disease. Genome wide association studies (GWAS) have identified many genetic loci associated with metabolic variation, but most of these studies have been conducted in individuals of European ancestry. Here, we conduct the first large scale GWAS of 1073 plasma metabolite levels in 8200 multi-ethnic, Asian individuals in the Health for Life in Singapore (HELIOS) study (53±12 years, female=59%, Chinese=67%, Malay=14%, Indian=19%). Linear mixed model-based ancestry

specific GWAS analysis was performed followed by inverse variance weighted fixed effect meta-analysis. We identified 2617 independent genetic variants across 502 genomic loci associated with 852 metabolites at genome-wide significance ($P<5x10^{-8}$), with multiple loci previously unreported in other metabolite GWAS studies. We observed that the genomic loci are highly pleiotropic with a median of two associations per locus. The most associations observed was with the Fatty Acid Desaturase (FADS) locus (116 lipids, 2 amino acids and 2 unknown metabolites). We also observed high polygenicity of associations with multiple metabolites (median of two independent associated loci per metabolite), with four metabolites having nine independently associated loci. Colocalization analysis with gene expression quantitative trait loci (eQTL) identified 2783 metabolite-gene pairs (637 metabolites and 726 genes) with shared causal variant (Coloc H4>0.6). Ancestry specific GWAS results also highlighted associations observed only in one population, including 63 independent variants associated with 82 metabolites in Indians and 41 variants associated with 42 metabolites in Malay, that were not observed in the Chinese sub-population. This study provides a valuable resource to understand the genetic architecture of metabolites in Asian populations and to help identify novel genetic loci previously undiscovered in large-scale European GWAS studies. It delivers insights into the roles of metabolites in human health, thereby, offering opportunities to determine novel diagnostic and therapeutic targets.

Session Title: Omics Technologies Poster Wednesday Session

Board 1135W: Genetically-predicted protein levels in UK Biobank and NIH *All of Us* identify thousands of associations of proteins and clinical phenotypes

Authors:

H. Taylor[1,2], Y. Xu[2], D. Loesch[3], J. M. Keaton[1], T. C. Tran[1], S. Lambert[2], S. Ritchie[2], X. Jiang[2], C. Foguet[2], E. Persyn[2], C. Ben-Eghan[2], Y. Liu[2], L. Gil[2], L. Lannelongue[2], D. Jin[2], B. Sun[4], H. Runz[4], C. Whelan[5], S. Petrovski[3], A. Nag[3], C. Zeng[1], D. Paul[3], F. S. Collins[1], A. S. Butterworth[2], J. Danesh[2], J. C. Denny[1], M. Inouye[2]; [1]Ctr. for Precision Hlth.Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]British Heart Fndn. Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [3]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, [4]Translational Biology, Res. and Dev., Biogen Inc., Cambridge, MA, [5]Data Sci., Janssen Pharmaceuticals, Princeton, NJ

Abstract:

Multi-omic studies provide an opportunity to profile and map the underlying molecular mechanisms driving human disease. However, such studies are difficult to perform due to the cost of the assays, availability of relevant tissues *in vivo*, and other factors. Through population-scale cohorts with vast genetic data and the improvement of genetic scoring methods to predict multi-omic information, human genetics offers a cost-efficient, reliable bridging strategy which can democratize multi-omics studies, reducing inequities for researchers without the resources to generate such data.
To explore the role of proteins in disease, we developed genetic scores for 2,683 proteins and performed phenome-wide association studies (PheWAS) across 1,113 clinical phenotypes using the UK Biobank (UKB) and NIH *All of Us* program (AoU). We trained genetic scores for each protein in the UKB Olink Explore panel measured in 54,219 participants using Bayesian ridge regression and a subset of the cohort. To assess the accuracy of the scores, we compared predicted to measured protein levels in a validation set of the UKB data which was withheld from score development. Of the 2,683 proteins, 1,712 had an $R^2 \geq 0.01$ and 166 had an $R^2 \geq 0.3$. We also assessed the scores in an external cohort and found similar performances. Finally, we performed ancestry-specific PheWAS with predicted protein levels in >600,000 participants from UKB and AoU and meta-analyzed the results.
We identified >9,000 protein-clinical phenotype associations (Bonferroni-adjusted P<0.05), including ≥1 association for 61.9% of the proteins and 25.3% of the phenotypes. Generally, we observed the greatest number of associations with phenotypes in the endocrine/metabolic, circulatory system, and digestive disease classes. Of the most associated clinical phenotypes, we identified several autoimmune diseases such as celiac disease, type 1 diabetes, and multiple sclerosis. Focusing on proteins, we found SSH3, ATP6V1D, and RAP1A to be highly pleiotropic, associating with several different classes of phenotypes. We recovered many well-established protein-clinical phenotype associations, such as the protective association of APOE with Alzheimer's disease. We also identified hundreds of potentially novel associations, including several not clearly identified via case-control GWAS, such as a protective association of PPY with type 2 diabetes. Together, these results provide additional insight into the genetics regulating disease pathophysiology and identify novel proteins with clinical relevance, underscoring the etiological value and accessibility of genetic prediction of proteomes.

Session Title: Omics Technologies Poster Wednesday Session

Board 1136W: Haplotype reconstruction using low-coverage genotyping-by-sequencing in genetically diverse mice

Authors:

S. Widmayer[1], S. Gude[2], L. Wooldridge[3], M. Barter[3], Q. Meng[3], B. Dumont[3], D. Gatti[3]; [1]The Jackson Lab., Farmington, CT, [2]seqWell, Beverly, MA, [3]The Jackson Lab., Bar Harbor, ME

Abstract:

Population and forward genetics studies require the collection of phenotypes and high-resolution genotypes from the same individuals. Genotyping microarrays have been the standard approach for collecting genotypes at thousands of loci. Microarrays are being replaced by low-coverage genotyping-by-sequencing (lcGBS) due to the continuing cost reductions of next-generation sequencing methods. However, one remaining bottleneck for high-throughput genotyping-by-sequencing is scalable library preparation. Here, we used the seqWell purePlex DNA Library Prep Kit automated on a Revvity Sciclone, and short-read sequencing to genotype 48 Diversity Outbred (DO) and 48 four-way cross (4WC) mice. We first genotyped the mice using the Giga Mouse Universal Genotyping Array (GigaMUGA), which contains 143,259 markers, and used these genotypes to reconstruct the genomes of the samples using *qtl2*. In parallel, we sequenced the same samples on an Illumina NovaSeq X plus 10B flow cell to a mean coverage of 1X. We then imputed roughly 39 million SNPs in each cross using *QUILT*, a software which performs precise SNP imputation in multiparent populations using reference haplotypes. The DO and the 4WC are derived from either 8 (DO) or 4 (4WC) inbred strains and the sample genomes can be reconstructed in terms of the founder haplotypes. This enables quantitative trait locus (QTL) mapping against the founder alleles instead of just the SNP genotypes. At full coverage, the concordance between founder haplotypes obtained using the GigaMUGA and lcGBS was at least 0.9. We downsampled the alignment files and found that concordance remained above 0.9 down to 0.1X coverage. We then applied the same experimental approach to an additional 183 DO samples and performed local expression QTL mapping. At similar coverage levels, lcGBS facilitated equivalent genetic mapping power and precision to the GigaMUGA at a fraction of the cost. Overall, we find that the lcGBS approach has a straight-forward bench protocol and produces equivalent results to array-based approaches at a lower price point.

Session Title: Omics Technologies Poster Wednesday Session

Board 1137W: Harmonized framework for RNA-seq-based rare disease diagnostics in a pan-continental consortium - Solve-RD

Authors:

V. Yepez[1], A. Esteve-Codina[2], J. Gagneur[3]; [1]Technical Univ. of Munich, Garching bei München, Germany, [2]CNAG, Barcelona, Spain, [3]Technical Univ. of Munich, Garching, Germany

Abstract:

RNA sequencing is a powerful and complementary approach to DNA sequencing to discover disease-causing gene regulatory defects in individuals affected by genetically undiagnosed rare disorders. However, as large rare disease consortia are implementing RNA-seq-based diagnostics, new data analysis challenges arise to cope with the heterogeneity of diseases, various tissues and sample sizes, and the multiplicity of interpreters. Here, we present new guidelines, algorithms, and methodologies that have been developed by the RNA-seq group of the Pan-European consortium Solve-RD to address these issues. The Solve-RD RNA-seq dataset consists of >450 samples collected from 4 tissues from individuals tested at more than 30 clinics from 6 European Reference Networks. The vast majority of the samples have matching DNA sequencing data and, often, of the parents. The samples were analyzed using the workflow DROP, which includes the recent release of the aberrant splicing caller FRASER 2.0. Two new RNA-seq quality control modules were added to DROP to confirm sex and tissue annotations. Both helped to detect a handful of mislabelled samples. To standardize results verification, we devised a consortium-wide annotation framework that integrates scoring of an aberrant event at four levels: i) raw data support of the outlier calls, ii) gene function, iii) variant(s) with a plausible regulatory mechanism (e.g. splice site or NMD), and iv) variant(s) segregation. Automatically generated reports were designed to concisely present all data needed for the annotators. To further increase the consistency of the annotations, we have conducted in-presence workshops, Solvathon, combining training and collective data inspection sessions, and offered weekly consultancy sessions. The analysis has led to 20 solved cases and >40 strong candidates including one new disease gene. Altogether, our new algorithms and processes contribute to improved RNA-seq-based rare disease diagnostics methodologies.

Session Title: Omics Technologies Poster Wednesday Session

Board 1138W: High throughput ancient DNA sequencing using the Ultima Genomics platform ⭐

Authors:

S. Mallick[1,2,3], I. Soifer[4], N. Rohland[1,3,5], M. Coole[3], P. Winer[4], D. Lipson[4], A. Bernier[3], M. Nawaz[1,2], M. Sexton[4], O. Barad[4], M. Michel[5,6], E. Brielle[5], T. Clark[4], B. Weller[4], A. Jaimovich[4], S. Pollock[4], K. Aguinaldo[4], N. Patterson[3,5], D. Bogumil[4], G. Soos[1], M. Mah[1,2,3], A. Herbig[6], S. Gabriel[3], G. Almogy[4], D. Reich[1,2,3,5]; [1]Harvard Med. Sch., Dept. of Genetics, Boston, MA, [2]Howard Hughes Med. Inst., Boston, MA, [3]The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Ultima Genomics

Inc., Fremont, CA, [5]Harvard Univ., Dept. of Human Evolutionary Biology, Cambridge, MA, [6]Max Planck Inst. for Evolutionary Anthropology, Leipzig, Germany

Abstract:

Ancient DNA (aDNA) sequencing has been challenging due to the naturally short fragments, the low percentage of the organism's DNA, and high microbial contamination. To address these issues, most samples have been analyzed using in-solution enrichment to target specific locations throughout the genome. However, this capture platform introduces ascertainment bias and cannot identify novel/structural variants or efficiently extract pathogens.
To overcome these limitations, we use the Ultima Genomics (UG) sequencing platform, which generates a high number of short sequences at ultra-low cost: a standard 30x human genome for $100. aDNA analysis is particularly well-suited to this platform as the naturally degraded aDNA fragments can be fully read by short single-ended reads with a run time less than 10 hours, enabling high throughput of up to 200B reads per week on one system.
In a collaboration between Harvard Medical School, the Broad Institute and Ultima Genomics we developed a sequencing workflow for aDNA, optimized in both wet-dry protocols, which reduces per base pair cost 3-fold relative to the current most cost-efficient Illumina platform. The method works with previously prepared libraries after conversion, and we are working on native library preparation protocols to make sequencing even more cost-efficient.
Alignment can be challenging for short noisy aDNA molecules. We implemented a variant aware alignment strategy with the UG-optimized aligner which is more than one order of magnitude faster than current approaches and decreases reference bias. All samples are demultiplexed and aligned on the sequencer directly to the hs37d5 reference sequence, with the RSRS inferred ancestral mitochondrial genome, and additional pseudo-chromosomes for screening pathogens. Users can switch in alternate reference genomes as needed for other studies. We have now sequenced thousands of samples also previously sequenced using Illumina, and found minimal biases or batch effects, with these data fully co-analyzable.
Making shotgun sequencing of aDNA economical for the great majority of analyzed samples has the potential to be revolutionary for ancient DNA analysis, making possible (1) analyses of population history unaffected by SNP ascertainment bias - a particular challenge for analysis of Africans, (2) analyses of population history based on the site frequency spectrum - not possible with capture methodology, (3) threading of ancient genomes through ancestral recombination graphs, (4) comprehensive study of structural variation, (5) more powerful analysis of adaptation over time, and (6) more powerful analysis of co-evolution of humans and pathogens.

Session Title: Omics Technologies Poster Wednesday Session

Board 1139W: High throughput single-cell TCR sequencing method on aging mouse T cells and antigen stimulation in human T cell powered by split pool combinatorial barcoding technology

Authors:

M. Curca[1], E. Papalexi[1], B. Hariadi[2], A. Suyama[1], S. Marrujo[1], A. Kumar[1], G. Geiss[1], C. Roco[1], A. Rosenberg[1]; [1]Parse BioSci.s, Seattle, WA, [2]Parse BioSci.s Inc., Seattle, WA

Abstract:

Single cell profiling of T cell receptors (TCRs) and gene expression enables the characterization of key T cell subsets with their antigen recognition and killing capabilities. However existing microfluidic-based methods have limited cell throughput making it impossible to capture the vast diversity of the TCR repertoire. To overcome these limitations, we developed an approach using combinatorial barcoding to simultaneously characterize the TCRs alongside the full transcriptomes of up to 1 million T cells from up to 96 unique samples. We demonstrate this approach in both human and mouse samples and profile millions of T cells. We first started with one million human T cells stimulated with tumor, viral, and autoantigens and included non-stimulated T cells from three donors as controls. Using our approach we recovered paired TCR clonotype information in 82% of all T cells. We estimated the clonotype diversity and found that the antigen-stimulated samples had lower diversity than the controls. Moreover, we used the top ten most expanded clonotypes in every sample, ran multiple sequence alignment and identified antigen-specific sequence motifs, some of which have been previously described. We also profiled 986,455 mouse spleen-resident T cells at different stages of life (1 to 27 month old mice). We performed transcriptome-based clustering and demonstrated that this approach identifies all major mouse T cell populations (CD4+, CD8+, Tregs, etc.) while efficiently detecting paired TCR sequences. Furthermore, consistent with previous findings, we detect shifts in memory and naive T cell proportions as well as differences in clonotype diversity that correlate with aging in mice. In summary, we present a highly flexible and scalable approach to single cell TCR profiling to allow researchers to profile up to 1 million of T cells in a single experiment. We envision our approach to be invaluable in investigating functional T cell responses during infection, cancer, autoimmunity, or therapeutic interventions.

Session Title: Omics Technologies Poster Wednesday Session

Board 1140W: Identification of non-coding RNAs as early prenatal screening biomarkers for congenital heart diseases by integrated analysis of Omics and clinical data

Authors:

T. Gong, X. Zhang, Y. Qiu, Z. Jin, Y. Gong, L. Zhang, H. Deng, C. Li, S. Zou, C. Sun, J. Huang, H. Sun, Q. Zhang, Z. Liu, C. Zhang; The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China

Abstract:

Congenital heart disease (CHD) is one of the most common birth defects worldwide. Numerous studies have demonstrated that non-coding RNAs (ncRNAs) may play pivotal roles in gene expression regulation during heart development and contribute to the pathogenesis of CHD. In this study, we performed an integrated analysis of the public and in-house omics and clinical data to identify novel potential ncRNA biomarkers in plasma samples of the pregnant women bearing fetuses with CHD, and then employed various bioinformatics tools and databases to analyze their target genes. We also developed an early non-invasive diagnostic model for CHD using machine learning algorithms, and subsequently identified three microRNAs (*miR-19b, miR-29c, miR-375*), two long non-coding RNAs (*LINC00598*, lncRNA *SNHG6*), and one circular RNA (*hsa_circRNA_100709*) that are associated with CHD. We are in the process to conduct a multi-center, large-scale study to validate these biomarker candidates. Our findings suggest that the maternal plasma-derived ncRNAs may not only play important roles in the pathogenesis of CDH, but also become potential diagnostic biomarkers for fetal CHD.

Session Title: Omics Technologies Poster Wednesday Session

Board 1141W: Improvements in Combinatorial scRNA Sequencing that Enable Scaling to Hundreds of Samples and Millions of Cells in a Single Experiment

Authors:

V. Tran[1], A. Sapre[1], E. Papalexi[1], B. Hariadi[2], A. Suyama[1], J. Pangallo[1], M. Curca[1], O. Kaplan[1], S. Schroeder[1], S. Marujo[1], A. Salvino[1], R. Koehler[3], G. Geiss[4], C. Roco[1], A. Rosenberg[1]; [1]Parse BioSci.s, Seattle, WA, [2]Parse BioSci.s Inc., Seattle, WA, [3]Parse BioSci.s.com, Seattle, WA, [4]Nanostring, Seattle, WA

Abstract:

As single cell RNA sequencing has become more ubiquitous across different research applications, the average number of cells profiled per project has exponentially increased year-over-year. The scope and ambition of these studies is increasing, with researchers routinely including multiple biological replicates, samples across time-course studies, as well as deep profiling of specific cell types to identify unique populations. As the scope and scale of single cell experiment grows, the sequencing infrastructure and computational tools are keeping pace. It's necessary to take advantage (and test the limits) of scRNA-seq and explore the possibilities of what large single cell datasets can do to elucidate biological insights. Given the limitations of conventional single cell approaches, many researchers are constrained in terms of how big they can scale their experiments in terms of samples and overall input. But the versatility of combinatorial barcoding technology has driven development of scaling at a much faster pace. There have been numerous developmental breakthroughs that offer distinct advantages over other methods. Here we present improvements in sensitivity, mappable reads and lower sample input requirements with large increases in sample multiplexing and the numbers of cells that can be processed in a single experiment. We demonstrate this approach by processing hundreds of samples in concurrence with scaling beyond 1 million cells in a single experiment. The process is compatible with automation using standard liquid handling instruments for even higher throughput applications. This approach democratizes the scaling of single-cell sequencing experiments to make processing 1 million cells the standard, inspires the scientific community to think bigger and position the field for significant advances in single-cell RNA sequencing applications.

Session Title: Omics Technologies Poster Wednesday Session

Board 1142W: Inferring phenotypes from tabular genotype data with pre-trained Large Language Model ⭐

Authors:

S. Yang[1], L. Joseph[1], X. Liu[2], J. Baik[1], D. Li[3], T. Chen[4], L. Shen[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]RIKEN, Yokohama, Japan, [3]Univ. of California, San Diego, San Diego, CA, [4]Massachusetts Inst. of Technology, Boston, MA

Abstract:

Background: Predicting phenotypes with complex genetics bases based on a small, interpretable set of variants remains a challenging task. Conventionally, data-driven approaches for feature selection and predictive model construction are commonly utilized for this task, yet with significant room for improvement. Recent advancements of Large Language Model (LLM) have

demonstrated unprecedented capabilities across various applications including diverse biomedical tasks: studies showed pretrained LLMs such as OpenAI's GPTs can leverage the knowledge encoded in the model weights and adapt to effectively perform cell type annotation, gene set selection and functional annotation, causal genes finding, and so on. Here, we assess the ability of LLM for phenotype inference given tabular genotype data and present a knowledge-driven framework. Methods: As a proof of concept, we choose OpenAI's latest GPT-4o model and focus on genetic ancestry prediction. We develop a knowledge-driven framework: (1) first use zero-shot chain-of-thought (CoT) to prompt GPT-4o to suggest a set of ancestry informative SNPs given ancestry phenotype labels from 1KGP ({African, American, East Asian, European, South Asian}). (2) extract SNP genotype tables and labels from 1KGP (got 2501 samples with relatively balanced class sizes) and serialize each row to text descriptions. (3) feed the descriptions to GPT-4o to make inference of the genetic ancestry with CoT and few-shots in-context learning and compare against k-nearest neighbor (kNN), random forest (RF), and support vector machine (SVM) using PCA for variants selection (equal number), with a range of few-shots training sample sizes. Results: Firstly, with 15-shots, our knowledge-driven inference already achieves 74.8% classification accuracy, significantly outperforming data-driven approaches (PCA variants + RF, kNN, SVM) that achieves 35.8% accuracy at best. Similar results were also observed for larger training sizes (20, 25, ...45). Moreover, as an ablation, we found the GPT-suggested variants also performed much better than PCA-derived variants for all kNN (82.8% vs. 50.8%), RF (83.0% vs. 52.5%), SVM (86.4% vs. 54.7%), e.g. with training size 45. Additionally, both GPT variants selection and ancestry predictions come with CoT reasoning steps, which provides interpretable results. Conclusions: Our results highlight the potential of LLMs for knowledge-driven phenotypes inference. For future work, we are expanding the framework with retrieval augmented generation and tool augmentation to tackle class biases, and we will also test its generalizability on cancer and Alzheimer's disease related phenotypes.

Session Title: Omics Technologies Poster Wednesday Session

Board 1143W: InFuse: a comprehensive framework for novel isoform and gene fusion detection from long-read RNA sequencing ⭐

Authors:

M. Ahsan[1], K. Rybacki[2], K. Wang[3]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]CHOP, Philadelphia, PA

Abstract:

Gene fusions and exon-skipping events are a major driver of diseases such as cancer and neurofibromatosis. Detection of functionally relevant gene fusions and novel transcript isoforms often requires transcriptome sequencing of effected tissue instead of genome sequencing. In particular, instances of exon-skipping that are not caused by structural variants may only be detectable through transcriptome sequencing. Long-read sequencing can assess full-length mRNA transcripts which allows for complete reconstruction of transcripts and more accurate read mapping in low complexity regions compared to short-read sequencing. However, full characterization of reads in terms of exons and detection of novel or aberrant transcripts, such as gene fusions and exon-skippings, are often hindered by alignment artifacts and the presence of multiple known isoforms for each gene. Currently, several long-read methods exist that tackle the detection of gene fusions and novel or aberrant transcripts individually, which can compound errors from multiple software. Here we present InFuse, a comprehensive framework for detection of novel isoforms and gene fusions from long-read RNA sequencing. InFuse uses a strand-specific interval tree representation of the transcriptome to characterize rapidly and fully each read in terms of the genes, transcripts and exons that it aligns to. This allows simultaneous detection of gene fusions, transcript readthroughs, novel isoforms and aberrant transcripts. Evaluation on 27 samples from CHOP Fusion Panel shows that InFuse can detect more gene fusions with a substantially higher read support than existing methods such as JAFFA, LongGF and FusionSeeker. This includes two fusions that were previously verified by clinicians as oncogenic: *STIL*::*TAL1* and *IGH*::*CRLF2* (fusion with under-annotated complex cluster), that other tools were not able to detect. InFuse was also able to detect and describe the structure of a complex exonic rearrangement in *EGFR* gene, consisting of exon-skipping and insertion, in one patient sample that had also been previously verified as oncogenic. In summary, InFuse allows a comprehensive characterization of the structure of novel isoforms and gene fusions from long-read RNA-seq, and can assist clinicians in interpretation of aberrant transcripts from patient samples.

Session Title: Omics Technologies Poster Wednesday Session

Board 1144W: Integrated Analytical Biobanking as a foundational component of NIH research in Neuroscience

Authors:

M. Sheldon; Sampled, Piscataway, NJ

Abstract:

Since 1999, SAMPLED, formerly known as RUCDR Infinite Biologics®/ Infinity BiologiX, has been awarded grants and contracts to operate the genetics resources for several agencies of the NIH, including the NIMH Repository & Genomics Resource, NIDA Center for Genetic Studies, NIDA Adolescent Brain Cognitive Development (ABCD) and HEALthy Brain and Child Development Study (HBCD) Studies and the NINDS Human Cell and Data Repository.These large-scale multinational efforts involve the collection, laboratory processing, analysis and storage of human biospecimens of many types, including blood, saliva, urine and feces, to name a few. The research projects funded by these institutions touch on numerous important areas of major public health concern, including psychiatric illnesses (e.g, Schizophrenia, Bipolar Disorder, Depression), addiction and substance abuse, neurologicaldisorders (e.g., Parkinson's, ALS, FTD) and human neurodevelopment (ABCD and HBCD). It is therefore of utmost importance to adhere to the highest standards in the collection and processing of these precious samples that result from the participation of hundreds of thousands of affected individuals. The facilities, infrastructure, expertise and training required to support these genetic resources are considerable. To this end, over the years SAMPLED has extended the conventional model of the biorepository as a facility for the deposit, storage and withdrawal of biospecimens to establish the world's first Global Integrated Analytical Biorepository (GIAB). The GIAB offers services that support large scale research or clinical projects from end to end, from sample collection kit assembly and shipping, sample accessioning, processing, multiomic analysis, to storage and global scale distribution with fully compliant cold chain logistics. This presentation will focus on the services, biospecimens, clinical and genomic data that are available to the neuroscience community through their auspices, with particular emphasis on the NINDS and NIMH resources managed by SAMPLED.

Session Title: Omics Technologies Poster Wednesday Session

Board 1145W: Integrating genomics, transcriptomics, and epigenomics for rare disease diagnosis

Authors:

S. Berger[1], J. Fraser[2], E. Delot[3], UCI-GREGoR, E. Vilain[4]; [1]Childrens Natl. Res. Inst., Washington, DC, [2]Children's Natl. Hosp., Washington, DC, [3]Children s Natl. Res. and Innovation Campus, Washington, DC, [4]Univ. of California, Irvine, Irvine, CA

Abstract:

Non-coding missense variants are challenging to classify with ACMG guidelines due to the limited data to predict functional effects, even more so for variants found in patients with

nonspecific phenotypes including neurodevelopmental disorders. To address this, the Pediatric Mendelian Genomics Research Center, part of the GREGoR Consortium, evaluates patients with high clinical suspicion of underlying genetic disorders with previous nondiagnostic clinical sequencing and uses multi-omic approaches to reach clinical diagnoses. Participants receive short-read genome sequencing, performed on trios when available, whole blood transcriptome sequencing, and long-read genome sequencing with methylation calling. To date, 814 samples have short-read genome sequencing, 336 RNA sequencing, and 105 long-read sequencing (using either Nanopore or PacBio methodologies). This includes 72 samples with data from all three and 286 samples with two datasets. Using phenotype-driven approaches, we integrate evidence across the datasets, identifying variants not identified by clinical sequencing and assessing their potential for pathogenicity. In one exemplar case, we integrated from trio short-read whole genome sequencing, whole blood transcriptomics, and genome-wide epigenomics from ONT long-read whole genome sequencing to reach a diagnosis of Coffin-Siris syndrome (CSS1) undetectable by clinical exome sequencing. This was a 7-year-old girl with neurodevelopmental delays and dysmorphic features referred after negative microarray, exome sequencing, biochemical evaluations, and karyotyping. Manual evaluation of short-read genome sequencing on the proband and parents identified a *de novo* deep intronic variant in *ARID1B*, ARID1B:c.3025+700C>G, classified as a variant of uncertain significance, with weak predictors of splice acceptor gain (0.31 at -143bp by spliceAI) and splice donor gain (0.19 at -5bp by spliceAI). Whole blood transcriptome demonstrated inclusion of a new exon predicted to introduce a premature stop codon. Using the genome-wide methylation calls from ONT, we compared the methylation at sites reported as differentially methylated in CSS1 for the patient's sample to 48 samples without suspected CSS1. Through permutation testing we demonstrated the patient's sample significantly ($p<0.001$) matched the CSS1 episignature confirming the diagnosis of Coffin-Siris Syndrome. Given the challenges of variant classification, particularly for noncoding missense variants, integration of multiple orthogonal omics techniques within standardizable pipeline is important for driving clinical rare disease diagnoses.

Session Title: Omics Technologies Poster Wednesday Session

Board 1146W: Integration of proteomics and CpG methylation sites identifies novel biomarkers and biology of early-stage heart failure

Authors:

K. Kottilil[1], M. Ramaker[1], N. Cauwenberghs[2], M. Nguyen[1], L. Kwee[1], S. Short[3], S. Plowman[3], K. Mahaffey[4,5,6], F. Haddad[4,5], S. Shah[1,7]; [1]Duke Molecular Physiology Inst., Durham, NC, [2]Res. Unit Hypertension and Cardiovascular Epidemiology, Dept. of Cardiovascular Sci., Univ. of Leuven,

Leuven, Belgium, [3]Verily Life Sci., LLC, South San Francisco, CA, [4]Div. of Cardiovascular Med., Stanford Univ. Sch. of Med., Stanford, CA, [5]Stanford Cardiovascular Inst., Stanford Univ., Stanford, CA, [6]Stanford Ctr. for Clinical Res., Dept. of Med., Stanford Sch. of Med., Stanford, CA, [7]Duke Clinical Res. Inst., Duke Univ. Sch. of Med., Durham, NC

Abstract:

Early-stage heart failure (HF), also termed Stage B HF, is the presence of structural heart abnormalities in individuals with HF risk factors (Stage A HF) prior to HF symptom onset. Identification of Stage B HF can aid in risk stratification; a circulating biomarker could have significant clinical utility for this purpose. Here, we aim to characterize novel circulating biomarkers of Stage B HF in the Project Baseline Health Study (PBHS). The study population consisted of 503 individuals with mass spectrometry-derived proteomics (289 proteins), microarray-derived methylation measurements, and echocardiograms at enrollment. HF staging was performed using American Heart Association (AHA) HF staging guidelines and echo phenotypes of cardiac structure and function. Univariate logistic regression was performed on Stage B HF cases (N=185) and Stage A HF controls (N=318); significant proteins ($p<0.05$) were used as inputs in elastic-net penalized regression models adjusted for HF risk factors. Of 289 proteins, 68 were associated with Stage B HF and input into the elastic-net model. The final elastic-net model contained 30 proteins and had good discrimination for Stage B HF (AUC 0.70, 95% CI [0.59, 0.81]). Top 5 weighted proteins included beta 2 microglobulin (B2M), EFEMP1, CRTAC1, GC, and CAMP. Four previously published protein quantitative trait methylation loci (pQTMs) were identified for the top proteins. These pQTMs were tested for association with Stage B HF, adjusting for age, race, sex, smoking status, and cell count proportions. One pQTM of B2M (*cg08099136*) was significantly associated with Stage B HF after covariate adjustment (FDR $P<0.05$). B2M is an inflammatory marker upregulated in cardiac tissue post ischemic damage, and *cg08099136* is known to disrupt immunoproteosome expression. This sheds light on potential dysregulation of inflammation and protein degradation in asymptomatic cardiac maladaptation. Overall, our findings identify novel molecular features with potential as therapeutic targets for early disease prevention in an intermediate phenotype of HF not studied at the molecular level.

Session Title: Omics Technologies Poster Wednesday Session

Board 1147W: Integrative Machine Learning Approaches for ASD Prediction using Gut Microbiome Analysis

Authors:

N. Ennajih, F. Annassiri, A. Moussa; Abdelmalek Essaadi Univ., the Natl. Sch. of Applied Sci., Tangier, Morocco

Abstract:

Applying machine learning techniques is a reliable approach to aid in the diagnosis of complex diseases. Recent studies have linked gut microbiota composition to the presence of autism spectrum disorder (ASD). However, despite a corresponding number of studies, it remains difficult to identify typical dysbiosis signatures in ASD patients.
This work proposes to use machine learning to investigate the composition of the gut microbiome and its role in distinguishing ASD from healthy controls. We collected 286 16S rRNA data samples from the gut microbiota of ASD and healthy controls. To this end, we used four machine learning algorithms, including random forest, XGBoost, logistic regression, and KNN. The models were evaluated in terms of AUC and predictive accuracy for identifying microbial biomarkers for ASD using a human metagenomic dataset. Consensus biomarkers were generated as potentially differentially abundant microorganisms.
Our research supports the understanding of the microbiome's role in the development of ASD and can help in the early diagnosis of children with ASD when using 16S rRNA data from the gut microbiome.

Session Title: Omics Technologies Poster Wednesday Session

Board 1148W: Isolation of single nuclei from cryosections of human cornea

Authors:

A. Wysocka[1], K. Jaskiewicz[1], M. Udziela[2], J. P. Szaflik[2], M. Gajecka[3,1]; [1]Inst. of Human Genetics, PAS, Poznan, Poland, [2]Med. Univ. of Warsaw, Warsaw, Poland, [3]Poznan Univ. of Med. Sci., Poznan, Poland

Abstract:

Background/objectives: Single-nuclei approaches paired with high-throughput sequencing techniques provide insight into the tissue's heterogeneity at single-cell resolution, and constitute the tool to study complex diseases. To perform such assessments, intact and debris-free nuclei should be provided, which may be daunting, as the selection of isolation procedure depends mainly on tissue type. Cell cultures and fresh, or snap-frozen tissues are being used as input material in many protocols. However, there are many techniques in which O.C.T.-embedded tissues are exploited but the protocol towards single nuclei isolation from those

materials has not been optimized yet. Here we present the approach to isolate nuclei of the human cornea cryosections. **Methods:** Cryosections of O.C.T.-embedded human corneas constituted the evaluated material. First, the cornea section's thickness and quantity were selected. Subsequently, we combined the chemical-, enzymatic- and mechanical-based methods for cryosection's dissociation to obtain a single-cell suspension, followed by the isolation of single nuclei, with the intact nuclear membranes. The modified workflow was based on a protocol for complex tissue (Demonstrated Protocol, CG000375, 10xGenomics). The quality control was performed using light microscopy, and acridine orange (AO)/propidium iodide (PI) counterstaining under fluorescence microscopy. The number of nuclei was counted using a manual hemocytometer. **Results:** We have established the workflow aimed at obtaining suspensions of single nuclei derived from corneal cryosections. A pool of 50 µm thickness cryosections was used per sample. The AO stains all cells, while PI which can enter the cell only upon disruption of the membrane, is a nuclei-specific dye, which allows the elimination of the cell debris from the microscopic image background. **Conclusion:** Our approach allows the use of cryosections of valuable tissue in the case of very limited biological material to obtain suitable input for single-nuclei-based analyses. **Grants:** Supported by the National Science Centre in Poland, grant no. 2021/41/B/NZ5/02245

Session Title: Omics Technologies Poster Wednesday Session

Board 1149W: Learning a sequence motif syntax from Transformer model attributions with SEMANTIC

Authors:

B. Slaff, N. Islas, Y. Barash; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Machine learning models that predict from DNA or RNA sequence show great promise for elucidating molecular mechanisms underlying genetics. Such models have been trained for a variety of tasks such as classification of regulatory regions and predicted levels of transcription factor binding, splicing, and expression. To derive biological insights from trained models, methods such as Enhanced Integrated Gradients (EIG) and Deep SHAP identify which RNA/DNA sequence features are most useful for making predictions. However, there remains a gap between identifying important sequence elements in a given sample and higher-level biological understanding of gene regulation. Specifically, in the context of the work presented here, we would like to move from individual nucleotides or short sequences (k-mers) to common groups of features (e.g. sequence or structure motifs), with logic (e.g. positional bias, motif

combination) which could then be connected with regulatory elements such as transcription or splice factors for condition-specific regulation.
To solve this problem we developed SEMANTIC, a computational pipeline which uses attribution scores from deep learning model predictions on DNA or RNA sequences to discover networks of interrelated sequence motifs and their syntax. As input, SEMANTIC takes sequence features and their corresponding importance scores. SEMANTIC then takes a data-driven approach to learn a minimal set of distinct, high-information motifs together with their co-occurrence, order, and distance patterns. In this way, SEMANTIC can be viewed as an attribution compression method that identifies minimal “networks” of co-occurring motifs. We first validated SEMANTIC by showing that it recovers the expected motif syntax from models trained on sequences that include simulated motifs. We then apply SEMANTIC to attributions from machine learning models trained on unperturbed DNA and RNA sequences and show that the recovered motif syntax recapitulates known biology and motivates new hypotheses.

Session Title: Omics Technologies Poster Wednesday Session

Board 1150W: Leverage generative GNN model to explore cellular heterogeneity in Perturb-seq data

Authors:

M. Yu; Yale Univ., New Haven, CT

Abstract:

The complexity of cell types in multicellular organisms arises not from a vast number of genes but from the combinatorial expression of a relatively small number of components. Understanding how gene regulation governs cell states is a fundamental question in molecular biology. Perturb-seq, which combines single-cell RNA sequencing with pooled CRISPR screens, has significantly advanced our understanding of genetic regulation of cell identity. However, despite experimental advancements, computational approaches are crucial for predicting perturbation effects due to the combinatorial explosion of potential gene perturbations.Existing computational methods, such as Compositional Perturbation Autoencoder (CPA), CellOracle, and GEARS, employ unique strategies to predict gene perturbation outcomes. However, these methods have limitations in leveraging prior knowledge and accurately modeling the heterogeneity of cell states. To address these gaps, we propose a multi-modal graph attention network to simulate gene expression levels and dynamics, capturing complex interactions within gene regulatory networks (GRNs). Additionally, we incorporate a diffusion-like model to

effectively represent cell state heterogeneity, enhancing our predictive capabilities.Our project has three primary aims:

1. Leverage existing single-cell multi-omic data to accurately represent the base GRN. We plan to integrate single-cell ATAC-seq and RNA-seq data to map biologically plausible transcription factor-gene connections, establishing a robust GRN adaptable to various cellular contexts.
2. Employ a multi-modal graph neural network to simulate perturbation effects on the GRN. Using a graph attention mechanism, we aim to refine network structures dynamically, improving the reliability of predictions across different cellular contexts.
3. Apply a generative model to capture cell state heterogeneity upon perturbations. By leveraging advances in generative modeling, we aim to infer distributions of possible cell states using Langevin dynamics, converting predictive models into generative ones.

We have obtained preliminary results demonstrating the feasibility of our approach and are exploring various models to optimize our predictions. Our ongoing work aims to refine these models to improve accuracy and robustness in predicting cellular responses to genetic perturbations.

Session Title: Omics Technologies Poster Wednesday Session

Board 1151W: Long Reads and Short Reads and Methylation Oh My: Options for low input/FFPE samples

Authors:

J. Gearhart, B. Marosy, M. Kokosinski, J. Paschall, K. Hetrick, P. Zhang, H. Ling, M. Mawhinney, K. Doheny; Ctr. for Inherited Disease Res. (CIDR), JHU, Baltimore, MD

Abstract:

Genomic DNA derived from Formalin Fixed Paraffin Embedded tissue is a common source for genomic studies but challenging for library prep and sequencing due to fixation damage, fragmentation and low mass. In response to the growing requests for epigenetic studies we recently evaluated an alternative chemistry for methylation detection using short read whole genome sequencing which provided a solution for low input, fragmented FFPE DNA sources. If wanting to acquire both WGS and methylation data for a tumor sample, long read sequencing using Oxford Nanopore would be ideal, given the capacity to acquire methylation information simultaneously with the nucleotide bases without any additional library prep or cost. Due to low

available quantities of DNA and fragmentation this technology is out of reach for most FFPE samples. Short read whole genome sequencing alone with bisulfite or enzymatic conversion would require two library preps and two sequencing runs in order to discern methylation status in addition to base calls. Pairing short-read WGS with a methylation array would be another option, however, the array requires a higher input of DNA material and the damaged FFPE material often does not yield useable array data. The BioModal evoC library prep method provides an alternative chemistry that would allow for standard whole genome sequencing with a low input requirement (<80ng) for DNA and simultaneously providing the distinction of 5mC/5hmC sites. Performance of this method (BioModal/NovaSeqXplus; 10B) was compared with our existing long read sequencing on Oxford Nanopore platform (LSK114; R10.4) and short read sequencing on Illumina NovaSeqXplus platform (Illumina DNA pcr free Prep; 10B) compared to GIAB. Using NA24385, recall and precision of SNPS in non-difficult genomic regions was 0.999013 and 0.999213 for BioModal; 0.999373 and 0.999477 for NovaSeqXplus; and 0.997486 and 0.998983 for Oxford Nanopore, respectively. In all regions (includes tandem repeats, homopolymer runs), recall and precision of SNPS was 0.981316 and 0.995854 for BioModal; 0.997173 and 0.998 for NovaSeqXplus; and 0.99634 and 0.994632 for Oxford Nanopore, respectively. Methylation recall and precision were determined from methylated lambda and non-methylated pUC19 spike in controls. NA24385 recall was 96.47% and precision was 99.6305 for modC calls. From our initial analyses, the low input evoC library prep method is a promising alternative for FFPE derived DNA in epigenomic studies.

Session Title: Omics Technologies Poster Wednesday Session

Board 1152W: Long-read RNA sequencing reveals isoform diversity in human immune cells and implications for diseases

Authors:

J. Inamo[1], A. Suzuki[2], M. Takahashi Ueda[1], K. Yamaguchi[1], H. Nishida[3], K. Suzuki[4], Y. Kaneko[4], T. Takeuchi[4], H. Hatano[2], K. Ishigaki[2], Y. Ishihama[3], K. Yamamoto[2], Y. Kochi[1]; [1]Tokyo Med. and Dental Univ., Tokyo, Japan, [2]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [3]Kyoto Univ., Kyoto, Japan, [4]Keio Univ., Tokyo, Japan

Abstract:

Recent advances in long-read RNA sequencing technology have enabled comprehensive characterization of full-length transcripts.Here, we applied this technology to profile the transcriptomes of 29 human immune cell subsets, uncovering an unprecedented level of isoform diversity. We identified 159,369 isoforms, over half of which were previously

unannotated, and compiled them into a database named TRAILS (Transcriptomic Resource of Immune cells using Long-read Sequencing). Intriguingly, many novel isoforms contained insertions of transposable elements, highlighting their role in shaping the transcriptional landscape of the human genome, particularly in immune-related genes. Furthermore, leveraging TRAILS, we identified isoforms associated with autoimmune diseases and neurological disorders. For example, an altered balance of specific *SIGLEC10* isoforms was linked to rheumatoid arthritis, while a potentially pathogenic novel read-through isoform from *APOE-TOMM40* locus was discovered in an Alzheimer's disease-associated gene. Utilizing TRAILS with genotype data revealed instances where SNPs modulate isoform expression, altering disease risk.Our study provides a valuable resource for understanding the diversity and functional specialization of human immune cells at isoform resolution, with broad utility for elucidating disease mechanisms and identifying therapeutic targets.

Session Title: Omics Technologies Poster Wednesday Session

Board 1153W: Massively parallel single-molecule protein quantitation using protein identification by short-epitope mapping (PrISM)

Authors:

J. Egertson[1], T. Gillies[1], M. Newman[1], V. Badamagunta[1], G. Bellesia[1], R. Grothe[1], K. Winters[1], J. Robinson[1], B. Nortman[1], M. Villancio-Wolter[1], J. Joly[1], S. Tonapi[1], S. Tan[1], I. Sprague[2], H. Boyce[3], **N. Nelson**[4], K. Lazaruk[1], M. Jouzi[1], S. K. Wilcox[1], T. Rinker[1], K. Juneau[1], P. Mallick[1]; [1]Nautilus Biotechnology, San Carlos, CA, [2]Nautilus Biotechnology, Seattle, WA, [3]Boyce Data Insights, The Hague, Netherlands, [4]Nautilus Biotechnology, San Diego, CA

Abstract:

Introduction: Massively parallel next-generation DNA sequencing improved access to genomic data by increasing the scale by which DNA could be read. To-date proteomic analyses have been severely limited in scale. Analyzing protein samples at single-molecule resolution promises to achieve improved dynamic range, reproducibility, and sensitivity. Using multi-affinity probes to interrogate immobilized, intact proteins, we aim to use Protein Identification by Short-epitope Mapping (PrISM) to analyze >95% of the proteome with broad dynamic range. PrISM has the potential to revolutionize multi-omics research by delivering highly accurate, large-scale proteomic analyses.
Methods: Single-molecule protein libraries of control protein mixtures and U2OS cells were prepared. Next, single-molecule protein libraries were deposited on a high-density flow cell at optically resolvable locations. PrISM utilizes non-traditional affinity reagents, multi-affinity

probes, designed with high affinity and low specificity to bind short epitopes present in multiple proteins in the proteome. Multi-affinity probes generate binding patterns for each protein molecule, which are translated to protein quantities using machine learning. We acquired PrISM data on U2OS cells, protein mixtures, and control samples.
Results: We demonstrate quantification of proteins from control mixtures and from the U2OS cell line using PrISM . The quantified proteins span a wide dynamic range of abundance. We also compare PrISM protein identification to mass spectrometry analysis. The data demonstrates improved quantitative performance defined by sensitivity, accuracy, reproducibility, and dynamic range. An in-depth analysis of binding data shows a high diversity among patterns generated and indicates an opportunity to continue to identify more proteins by adding more multi-affinity probes.
Conclusions: The PrISM approach has resulted in protein detections in a human cell line sample at single-molecule resolution. With this capability PrISM can generate unique insight into the proteome and human biology.

Session Title: Omics Technologies Poster Wednesday Session

Board 1154W: Measuring enhancer-enhancer cooperativity using single-cell CRISPR perturbations and single-cell multiome sequencing data

Authors:

K. Guruvayurappan[1], J. Zhou[2], R. Luo[1], S. Toneyan[2], N. Hu[1], H. V. Chen[3], A. R. Chen[4], P. K. Koo[2], D. Huangfu[1], K. Dey[1], G. McVicker[5]; [1]Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [3]Agency for Sci., Technology, and Res. (A*STAR), Singapore, Singapore, [4]Stony Brook Univ., Stony Brook, NY, [5]Salk Inst. for Biological Studies, La Jolla, CA

Abstract:

Enhancers are short genomic sequence elements that regulate transcription. A single gene may be regulated by multiple enhancers, and in-depth characterization of enhancers at individual gene loci indicates that enhancers regulate gene expression in a non-linear manner. To interrogate non-linear enhancer cooperativity genome-wide, we developed a statistical framework using generalized linear models that evaluates the regulatory effects of enhancer pairs on a common target gene and quantifies enhancer cooperativity through an interaction term. We applied these models to two different data modalities capturing enhancer cooperativity - a multiplex CRISPR enhancer-gene screen and single-cell RNA+ATAC multiome data.

We first reanalyzed a single-cell CRISPR experiment in K562 cells that delivered random combinations of enhancer-targeting perturbations to each cell and measured cooperative effects on gene expression for 46,166 candidate enhancer pairs. We observed that enhancer effects generally combine in a log-additive model without cooperativity, except for 31 enhancer pairs with significant interaction effects (FDR < 0.1). 20 of these signals were detected as outliers and the remaining 11 interactions provided limited biological evidence for cooperativity. However, our CRISPR analysis was limited by two experimental factors: the low number of cells with two enhancers for the same gene perturbed (median=12) and high multiplicity of infection (MOI) (median=28 guides/cell).
We next applied our approach to single-cell multiome data by associating pairs of chromatin accessibility peaks with gene expression. We reasoned that multiome data can detect enhancer cooperativity in cell types where CRISPR data is limited. Since enhancer cooperativity has been observed in differentiation, we applied our models to a multiome dataset of 7,395 definitive endoderm cells, which are the second stage of embryonic differentiation to pancreatic beta cells. We observed significant enhancer cooperativity between 183 out of 9,233 tested enhancer pairs. Enhancers with cooperative associations were 2.2x enriched for fine-mapped causal variants from genome-wide association studies across 94 UK BioBank diseases and traits. Genes displaying enhancer cooperativity showed enrichment in developmental pathways such as endoderm differentiation; one notable such gene was *SOX17*, a transcription factor known to influence cell fate determination and differentiation. In summary, our proposed model reveals that most enhancer pairs follow a log-additive model of gene regulation, but we also identify cases of enhancer cooperativity in differentiation-related genes.

Session Title: Omics Technologies Poster Wednesday Session

Board 1155W: Medical Action Ontology (MAxO) enables prioritization of diagnostic procedures to support rare-disease differential diagnosis

Authors:

M. Beckwith[1], L. Carmody[1], D. Danis[2], M. Gargano[1], P. Robinson[2]; [1]The Jackson Lab., Farmington, CT, [2]Berlin Inst. of Hlth.at Charite, Berlin, Germany

Abstract:

Over 10,000 rare diseases have been identified to date, collectively affecting between 3.5% and 8% of the population, yet many patients experience a long diagnostic odyssey of 5-7 years. Next-generation sequencing technologies (NGS), including exome sequencing and genome sequencing, have dramatically improved the accuracy and utility of genetic testing, but the

overall diagnostic yield still has not exceeded 50% in most large cohort studies. Therefore, supplementary approaches are needed to continue to improve our ability to identify the correct diagnosis in individuals with rare genetic disease. The Medical Action Ontology (MAxO) is a tool for annotating and analyzing treatments, diagnostic modalities, and other actions taken for the clinical management of human disease. The diagnostic procedure (MAXO:0000003) branch is the second largest branch of MAxO and currently contains 774 terms, such as magnetic resonance angiography procedure (MAXO:0035088). MAxO is tightly integrated with the Human Phenotype Ontology (HPO), which has become the de facto standard for deep phenotyping in human genetics and related fields and is an integral component of nearly all academic and commercial software for diagnostic genomic sequencing for individuals with rare diseases. Currently, there are 18,505 diagnostic annotations that specify which diagnostic procedures are able to identify 16,389 unique HPO terms. Here, we introduce a mathematical model to estimate the quality of a differential diagnosis and to provide a metric for the expected change brought about by a diagnostic procedure represented by a MAxO term. The model leverages the over 150,000 HPO annotations for roughly 8500 rare genetic diseases to investigate whether a diagnostic test is likely to change the order of the current rankings of a differential diagnosis generated by tools such as LIRICAL or Exomiser, in order to suggest to clinicians which investigation is likely to be the most useful and why. We provide an HTML web interface that can be used to run the maxodiff algorithm.

Session Title: Omics Technologies Poster Wednesday Session

Board 1156W: Meta-analysis of RNAseq data from two large atrial tissue repositories identifies more than 8000 transcripts associated with atrial fibrillation and suggests a prominent role for mitochondrial ADP/ATP processing: the CATCH ME and RACE V Consortia

Authors:

A. Isaacs[1], J. Winters[1], S. Zeemering[1], M. Kawczynski[1], M. Gilbers[2], B. Maesen[3], J. Maessen[3], E. Bidar[3], B. Boukens[1], B. Hermans[3], A. van Hunnik[1], B. Casadei[4], L. Fabritz[5], W. Chua[6], L. Sommerfeld[5], E. Guasch[7], L. Mont[8], M. Batlle[7], S. Hatem[9], P. Kirchhof[10], R. Wakili[11], M. Sinner[12], S. Kääb[12], S. Verheule[1], I. Van Gelder[13], M. Rienstra[13], M. Stoll[14], U. Schotten[1]; [1]Maastricht Univ., Maastricht, Netherlands, [2]Medisch Spectrum Twente, Enschede, Netherlands, [3]Maastricht Univ. Med. Ctr.+, Maastricht, Netherlands, [4]Univ. of Oxford, Oxford, United Kingdom, [5]Univ. Med. Ctr. Hamburg, Hamburg, Germany, [6]Inst. of Cardiovascular Sci., Birmingham, United Kingdom, [7]Inst. of BioMed. Res. August Pi Sunyer (IDIBAPS), Barcelona, Spain, [8]Univ.t de Barcelona, Barcelona,

Spain, [9]Sorbonne Univ., Paris, France, [10]Univ. Hosp. Hamburg Eppendorf, Hamburg, Germany, [11]Universitäres Herz- und Gefäßzentrum Frankfurt, Frankfurt, Germany, [12]Univ. Hosp. of Munich, Munich, Germany, [13]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [14]Univ. of Münster, Münster, Germany

Abstract:

Atrial fibrillation (AF) is the most common arrhythmia in adults and contributes to outcomes such as heart failure and stroke; despite this, many underlying molecular mechanisms remain to be elucidated. This analysis used the largest-scale atrial RNAseq dataset available to date to illuminate the AF transcriptome.

Two large atrial tissue biobanks, from the European CATCH ME and Dutch RACE V consortia, provided 202 right and 238 left atrial appendage biopsies. All underwent identical RNA extraction, selection of poly-A tailed transcripts, library preparation, sequencing, alignment, and analysis pipelines. Differential expression (DE) analyses were performed independently in each study, stratified by atrial side, and two comparisons were made in each stratum: paroxysmal AF (pAF) vs. no AF history (noAF) and persistent AF (persAF) vs. noAF. Results were meta-analyzed using random effects models and *P*-values were adjusted using the Benjamini & Hochberg method (P_{FDR}). Transcripts with a $P_{FDR} \leq 0.05$ and a concordant effect direction were considered significant.

In total, 24846 transcripts were analyzed after stringent filtering; 8497 showed evidence of DE in at least one comparison (n = 5255 for persAF vs. noAF and n = 1573 for pAF vs. noAF in right atria; and n = 4927 for persAF vs. noAF and n = 2365 for pAF vs. noAF in left atria). The majority of these (77.7%) were protein-coding, with most of the remainder (17.3%) long, non-coding RNA. Approximately 45% of DE transcripts were associated in more than one comparison, including 287 (3.4%) which differed in all four. Gene ontology (GO) term enrichment for these 287 yielded many enriched terms, including those related to mitochondrial membrane organization and outer membrane translocase complexes, Z-disc, cell-substrate junction, and actin and kinesin binding. Protein-protein interaction networks visualized for these molecules in STRING identified networks mostly centered around mitochondrial function and energy metabolism. The largest network contained the three genes with the highest connectivity (*VDAC1*, *TOMM40*, and *HSPA9*) and was focused on mitochondrial ADP/ATP handling. *TOMM40* and *VDAC1* encode β-barrel pores on the mitochondrial membrane; both are involved in ADP/ATP exchange. Other members of the network encode proteins in the mitochondrial translocase complex and ADP/ATP chaperones.

These large-scale RNAseq analyses dramatically expand the number of transcripts robustly associated with AF. They directly implicate mitochondrial ADP/ATP processing, either as a cause, or consequence, of AF. These results suggest the need for functional investigation of mitochondrial function as an AF mechanism.

Session Title: Omics Technologies Poster Wednesday Session

Board 1157W: Modification of an early barcoding bulk RNA-sequencing method to capture long non-polyadenylated transcripts

Authors:

R. Flamio[1,2], A. Nimoni[1], Y. Demirci[3], T. B. Niewold[1,2]; [1]Hosp. for Special Surgery, New York, NY, [2]Weill Cornell Med. Coll., New York, NY, [3]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Early barcoding, the addition of sample-specific tags during cDNA synthesis, has been successfully implemented in both single-cell and bulk RNA sequencing (RNA-seq) library preparation methods to reduce costs without compromising data quality. Reduced costs facilitate the testing of additional samples and the increase of study power. Prime-seq is an early barcoding bulk RNA-seq proven to be effective and affordable for mRNA expression analysis. However, the method is limited to polyadenylated transcripts due to reliance on oligo(dT) primers. Long noncoding RNAs (lncRNAs), many of which are not polyadenylated including circular RNAs (circRNAs), have been increasingly found to be important gene regulators in both health and disease, including systemic lupus erythematosus (SLE). The objective of the study is to modify the prime-seq protocol to capture long non-polyadenylated transcripts. Naïve B cells were isolated from the venous blood of three subjects and stimulated for 6 hours with interferon-α. RNA was extracted from 50,000 cells using the prime-seq bead-based isolation. Prime-seq libraries were prepared using three different protocols with pre-cDNA synthesis modifications: 1) standard prime-seq, 2) prime-seq with preparatory ribosomal RNA (rRNA) depletion and poly(A) tailing, and 3) prime-seq with preparatory rRNA depletion, poly(A) tailing, and 5’ capping. A fourth RNA-seq library was prepared using a commercial total RNA prep kit for comparison. DNA oligos were used to deplete rRNA while RNase H simultaneously degraded rRNA in rRNA:DNA hybrids and linearized circRNA. Poly(A) tailing allowed the capture of non-polyadenylated transcripts, which was further enhanced with 5’ capping. RNA and cDNA were tested for quantity and quality using a Qubit fluorometer and a 4200 TapeStation, respectively. Sequencing is in progress on a NovaSeq6000 system, but preliminary QC analyses are promising. Across protocols, omission of proteinase K digestion promoted RNA integrity, and cDNA was amplified at sufficient quality and quantity for sequencing. A custom bioinformatics pipeline will compare captured lncRNAs across different library preparation methods. We anticipate libraries prepared by rRNA depletion and poly(A) tailing will capture more diverse lncRNAs including those missed in standard prime-seq. 5’ capping may retain an even larger number of lncRNAs due to enhanced binding of the template

switch oligo during reverse transcription. Overall, modified prime-seq may represent an affordable and reliable method for studying differential gene expression and transcriptional regulation in diseases to which lncRNAs are suspected to contribute.

Session Title: Omics Technologies Poster Wednesday Session

Board 1158W: More accurate and faster expression and splicing methods improve DROP - the Detection of RNA Outliers Pipeline

Authors:

C. Mertes[1,2], V. Yepez[1], I. Scheller[1], F. Brechtmann[1], B. Andrew[1,3], S. Cristian[1], N. Smith[1], GHGA consortium, J. Gagneur[1,2]; [1]Computational Molecular Med., Technical Univ. of Munich, Garching, Germany, [2]Inst. of Humangenetics, Klinikum rechts der Isar, Munich, Germany, [3]Munich Data Sci. Inst., Technical Univ. of Munich, Munich, Germany

Abstract:

Background: RNA-seq is a powerful approach to prioritize candidate genes for individuals affected with a rare genetic disorder and increases diagnostic yield by approximately 15%. DROP has a fast growing user base worldwide and is getting implemented in clinical settings for RNA-seq-based diagnostics. Learning from its application in many large-scale projects, such as Solve-RD, UDN, GREGoR, Genomics England, and Bavarian Genomes, we have updated and improved DROP and integrated recently developed approaches, accelerating the detection of aberrant events in RNA-seq data.
Method: We benchmarked the optimal hard thresholding procedure to determine the optimal latent space dimension, introduced by Gavish and Donoho, and implemented it in OUTRIDER. We further integrated FRASER 2.0 for splicing analyses and extended QC measures, including a sex verification step. Additionally, we implemented a continuous and prompt support via GitHub Issues for our users.
Results: DROP accelerates the adoption of RNA-seq-supported diagnostics in human genetic centers and has demonstrated success in expediting rare disease diagnoses. With the new improved version of DROP, we were able to support the identification of 20 new disease causing variants in the Solve-RD cohort. On the technical level OUTRIDER's updates reduce the compute time for the expression outlier analysis by 15-fold. FRASER 2.0 uses the new intron Jaccard index that reduces the false positives by 8-fold while maintaining high sensitivity and simplifying the splicing analysis overall. By porting DROP to nf-core/Nextflow, we co-develop with the German Human Genome-Phenome Archive (GHGA) a containerized, scaleable, and cloud-ready workflow that is easy to deploy by its users.

Conclusion: The improvements in DROP, driven by large-scale projects and its integration into the nf-core ecosystem, significantly enhance DROP's applicability in RNA-seq-supported rare disease diagnostics. With robust maintenance and support through GHGA (283 closed issues), healthcare professionals and researchers can now more effectively utilize RNA-seq to identify disease-causing variants, thereby improving diagnostic outcomes and accelerating the discovery of novel gene-disease associations.
Grants: German Research Foundation (DFG) 441914366 (NFDI 1/1).

Session Title: Omics Technologies Poster Wednesday Session

Board 1160W: New Genome In A Bottle HG002 T2TQ100 diploid assembly-based small and structural variant benchmark sets.

Authors:

N. Olson[1], N. Dwarshuis[1], J. McDaniel[1], J. Wagner[2], J. Zook[2], Genome In A Bottle (GIAB), Telomere-to-Telomere (T2T); [1]NIST, Gaithersburg, MD, [2]Natl. Inst. of Standards and Technology, Gaithersburg, MD

Abstract:

Genome in a Bottle was started over ten years ago to develop well-characterized human genome reference materials for validating genome sequencing and variant calling methods. GIAB has produced genomic DNA reference materials and benchmark sets composed of high-confidence variant calls and regions confidently identified as homozygous reference or as high-confidence variants. These benchmark sets are widely used to evaluate and train variant calling methods. Traditionally, these benchmark sets are developed by integrating variant callsets generated using different sequencing technologies and variant calling algorithms. However, this process involves mapping reads to a reference, which is limited in its ability to characterize more complex genomic regions. With recent advances in sequencing methods and genome assembly algorithms, creating high-quality diploid de novo genome assemblies is possible. A subgroup of the T2T consortium that generated the first complete human genome assembly, CHM13, is working on a highly polished and curated T2T diploid genome assembly of GIAB HG002 (son of the Ashkenazi Jewish trio.) While previous GIAB benchmark sets have used diploid assemblies for particular genomic regions, we use the HG002 Q100 T2T diploid assembly to generate the first whole genome assembly-based small and structural variant benchmark sets. These benchmark sets were generated using a new Snakemake-based pipeline and include regions of the genome that were excluded when using read-mapping-based benchmark generation methods, e.g., the highly polymorphic MHC, gene conversions, and more complex SVs. The

small variant benchmark set includes ~ 89.8% (2.74 Gbp) of the HG002 assembly with >3.6 million SNVs and ~950,00 indels (>50bp), over 700,000 more variants than the previous v4.2.1 benchmark. The structural variant benchmark set covers nearly 90.3% (2.75 Gbp) of the HG002 assembly and nearly 30,000 variants. This represents a significant improvement over the previous v0.6 HG002 SV benchmark, which covered only 2.51Gbp and included ~10k variants. Draft versions of the benchmark sets are publicly available and undergoing external validation before official release.

Session Title: Omics Technologies Poster Wednesday Session

Board 1161W: New quantitative and qualitative analytical framework of scRNAseq data reveals the pathophysiology of systemic lupus erythematosus ⭐

Authors:

M. Nakano[1], M. Kono[1,2], H. Hatano[1], K. Asahara[1], T. Nishino[1], H. Takahashi[1,2], B. Natsumoto[1], K. Ishigaki[1,2,3]; [1]Lab. for Human Immunogenetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [2]Dept. of Microbiol. and Immunology, Keio Univ. Sch. of Med., Tokyo, Japan, [3]Keio Univ. Human Biology-Microbiome-Quantum Res. Ctr. (WPI-Bio2Q), Tokyo, Japan

Abstract:

Systemic lupus erythematosus (SLE) is a complex autoimmune disease with unknown etiology. While we previously identified candidate key gene signatures of SLE using bulk RNAseq from 27 major immune cell types (*Cell* 2022), we failed to identify granular disease-relevant cell states within each cell type. Although single-cell RNA-seq (scRNAseq) is an ideal tool to address this problem, most of the recent single-cell studies ended up profiling only ~20 coarse cell types and depend on cell-type level pseudobulk analysis, which cannot distinguish quantitative change (e.g., differential expression) and qualitative change (cell state abundance within each cell type). To overcome these limitations, we developed a new analytical strategy to efficiently dissect cell state abundance (quantitative change) and gene expression signature (qualitative change) in patients at a single-cell level. We applied this framework to the largest-scale scRNAseq data of SLE with ~1.4 million cells from 269 donors. After we successfully identified 95 granular cell states in 24 cell types using the fine clustering method, we revealed that previously uncharacterized cell states such as HLA^{hi}GZMK^{hi}cytotoxic CD8^{+} T cells and C1Q^{hi} nonclassical monocytes expanded in SLE patients, especially in severe phenotype (Bonferroni-corrected global $P < 0.05$ in co-varying neighborhood analysis). Next, our new single-cell level differential expression model successfully dissected ~2,000 true differentially expressed genes (FDR < 0.05)

for each cell type (qualitative change) without being biased by cell state abundance (quantitative change). Intriguingly, while signature genes reflecting quantitative and qualitative change showed different biological processes (e.g., pathway enrichment), stratified LD score regression revealed that polygenic signals were enriched in both signature genes, highlighting the diverse pathogenic functions of GWAS variants. Our findings can be a foundation for deeper characterization of disease-relevant cell states and signature genes of SLE and contribute to a better understanding of pathogenesis in complex diseases.

Session Title: Omics Technologies Poster Wednesday Session

Board 1162W: Omics Wide Association Analysis Platform (OWAS) to identify personalized genetic risk in CVD

Authors:

A. Jha[1], **Y. Singh**[2], N. Sharma[3], K. Raghuvanshi[4], J. Kwan[5]; [1]Weill Cornell Med., NY, NY, [2]Khan Labs, Santa Clara, CA, [3]The Harker Upper Sch., San Jose, CA, [4]Khan Lab High Sch., Mountain View, CA, [5]Yale, New Haven, CT

Abstract:

Various platforms have been developed to simplify the process of differential expression analysis using RNA-seq data. However, very few platforms can perform pathway-specific differential expression and integrate that with Genome-Wide Association Analysis (GWAS) results. Here, we have developed OWAS (Omics Wide Association Analysis Platform) to integrate Transcriptome-Wide Association Study (TWAS) data from publicly available sources such as NCBI GEO for all cardiovascular diseases. OWAS identifies differentially expressed pathways, focusing on heart failure (HF) and coronary artery disease (CAD). In the first step, we combined all the CAD and HF datasets from NCBI GEO, applied pre-processing methods for batch effect removal, and calculated the effect weight for each gene by combining 18 studies containing 208 control and 419 heart failure samples. We used DEseq2 to calculate differentially expressed genes (DEGs) and compute Z-scores. Next, we identified the list of key pathways for HF and CAD and calculated the average effect size. We identified 12 pathways for HF and 19 pathways for CAD. Further, we examined GWAS data from the GWAS Catalog, employing a machine learning (ML) model to identify transcriptome-variant associations. The key factor here is to pinpoint regions of somatic mutations in GWAS data and identify impacted pathways as well as differentially changed pathways due to these mutations. This comprehensive approach allows for a deeper understanding of disease progression and treatment response in cardiovascular diseases (CVD). We have developed a Shiny app to access NCBI GEO and GWAS data, allowing researchers to

identify DEGs and associated pathways due to differential changes. The ongoing app is expected to incorporate a ML model to stratify the list of pathways for HF and CAD. In the subsequent version, we will add somatic mutation identification, pathways specific to somatic genes, and association with GWAS data, along with a list of differentially expressed pathways and possible drug target correlations. The app will allow Omics-wide association and provide a platform for translational research by identifying disease-causing pathways, the contribution of specific pathways to disease outcomes, and testing models for specific patients by uploading patient data and identifying causal pathways. This will enable the addition of Omics-based risk stratification in CVD. The Shiny app is easy to use with minimal user intervention, and further functionality will make it an essential tool for individual risk identification, somatic mutation burden assessment, and possible treatment and drug response planning in CVD.

Session Title: Omics Technologies Poster Wednesday Session

Board 1163W: Paired single-cell protein and RNA comparison in developing human brain reveals biological gap in current single-cell research

Authors:

T. Wu; UCSF, San Francisco, CA

Abstract:

The core focus of medical and biological research is to understand the mechanisms underlying protein changes. Researchers currently rely on single-cell RNA measurements to infer protein expression change, with the implicit notion that RNA and protein levels will exhibit a high degree of consistency. However, we provided that is an unsupported hypothesis. In our study, we apply optimized and highly sensitive mass spectrometry to profile protein in over thousands single cells in developing brains from three distinct human samples, we also conducted paired single-cell RNA sequencing for the same samples to perform a fair comparison in single cell resolution. Despite consistent expression trends in some key cell marker genes, like TNC for Radial glia, EOMES for IPC-EN and TBR1 for Excitatory Neurons, our comparative analysis revealed a low overall correlation between proteins and RNAs expression in developing brain (cor = 0.26). A further cell type specificity comparison is performed to remove the potential effect caused by cell type proportion, results again highlighted the gap between RNA and Protein data, the most inconsistency is from Microglia cells, with only 0.19, and the most consistent cell type is RG, with cor = 0.36. The gap suggests that current biological conclusions based on single-cell RNA inference may not align with actual cellular activities. We then

investigate in disease genes associated with Autism spectrum disorder, Tau evaluation revealed strong cell type-dependent expression patterns in disease genes' protein, contrasting with unspecific expression at the single cell RNA level (median Tau 0.856 vs 0.395), the discrepancy between RNA and protein in developing brain underscores the need of prompting directly single cell proteins profile in development and disease research in current single cell Omics Era to gather more accurate understanding of the true cellular state.

Session Title: Omics Technologies Poster Wednesday Session

Board 1164W: Plasma metabolites are associated with bone density, microarchitecture, and bone strength; The Framingham Heart Study

Authors:

P. Okoro[1], Z-Z. Chen[2,3], C-T. Liu[4], C. Clish[5], M. L. Bouxsein[6], A. B. Dufour[1,3], S. Sahni[1,3], Y-H. Hsu[1], D. Kiel[1,3]; [1]Marcus Inst. for Aging Res., Hebrew SeniorLife, Boston, MA, [2]Div. of Endocrinology, Diabetes, and Metabolism, Beth Israel Deaconess Med. Ctr., Boston, MA, [3]Dept. of Med., Beth Israel Deaconess Med. Ctr. and Harvard Med. Sch., Boston, MA, [4]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, Boston, MA, [5]Metabolomics Platform, Broad Inst., Cambridge, MA, [6]Dept. of Orthopedic Surgery, Beth Israel Deaconess Med. Ctr. and Harvard Med. Sch., Boston, MA

Abstract:

Age-related loss of skeletal integrity increases fracture risk due to deteriorations in bone density and microarchitecture. Circulating metabolites associated with cortical and trabecular bone density, microarchitecture, and strength (failure load) may provide insight into bone metabolism beyond genetic variation, as they amplify genetic variants governing molecular changes difficult to detect in genomics studies.

Here, we profiled 643 non-redundant circulating metabolites using liquid chromatography-mass spectrometry in plasma and identified their cross-sectional association with skeletal density, microarchitecture and strength - measured using high resolution imaging of the distal tibia and radius - in two Framingham Heart Study (FHS) cohorts. Controlling for age, BMI, alcohol, smoking, menopause, estrogen, glomerular filtration rate, diabetes, physical activity, and glucocorticoids, associations were modeled using multivariable linear regression with discovery in the FHS Offspring cohort (n=865, FDR<0.25) and replication in the FHS Generation 3 cohort (n=1016, P< 0.05, concordant direction of effect).

A total of 193 bone-metabolite associations were found in the discovery cohort, involving 117 distinct metabolites. Of these, 49 associations were replicated. Higher levels of ~30% of these

bone-associated metabolites were beneficial. Tibial failure load had the most metabolite associations (n=22), followed by trabecular bone mineral density (BMD, n=19). The replicated bone-metabolites were enriched for members of unsaturated fatty acids biosynthesis (q=9.56E-4), and branched chain amino acid (BCAA, valine, leucine, and isoleucine) biosynthesis pathways (q=0.09) in KEGG analyses. Findings include, the positive association of BCAA with BMD, which is consistent with previous reports which showed that degradation of BCAA is a contributor to fracture risk. We also found over 20 fatty acids including myristoleate, palmitoleate, oleate, that were deleterious to bone strength, suggesting the possibility that these fatty acids may alter osteoblasts functions.
Given that bone strength is the strongest predictor of incident fracture, we conclude that metabolites can serve as biomarkers and potential targets for the preservation of skeletal integrity. Additionally, the large number of associations that we found will allow us to nominate multiple metabolic pathways, after additional prioritization including assessment of causality with Mendelian Randomization, for further study to identify novel targets to improve bone health.

Session Title: Omics Technologies Poster Wednesday Session

Board 1165W: Plasma protein signature as easily accessible biomarkers to predict future MCI to dementia conversion

Authors:

J. Xi, Y. Zhang, B. Smets, S. Miller, Z. Saad, C. Whelan, S. Li; Johnson & Johnson Innovative Med., Spring House, PA

Abstract:

Background: Current available biomarkers for the diagnosis and progression of Alzheimer's disease (AD) are Amyloid beta and Tau, typically measured in brain by positron emission tomography (PET) or cerebrospinal fluid (CSF) by immunoassays. These methods are either expensive or invasive which causes operational challenges due to high screening failure rate. Therefore, there is a significant need for easily accessible approaches utilizing plasma biomarkers to capture mechanistic heterogeneity enabling the early identification of individuals at risk for AD. **Method:** Fundació ACE (FACE) is a memory clinic cohort based in Barcelona, Spain. A total of 755 patients with mild cognitive impairment (MCI) at baseline were followed up for the progression to dementia. Baseline plasma proteomics were profiled using Somalogic 7k protein panel. Univariate Cox proportional hazards regression was employed to identify proteins associated with MCI-to-dementia conversion. Random survival forest and gradient

boost were used to develop the prediction models. Three prediction models with different feature sets were constructed: Model 0 included only demographics features; Model 1 included only proteomics features; Model 2 included both demographics and proteomics features. Harrell's Concordance index (c-index) and time dependent Area under the ROC Curve (AUC) were used for model evaluation. **Result:** A total of 368 (49%) patients had converted from MCI to dementia with a mean follow-up of 2 years. 136 plasma proteins, enriched in extracellular matrix pathways, were significantly associated with MCI-to-dementia conversion (FDR<0.05). Model 0 (demographics only) achieved a c-index 0.67 on the test set, a median time dependent AUC of 0.71 and a max AUC of 0.75 at 5 years. Model 1, with 41 proteins, achieved a c-index of 0.67 on the test set, a median time dependent AUC of 0.75, and a max AUC 0.76 at 3.5 years. Model 2, with 39 proteins and demographics achieved the best performance on test set. This model showed a c-index of 0.68, a median time dependent AUC of 0.78, and a max AUC of 0.80 at 5 years. Bootstrapped test showed that adding protein signature to demographic-only model significantly (p-value < 0.05) improved the prediction performance. **Conclusion:** We identified protein signatures that can predict the near-future MCI-to-dementia conversion, suggesting their potential as early and easily accessible prognostic biomarkers, although further validation in independent cohorts is still required.

Session Title: Omics Technologies Poster Wednesday Session

Board 1166W: ProGenAC: Capturing deep phylogeny through the hyperbolic embedding space of a Variational AutoEncoder

Authors:

J. Lamanna[1], M. Mowlaei[2], P. English[3], X. Shi[2], V. Carnevale[1]; [1]Temple Univ., Philadelpia, PA, [2]Temple Univ., Philadelphia, PA, [3]Univ. of Utah, Salt Lake City, UT

Abstract:

Proteins from the same homologous family share a common evolutionary origin and constrain their sequence variability to intricate mutational patterns that can be learned by deep models. Tracing the evolutionary history of these proteins and building phylogenetic trees are challenging tasks because there are no known methods to precisely quantify the evolutionary distance between protein sequences. Commonly, the Hamming distance between sequences, or variants of it, is used to estimate evolutionary distance and build phylogenetic trees. However, these measures struggle with deep phylogenies by saturating with sufficient depth, losing all notion of relatedness between proteins. To infer the deep phylogenetic history present in homologous protein families, a method to better estimate the number of substitutions between

protein sequences is needed. Here, we present ProGenAC, a Variational AutoEncoder (VAE) equipped with adversarial training, a composite loss function, and hyperbolic latent space geometry, which enhances sequence generation and enforces a correlation with evolutionary distances in the embedding space. The distance between points in this space provides a more accurate estimate for the number of substitutions along a tree when compared to the Hamming distance. We demonstrate that ProGenAC outperforms state-of-the-art generative protein sequence models in various measures and produces evolutionarily meaningful low-dimensional representations of protein sequences. Our work paves the way for the rational design of protein sequences and potential ancestral sequence generation.

Session Title: Omics Technologies Poster Wednesday Session

Board 1167W: Protein quantitative trait locus analysis in African American and non-Hispanic White individuals

Authors:

Y. Cai[1], M. Johnson[1], J. Haessler[1], A. J. Molstad[2], S-J. Hwang[3], R. Joehanes[3], J. Murabito[4], U. Tahir[5], N. Franceschini[6], R. Gerszten[5], W. Sun[1], D. Levy[3], L. Raffield[6], C. Kooperberg[1], L. Hsu[1], A. P. Reiner[1]; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Univ. of Minnesota, Minneapolis, MN, [3]Framingham Heart Study, Framingham, MA, [4]Boston Univ. Sch. of Med., Boston, MA, [5]Beth Israel Deaconess Med. Ctr., Boston, MA, [6]Univ North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Background: Substantial efforts have been dedicated to exploring the link between genetic regulation and the proteome, informing studies of complex trait mechanisms. Most of these efforts have been limited to populations of European ancestry.
Results: We conducted an Olink protein quantitative trait locus (pQTL) analysis on 1,245 proteins involving 1,033 self-identified African American (AA) and 1,764 non-Hispanic White (NHW) participants from the Women's Health Initiative and Framingham Heart Study. For replication of candidate pQTLs, we used data from 534 self-identified AA adults from the Jackson Heart Study and protein genome-wide association analysis statistics from the UK Biobank Pharma Proteomics Project, including 54,219 participants, of whom 931 are of African ancestry. In total, we identified 5,103 pQTLs (4,496 or 88% cis- and 602 or 12% trans-pQTLs) for 983 proteins. Among these, 195 were previously unreported, with most (166 or 85%) identified in our AA sample, many of which were essentially monomorphic in European reference populations. Several of these newly identified African ancestry-specific pQTLs have been reported in ClinVar; our results suggest impact on circulating protein levels, potentially

bolstering evidence for clinical significance. We identified a “cis pQTL hotspot” within the leukocyte receptor gene cluster on human chromosome 19q13.4. We also provide examples where a particular cis-pQTL, identified through conditional analysis, offers biological insights into an overlapping GWAS signal for disease susceptibility.
Conclusion: The identification of previously undescribed African ancestry-specific pQTLs contribute to understanding protein genetic regulation and highlights the significance of proteomic analysis in diverse populations.

Session Title: Omics Technologies Poster Wednesday Session

Board 1168W: Proteoform quantitation to understand isoforms and protein heterogeneity using a single-molecule proteome analysis platform.

Authors:

A. Huhmer; Nautilus Biotechnology, San Carlos, CA

Abstract:

Alzheimer’s disease (AD) and related neurodegenerative diseases are a significant burden in healthcare due to growing prevalence. One central element in these diseases, particularly in tauopathies, is the microtubule-associated protein Tau. Six primary splicing isoforms of the Tau protein and over 70 different post-translational modifications (PTMs) have been identified on Tau that are believed to impact fibril formation and ultimately disease progression. Though the general prevalence of these alterations has been studied, there remains a significant knowledge gap with regards to which proteoforms (defined by splice variants and one or more PTMs) are most prevalent, how diverse proteoforms may interact with each other, and how the proteoform landscape of tau drives disease progression. Current methods do not adequately address the molecular diversity of Tau proteoforms in neuronal tissue and genomic analysis alone does not provide insight into these types of molecular changes. Here we introduce and apply a single-molecule approach to the quantification of Tau proteoforms from iPSC- derived neurons. The method has three steps. First, single protein nanoparticle-conjugates are deposited on the flow cell. In a second step, billions of Tau molecules are probed using a fluorescently-labeled commercially-available antibody. Antibodies target either splice-variants or site-specific PTMs. In the third step, bound antibody is imaged by fluorescence, localizing each antibody binding event to an individual protein molecule at a specific location. The cycle of binding, imaging, and removal is repeated multiple times to probe for Tau epitopes of interest. Using control samples, we validated that the platform exhibits a linear analytic response and

sensitivity into the yoctomole range. Applying a panel of 8 antibodies (256 potential proteoforms) covering several Tau isoforms and phosphoproteoforms we quantify the relative amounts of Tau proteoform in differentiated iPSC cells confirming the relative amounts of Tau isoforms indicated by Western blot. In addition, the single-molecule nature of the platform also allows us to determine relative quantities of phosphorylated proteoforms present in those cells. These results clearly demonstrate the unique capability of the platform to quantify biologically relevant Tau proteoforms accelerating our understanding of the role of Tau proteoforms in Tau aggregation, a hallmark of AD.

Session Title: Omics Technologies Poster Wednesday Session

Board 1169W: Proteogenomic approaches to advance rare disease diagnostics: a pilot study in the 100,000 Genomes rare disease cohort

Authors:

J. Carrasco Zanini Sanchez[1,2], M. Pietzner[3], A. Kousathanas[4], J. Jacobsen[5], J. Lord[6], D. Baralle[7], N. J. Wareham[2], G. Elgar[4], M. Brown[4], M. Caulfield[5], D. Smedley[5], C. Langenberg[1]; [1]Precision Hlth.care Univ. Res. Inst., Queen Mary Univ. of London, London, United Kingdom, [2]MRC Epidemiology, Univ. of Cambridge, Cambridge, United Kingdom, [3]Computational Med., Berlin Inst. of Hlth.at Charité-Univ.smedizin Berlin, Berlin, Germany, [4]Genomics England, London, United Kingdom, [5]William Harvey Res. Inst., Barts & The London Sch. of Med. & Dentistry, Queen Mary Univ. of London, London, United Kingdom, [6]Sheffield Inst. for Translational NeuroSci. (SITraN), Univ. of Sheffield, Sheffield, United Kingdom, [7]Univ. of Southampton, Faculty of Med., Southampton, United Kingdom

Abstract:

Rare disease patients experience long delays before receiving a diagnosis. For a large proportion a genetic cause is never identified even through whole genome sequencing (WGS), providing a molecular diagnosis for only 25 to 50% of patients. We designed a pilot study in the 100,000 Genomes project (100kGP) to explore the potential of proteomics to improve diagnostic yield in patients with rare diseases. We performed serum proteomic profiling of 495 patients in the 100kGP using the Olink Explore 1536 platform targeting 1463 unique proteins. We included 71 patients with a genetic diagnosis (positive controls), 66 patients with a variant of uncertain significance (VUS) prioritised by clinicians, and 358 undiagnosed patients across more than 10 disease specialties. The Olink assay covered the protein encoded by the affected gene (cis-protein) for 9 of the diagnosed patients and 14 patients carrying VUS. We performed outlier analysis for 100kGP individuals by comparing their protein levels to a reference population-

based cohort (EPIC-Norfolk) of 771 individuals measured with the same Olink assay. We observed strong under-expression (p-value < 3.3×10^{-5} & z-score < -2) of the cis-protein in 2 of the 9 diagnosed patients, and 7 of the 14 patients carrying VUS, compared to the population reference range. For 19 undiagnosed patients, we identified converging evidence for genes through both under-expression of the cis-protein and independent Exomiser variant prioritisation. For example, we identified a missense variant in the *SETMAR* gene present in only two unrelated patients of the 100kGP. Both patients were affected by rhabdomyolysis and had strong under-expression of the SETMAR protein (z-scores = -3.69 and -3.13 standard deviations away from the reference population mean in each proband). We further identified a missense variant in *TIE1* in a patient with strong under-expression of the TIE1 protein (z-score = -5.12) and affected by an ultra-rare monogenic cardiac disorder. We identified this variant also only to be present in the father, who was also affected by this ultra-rare disorder (but not included in the 100kGP proteomic pilot). For 5 of 25 patients with Hereditary Haemorrhagic Telangiectasia, that remained undiagnosed through WGS, we were able to identify a candidate diagnostic variant through integration of proteomics. Across disease indications, missense variants accounted for over half of novel candidate diagnostic variants, followed by splice region variants (~28% of findings). In summary, our proof-of-concept 100kGP pilot study demonstrates the potential diagnostic value of a proteogenomic approach across several rare disorders.

---PAGEBREAK---

Session Title: Omics Technologies Poster Wednesday Session

Board 1170W: Rare variants in non-*MECP2* genes modulating clinical outcome in Rett syndrome.

Authors:

S. Basso[1,2], K. Zguro[1,2], C. Dimatteo[1,2], S. Croci[1,2], G. Brunelli[2], R. Canitano[3], S. Buoni[3], H. Huseynli[1,2], S. Furini[4], C. Lo Rizzo[1,2,5], C. Fallerini[1,2], A. Renieri[1,2,5]; [1]Med. Genetics, Univ. of Siena, Siena, Italy, [2]Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies,Univ. of Siena, Siena, Italy, [3]Neuropsychiatric Unit, Azienda Ospedaliera Senese, Siena, Italy, [4]Dept. of Electrical, Electronic and Information Engineering “Guglielmo Marconi”, Univ. of Bologna, Bologna, Italy, [5]Genetica Medica, Azienda Ospedaliera universitaria Senese, Siena, Italy

Abstract:

Rett syndrome (RTT) is a neurodevelopmental disorder that represents one of the most common genetic causes of mental retardation in girls. Mutations in *MECP2* gene are responsible for RTT but the spectrum of clinical phenotype is wide, suggesting that genetic background or

additional factors may contribute to the syndrome's variable expressivity. In the aim to understand this clinical variability of RTT we investigated 245 *MECP2*-mutated RTT girls with balanced X-chromosome inactivation selected from the Italian Rett Database. Mutations were categorized based on their respective position within the three regions of *MECP2* gene: Methyl CpG Binding Domain (MBD), Transcriptional Repression Domain (TRD) and C-Terminal Domain (CTD). The Clinical Severity Score (CSS) was used for both global and specific sub-areas clinical assessment. Statistical analyses were performed using the R Studio software. Both Polygenic Risk Score (PRS) for the Intellectual Disability (ID and Epilepsy were applied employing regression models. Whole Exome Sequencing (WES) analysis was performed on NovaSeq6000. Variants calling was performed according to the GATK best-practices. ANNOVAR was used for variants annotations on GRCh38 genome build. We found that females with mutations in the CTD had the lowest total CSS compared to ones with mutations in the MBD or the TRD, confirming previous studies about involvement of mutation types in the severity of Rett Syndrome. PRS related to cognitive ability or epilepsy failed to demonstrating involvement of common genomic polymorphisms. Finally, at least one secondary rare pathogenic/likely pathogenic medically relevant variant was detected in 26% RTT patients in non-*MECP2* genes. Mutated genes were involved in neurodegeneration, epilepsy, ID, Attention-Deficit/Hyperactivity Disorder (ADHD), cardiovascular disease, scoliosis, metabolic disease, infectious disease, coagulation defects, dermatitis as well as cancer susceptibility. The most severe epilepsy score was found in patients heterozygous in the epilepsy related genes: *ADPRHL2*, *CC2D2A*, *GRN, PMM2*. Patients affected by severe scoliosis carried *ABCC6* secondary variant. Our results confirm the correlation between the CSS and the type of *MECP2* variants, while PRS apply failed to demonstrate any influence. Unexpectedly, pivotal role of rare variants in additional genes rather than *MECP2* was shown in the RTT clinical phenotype in more than ¼ cases.

Session Title: Omics Technologies Poster Wednesday Session

Board 1171W: Reconstructing refined spatial tissue architecture in single-cell RNA-seq with REPAIR

Authors:

S. Jiang, J. Jin, R. Xiao, M. Li; Dept. of Biostatistics, Epidemiology and Informatics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

Abstract:

A vast amount of single-cell RNA-sequencing (scRNA-seq) data has been generated, but the absence of spatial information has become a significant drawback. Spatial transcriptomics (ST) technologies can overcome this limitation by profiling gene expression while retaining cell location information, making it possible to recover missing cell locations in scRNA-seq. Several algorithms such as NovoSpaRc, CellTrek, and CeLEry, have been developed to leverage the gene expression-spatial location relationships learned from ST to predict the 2D spatial origins of cells in scRNA-seq. However, a common limitation is that they solely utilize gene expression as model features and fail to consider cellular neighborhood information. To address the limitations, we present REPAIR (**RE**fined s**P**atial **A**rch**I**tecture **R**econstruction), a deep learning algorithm for reconstructing cell locations with refined spatial architecture in scRNA-seq. For a given cell, REPAIR employs gene-gene covariance matrices between spatially neighboring cells to effectively represent its neighborhood information. Consequently, REPAIR treats the concatenation of gene expression and flattened neighboring gene-gene covariance matrices as model features and uses a deep neural network to infer the 2D location of each cell. Moreover, we extend the model framework to 3D spatial reconstruction of single cells. To evaluate the performance of REPAIR, we conducted comprehensive benchmarking evaluations on multiple datasets involving brain and cancer tissues, ensuring that the training and testing data were generated from different subjects, studies, and platforms. Our results show that REPAIR robustly retains the pairwise distance relationships between cells and accurately reconstructs the refined tissue architecture, such as the hippocampus region in the mouse brain. By contrast, existing methods can only maintain the rough tissue shapes without detailed microenvironment structures. We demonstrate that the location-recovered scRNA-seq data can be used for cell-cell communication analyses with spatial constraints. Furthermore, based on the location-recovered scRNA-seq data, REPAIR can impute missing gene expression for imaging-based ST data by identifying the physical neighborhoods in scRNA-seq and transferring the full-transcriptome knowledge from these neighbors to the ST cells. Lastly, we demonstrate that REPAIR is computationally fast and memory efficient, making it a promising tool for large-scale studies.

Session Title: Omics Technologies Poster Wednesday Session

Board 1172W: ResSAT: Enhancing Spatial Transcriptomics Prediction from H&E-Stained Histological Images with Self-Attention Transformer Networks

Authors:

A. Liu, Y. Zhao, H. Shen, Z. Ding, H-W. Deng; Tulane Univ., New Orleans, LA

Abstract:

The rapid advancement of spatial transcriptomics technology has revolutionized RNA abundance quantification, offering remarkable spatial resolution for concurrent gene expression profiling, spatial spot localization, and analysis of matched histology images stained with hematoxylin and eosin (H&E). H&E images provide invaluable insights into tissue structure and composition at a microscopic level, complementing gene expression data to elucidate molecular mechanisms underlying tissue organization. However, spatial transcriptomics profiling can be resource-intensive and time-consuming. Current methods for spatial expression prediction from H&E images often overlook spatial relationships between different locations. In this study, we introduce ResSAT (Residual networks - Self-Attention Transformer), a novel framework capable of generating spatially resolved gene expression profiles from whole-slide H&E-stained histology images. Our approach first captures diverse tissue structures and cellular compositions depicted in the images. It then employs a self-attention transformer mechanism to identify and cluster spots with high correlation, enabling the model to focus on interactions between spots and enhance spatial gene expression prediction performance. We demonstrate the effectiveness of ResSAT in spatial transcriptomics prediction by benchmarking its performance on two different datasets captured via the 10x Genomics Visium platform. Our results show significant improvements over existing methods. For example, the mean correlations across all genes in predicted expression versus ground truth expressions in a public sagittal anterior mice brain dataset obtained from 10X Genomics yields a value of 0.2218, and the mean correlations across all genes in the sagittal posterior mice brain dataset obtained from 10X Genomics yield 0.2437. These values notably surpass those obtained using alternative methods. The proposed framework has the potential to substantially reduce the time and cost associated with spatial transcriptomics profiling, paving the way for rapid acquisition of numerous spatial transcriptomics profiles. This advancement holds promise for reconstructing comprehensive 3-dimensional (3D) spatial transcriptomics from adjacent 2-dimensional (2D) spatial profiles.
Grant Acknowledgement: This work is partially supported by grants from the NIH U19AG05573, R01AG061917, R01AR069055, and P20GM109036.

Session Title: Omics Technologies Poster Wednesday Session

Board 1173W: Ribbit: Accurate identification and representation of complex tandem repeats in genomes

Authors:

A. Avvaru, A. Sharma, D. Sowpati; CSIR Ctr. for Cellular and Molecular Biology, Hyderabad, India

Abstract:

DNA tandem repeats (TRs) are nucleotide sequences with a sequence motif repeated contiguously. These elements are abundant across eukaryotes with established functional associations in gene regulation and genome organization. Based on motif size, TRs are categorized as short tandem repeats (STRs, 2-6 bp motifs) and variable number tandem repeats (VNTRs, 7-100bp motifs). These are highly polymorphic sequences showing characteristic length polymorphism due to errors introduced during replication. A recent benchmark study established that though TRs cover 8.1% of the human genome they account for 29% of variants per individual. Abnormal length variations in TR loci are linked to more than 60 neurodegenerative diseases in humans. Tandem repeat sequences are defined based on their structure characteristics including motif sequence, motif length, repetition count and the sequence identity between motifs i.e. TR purity. While studies focused on the understanding impact of motif and repeat length, the relevance of TR purity remains underexplored. This is mainly due to various limitations of existing tools. Firstly, the definition of TRs, especially that of purity, varies across different tools. Also, these tools work well with simple repeats but perform ambiguously at identifying complex repeat loci such as nested and compound TRs. Finally, these tools suffer in accurate representation of TR motifs; smaller motif sizes are usually misrepresented as their longer multimer versions. Here we propose Ribbit, which uses 2-bit representation of DNA sequences for identifying tandem repeats of 2-100 bp motifs with improved accuracy and comprehensive delineation of complex TR loci. Compared to existing tools, Ribbit uses an order of magnitude lower memory while not sacrificing speed, and outperforms existing tools both on simulated and real datasets. We showcase the utility of Ribbit and improved representation of TR loci across different public datasets including established disease associated TR loci, benchmark TR loci, and four other TR databases defined for hg38 and T2T-CHM13 versions of human genome.

Session Title: Omics Technologies Poster Wednesday Session

Board 1174W: ScRNAseq reveals altered stemness and reduced genesis of GnRH neurons in mutated *SMC3* patient-derived iPSCs

Authors:

J. Zhai[1,2], D. Berdous[2], Z. Kolesinska[2], C. P. Lopez[2], Y. Zouaghi[2,1], A. Boizot[3], F. De-Azevedo-Correa[1,2], N. Niederlander[2], F. Santoni[2,1], A. Messina[2,1], N. Pitteloud[2,1]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Lausanne Univ. Hosp., Lausanne, Switzerland, [3]Univ. Hosp. of Lausanne, Lausanne, Switzerland

Abstract:

Background: Congenital Hypogonadotropic Hypogonadism (CHH) is a rare genetic disorder characterized by absent puberty caused by gonadotropin-releasing hormone (GnRH) deficiency and shares phenotypic features with Cornelia de Lange Syndrome (CdLS). While mutations in cohesin genes are known to cause Cornelia de Lange Syndrome (CdLS), their specific impact on the pathogenesis of CHH is uncharted. This study investigates the effects of a *de novo* variant in *SMC3* on GnRH neuron development in a patient with both CdLS and CHH using patient-derived induced pluripotent stem cells (iPSCs) and single-cell RNA sequencing (scRNA-seq) technology. **Methods**: iPSCs were derived from patients and controls and differentiated into neural progenitor cells (NPCs) and subsequently into GnRH progenitors/neurons. Patient and control cells were collected at each stage of differentiation and subjected to split-pool based scRNA-seq using Parse Biosciences Evercode Whole-Transcriptome technology. Demultiplexing of cells (n=8665) from the patient and controls was performed using scSplit. Standard single-cell data processing was conducted using an in-house pipeline that employs the Seurat package. **Results**: Our data demonstrated a marked reduction in stemness, evidenced by an increased rate of spontaneous differentiation in *SMC3* proband-derived iPSCs (54.23%) compared to controls (20.65%; p=4.22e-50). This suggests that *SMC3* mutation may disrupt key regulatory networks maintaining stem cell pluripotency. Further analysis highlighted a notable impairment in the differentiation of GnRH neurons within the patient-derived cultures. The fraction of GnRH neurons in *SMC3* proband-derived cultures was significantly lower (12.37%) compared to controls (63.56%; p=4.77e-79). **Conclusion**: This study demonstrates that *SMC3* mutations can significantly impact human GnRH neuron development *in vitro* and thus contribute to the pathogenesis of CHH**.**

Session Title: Omics Technologies Poster Wednesday Session

Board 1175W: Sensitivity and dynamic range of massively parallel single-molecule protein quantitation illustrate the capabilities of protein identification by short-epitope mapping (PrISM).

Authors:

S. Wilcox[1], T. Gillies[1], R. Grothe[1], V. Budamagunta[1], K. Winters[1], M. Newman[1], M. Villancio-Wolter[1], P. Hao[1], S. Tonapi[1], K. Lazaruk[1], K. Juneau[1], J. Egertson[2], A. Huhmer[1], P. Mallick[1]; [1]Nautilus Biotechnology, San Carlos, CA, [2]Nautilus Biotechnology, Seattle, WA

Abstract:

Many biological processes in the cell cannot be explained by genomics alone and are ultimately driven by protein abundance changes across the full complement of the cellular proteome. Detection of those changes requires a proteomic measurement capable of detecting changes over a wide dynamic range. Protein Identification by Short-epitope Mapping (PrISM) is a novel single-molecule based proteome analysis approach that may enable identification and quantification of nearly the entire human proteome in a single experiment. PrISM has important advantages over other proteome analysis methods due to its ability to directly measure intact proteins. The key attributes of this novel protein measurement platform include the ability to provide comprehensive coverage of sample constituents, the exceptional sensitivity at the single-molecule level, and a wide dynamic range suitable to quantify high and low-abundant proteins in the proteome with high reproducibility and robustness. Here we assess the dynamic range associated with the PrISM approach utilizing a simple mixture of standard proteins. Two proteins from a 21-protein standard mixture were deposited in separate control lanes on the flow cell and in a sample lane as part of the 21-protein mixture. In a second set of lanes, the two proteins were varied over multiple orders of magnitude. PrISM data were acquired on the flow cell using 150 sequential incubate/rinse/image/remove cycles with 50 multi-affinity probes, which are designed to target short amino acid sequences (~3 amino acids). Frequency of binding of the probes to the standard proteins in the control lanes compounded over multiple cycles of measurement provided the basis for confident identification and quantification of the two proteins in the sample lane. The results indicate that the two standard proteins were reproducibly quantified over multiple orders of magnitude in the presence of the mixture with one standard varying as much as one-million-fold vs the other standard. We also demonstrate that the protein decoding signal has a wider dynamic range than quantitation from any single multi-affinity probe due to the ability of the decode algorithm to increase the fidelity of quantitation by accumulating information across multiple multi-affinity probes measurement events. Furthermore, we demonstrate assay sensitivity into the yoctomole range. These attributes showcase the platform's robust capability for precise and sensitive quantification of protein variations in complex biological samples, which will complement genomic analysis with a robust proteomics measurement.

Session Title: Omics Technologies Poster Wednesday Session

Board 1176W: Short tandem repeat genotyping in an immune disorder cohort

Authors:

E. Karlins[1], J. Berghout[1], R. Ghosh[2], B. Seifert[2], M. Tokita[2], J. Yan[2], M. Similuk[2], A. Oler[1], M. Walkiewicz[2]; [1]Bioinformatics and Computational BioSci.s Branch, Office of Cyber Infrastructure and Computational Biology, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]NIAID Centralized Sequencing Program, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD

Abstract:

Background: Short tandem repeats (STRs) are polymorphic repetitive units of DNA in the human genome, known to be important in many disorders, especially movement disorders. In disorders like Huntington's disease and ataxias, expansions of repeat units beyond a pathogenic threshold can lead to defects in the gene involved and cause the disorder. To our knowledge there are no known STRs that lead to immune disorders. We sought to evaluate a cohort of subjects referred to the NIH for immune disorders for genome wide STRs.
Methods: To genotype STRs potentially involved in immune disorders we used the tool ExpansionHunter, querying a catalog of over two million known STRs in Illumina short-read whole genome sequencing data from 2918 individuals with immune disorders (~30 % with molecular diagnosis) and 188 healthy controls. We validated our calls with positive controls for a subset of these with 100% concordance for repeats with fewer than 200 units, and 75% concordance for all repeats. From the distribution of the number of repeat units of each STR in the healthy controls, we set a threshold of three standard deviations above the mean in the control cohort to look for outliers in the immune disorder cohort.
Results: We detected 15 STRs with outliers in multiple patients. Further investigation is necessary to determine if any of these STRs contribute to the immune disorders.
Conclusion:We explored the prevalence of expansion of STRs in a cohort of primary immune disorders. Our method detected novel expansions enriched in the patients in genes with potential clinical significance.

Session Title: Omics Technologies Poster Wednesday Session

Board 1177W: Simultaneous profiling of 1M single-cell BCRs and whole transcriptomes from diseased and healthy B cells in a single experiment using split-pool combinatorial barcoding ⭐

Authors:

A. Sapre, E. Papalexi, A. Salvino, G. Gallareta-Olivares, S. Marrujo, G. Geiss, C. Roco, A. Rosenberg; Parse BioSci.s, Seattle, WA

Abstract:

Many autoimmune diseases are characterized by aberrant B cells which play a central role in pathogenesis. Each B cell expresses a B cell receptor (BCR), which determines its specificity in binding to different epitopes and downstream antibody generation. For autoimmune diseases such as Type 1 Diabetes, Rheumatoid Arthritis, Multiple Sclerosis, Crohn's and Celiac disease our current understanding of the role of unique BCR repertoires is poorly understood; however, recent advances in single-cell RNA sequencing has allowed the field to simultaneously profile both the BCR and whole transcriptome leading to the characterization of clonal diversity of B cell subtypes. Existing BCR profiling methods using microfluidic barcoding are low throughput (1,000s-10,000s of cells) and costly.
To overcome these limitations, we developed a method to simultaneously characterize the BCR repertoire and full transcriptomes of up to 1 million cells from up to 96 samples in one experiment using combinatorial barcoding. We then applied this method to profile 12 healthy and 12 diseased samples. From the combined whole transcriptome and BCR data, we defined BCR isotypes, captured on average 82% of cells with paired chains, detected full length chains, distinguished expanded & unique clones resulting in over 900,000 clonotypes detected, and identified all major B cell subtypes. The dataset is publicly available for researchers to further investigate and parse out deeper insights into immunology. This technology paves the way for high-throughput and sensitive screening of BCR repertoires to further the understanding of B cell biology and disease.

Session Title: Omics Technologies Poster Wednesday Session

Board 1178W: Single-nucleus multiomics in three brain regions reveals the regulatory programs for sporadic early-onset Alzheimer's Disease

Authors:

A. Liu[1], C. Citu[1], N. Enduru[2], X. Chen[1], A. M. Manuel[3], T. Sinha[1], D. A. Gorski[1], B. S. Fernandes[1], M. Yu[1], P. E. Schulz[1], L. M. Simon[4], C. Soto[1], Z. Zhao[5]; [1]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]UTHlth., Houston, TX, [3]The Univ. of Texas Hlth.Sci. Ctr., Houston, TX, [4]Baylor Coll. of Med., Houston, TX, [5]Univ Texas HSC Houston, Houston, TX

Abstract:

Background: Sporadic early-onset Alzheimer's Disease (sEOAD) represents a significant but less-studied subtype of Alzheimer's Disease. The integration of single-nucleus RNA sequencing (snRNA-seq) and single-nucleus chromatin accessibility, using the Assay for Transposase-Accessible Chromatin sequencing (snATAC-seq) has proven invaluable in identifying

candidate *cis*-regulatory elements (cCREs) and regulatory networks associated with late-onset Alzheimer's Disease (LOAD). However, comprehensive analyses of dysregulated regulatory circuits in sEOAD remain limited. To bridge these research gaps and delineate the molecular landscape of sEOAD, we generated single-nucleus multiome data in this study. **Methods:** We generated single-nucleus multiome profiles of 71,763 nuclei from the postmortem prefrontal cortex, entorhinal cortex, and hippocampus of four sEOAD patients and five control donors. Nuclei were annotated using snRNA-seq data from human brains. We conducted systematic analyses to detect cellular compositional changes, transcriptome alterations associated with sEOAD, and cell-type specific cCREs influencing gene expression across different brain regions. We identified dysregulated genes targeted by transcription factors (TFs) acting through distal enhancer-gene connections in sEOAD brains. We also analyzed altered intercellular communications between non-neuronal and neuronal cells, pinpointing potential target genes in receiver cells. Lastly, we fine-mapped genetic variants associated with LOAD and other traits to cell type-specific chromatin accessible peaks in sEOAD brains. **Results:** Our analyses identified cell-type-specific transcriptome alterations in sEOAD and linked cCREs across brain regions. We identified and prioritized seven conservative TFs in glial cells in multiple brain regions, including RFX4 in astrocytes and IKZF1 in microglia, which are implicated in the regulation of sEOAD-associated genes, involving in dysregulated synaptic plasticity and neuroinflammation. Additionally, we prioritized altered intercellular signaling pathways and their regulatory impacts on sEOAD DEGs. Our genetic fine-mapping analysis revealed that LOAD-related risk variants are located in open chromatin regions associated with sEOAD, contributing to transcriptome changes in microglia. **Conclusion:** This study enhances our understanding of the transcriptional and chromatin dynamics in sEOAD and provides a foundational resource for further pathological investigations.

Session Title: Omics Technologies Poster Wednesday Session

Board 1179W: Streamlining NGS Library Normalization: A Novel Method for Efficient and Flexible Equimolar Pooling

Authors:

D. Gelagay, **D. Hite**; Watchmaker Genomics, Boulder, CO

Abstract:

Library quantification and normalization in NGS workflows are crucial for optimizing sequencing throughput, but the process can be time-consuming and error-prone - especially when processing a large number of samples. Traditional protocols require individual library

quantification and adjustments for equimolar pooling using various methods, which can be time-consuming and may preclude resequencing due to library destruction.
To address these challenges, we introduce a novel NGS library normalization method. Our approach starts with addition of normalization reagent to libraries of variable and unknown concentrations. The normalization reagent associates with library molecules in a stoichiometric ratio. Magnetic beads are then added and bind to the normalization reagent effectively pulling out a predetermined number of library molecules for each sample. During elution, the normalization reagent dissociates from the library molecules, resulting in a supernatant containing the normalized library molecules. This normalization method is adaptable and allows for processing of all libraries individually or through an early pooling protocol to reduce consumable use and bead handling. Additionally, the method can be used immediately prior to Illumina sequencing or upstream for normalization of libraries for pre-capture pooling.
We compared our method with a quantitative approach based on the Agilent TapeStation and assessed read counts by sequencing. Libraries normalized using our system exhibited demultiplexing read count variances comparable to traditional methods, with identical sequencing data quality and secondary analysis results.
In conclusion, our NGS library normalization module offers a simplified, streamlined, and non-destructive method that significantly reduces sample processing time.

Session Title: Omics Technologies Poster Wednesday Session

Board 1180W: SVTopo: enhance complex structural variation analysis from highly accurate long reads with human-readable images

Authors:

J. Belyeu[1], W. J. Rowell[1], J. Lake[1], J. M. Holt[2], Z. N. Kronenberg[1], C. T. Saunders[3], M. A. Eberle[4]; [1]Pacific BioSci.s, Menlo Park, CA, [2]PacBio, Menlo Park, CA, [3]PacBio, San Diego, CA, [4]PacBio, Oceanside, CA

Abstract:

Structural variants (SVs) are alleles that differ from the reference genome by at least 50 nucleotides. SVs are common in the human genome and play a major role in both phenotypic diversity and human health. Many are deletions or duplications of genomic material resulting from a single non-reference end-joining event. These are easily identified and visualized from high quality long reads by existing software tools. Other SVs, which we define as complex SVs, are less easily categorized and remain difficult to interpret.
Complex SVs often lead to convoluted signals in both coverage and break-ends. Complex SVs

may combine multiple copy-number alterations in tandem with duplicated inversions, creating genomic rearrangements that may visually appear as several nearby changes in coverage. Inversion events, often appearing with inconsistent coverage and mapping abnormalities, can be difficult to visualize with popular genome browsers in the best of cases but are even more challenging when appearing in tandem with CNVs.
SVTopo addresses the challenge with a dedicated complex variant plotting approach. It uses haplotagged HiFi reads to identify genome alignment breakpoints relative to a reference genome, connects these into multi-locus SVs via shared chimeric alignments, and presents the supporting evidence in easily understood images.
In an analysis of complex SVs within the Platinum Pedigree (CEPH-1463, a family of 28 samples), SVTopo characterized 469 distinct complex SVs. Of these, 34 were solitary inversion events, and 112 were triplications. An additional 87 inversions were found with a flanking deletion on one or both sides, and 112 inversions were found with other complexities such as duplications of nearby sequences, the inverted sequence, or both. SVTopo also found 124 other SVs, such as non-tandem duplications, deletions followed by insertion of a non-tandem duplication, paired deletion/duplication events, and re-ordering of multiple genomic blocks. In many of these cases, SV callers using high quality long reads are able to identify individual SV components, but variant interpretation with SVTopo clarifies multiple calls into a single complex rearrangement. The prevalence and complexity of these variants in many samples without genetic disorders, makes them an important but challenging target of human genomic research, highlighting both the importance of long reads for their identification and the role of targeted software tools like SVTopo to better understand the signals they produce.

Session Title: Omics Technologies Poster Wednesday Session

Board 1181W: Telo-seq: Native, multiplexed telomere sequencing of individual chromosomes using 3’ end adapter ligation ⭐

Authors:

S. Hickey[1], C. Tyer[1], T. Schmidt[2], P. Rughani[1], C. Haggblom[2], J. R. Jones[2], X. Dai[1], K. Frazer[3], F. Gage[2], R. King[4], A. Hodson[4], S. Schwenk[4], A. Jarvelin[4], D. Stoddart[4], S. Juul[1], J. Karlseder[2]; [1]Oxford Nanopore Technologies Inc, New York, NY, [2]Salk Inst. for Biological Studies, La Jolla, CA, [3]UC San Diego, San Diego, CA, [4]Oxford Nanopore Technologies, Oxford, United Kingdom

Abstract:

Telomeres are protective structures at the ends of linear eukaryotic chromosomes consisting of repetitive sequence motifs. Accurate telomere length measurement is important to understanding cellular aging and diseases such as cancer. Long telomere repeat sequences pose challenges for measurement by traditional methods, but Oxford Nanopore Technologies' native long-read sequencing enables precise characterization of haplotype-specific telomere length, sequence composition, and methylation at single chromosome resolution. Here, we present Telo-seq for scalable sequencing and analysis of telomeres. Sequencing adapters are ligated directly to the single-stranded telomeric 3' G-overhang. Full-length telomeres and subtelomeres are sequenced, allowing for unambiguous mapping and chromosome arm assignment of telomeres ranging from 500 bp to over 100 kb in length. Additionally, up to 12 samples can be multiplexed per MinION flow cell, yielding between 1,000 - 2,000 telomeric reads per sample for downstream analysis. Raw telomeric reads were basecalled with Dorado and analyzed using the wf-teloseq bioinformatics analysis workflow. The pipeline assigns reads to individual chromosome arms using T2T references, and also performs de novo guided reference assembly validated using HG002 and HG005 cell lines. The pipeline accurately measures average bulk telomere length with a minimum of 300-500 telomeric reads per sample, or haplotyped chromosome arm-specific telomere length with 10x minimum coverage per telomere arm. Using this workflow, telomere lengths were measured on donor-derived primary fibroblasts, matched iPSCs, and cancer cell lines with different telomere maintenance mechanisms (TMMs). We observed shorter telomeres with increasing donor age and found that telomeres in iPSCs were ~3kb longer than in their matched primary fibroblasts, on average. Additionally, we identified distinct telomere length distributions between ALT-positive and telomerase-positive cancer cell lines, compared to normal cells. Overall, our analysis suggests that individual telomere arm length heterogeneity offers a higher resolution picture of telomere biology in aging and disease.

Session Title: Omics Technologies Poster Wednesday Session

Board 1182W: The blended genome and exome sequencing method captures genetic variation in an unbiased, high-quality- and cost-effective manner ⭐

Authors:

T. Boltz[1], b. chu[2], J. Sealock[1], C. Liao[1], R. Ye[3], L. Zhan[4], L. Majara[3], M. Yohannes[3], D. Howrigan[3], S. Chapman[3], M. DeFelice[3], J. Grimsby[3], S. Rubinacci[3], D. Akena[5], E. Atkinson[6], A. Lukoye[7], T. Bigdeli[8], N. Freimer[4], B. Gelaye[9], K. C. Koenen[9], S. Medland[10], C. R. Newton[11], S. M. Kariuki[11], Z. Zingela[12], A. Stevenson[1], C. N. Pato[13], M. T. Pato[13], D. Stein[14], S. Teferra[15], S. Service[4], L. Olde

Loohuis[4], R. Ophoff[16], B. Neale[1], A. Martin[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Stanford Univ., MENLO PARK, CA, [3]Broad Inst., Cambridge, MA, [4]Univ. of California Los Angeles, Los Angeles, CA, [5]Makerere Univ. Coll. of Hlth.Sci., Kampala, Uganda, [6]Baylor Coll. of Med., Houston, TX, [7]Aga Khan Univ., Karachi, Pakistan, [8]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [9]Harvard Sch. of Publ. Hlth., Cambridge, MA, [10]QIMR Berghofer, Brisbane, Australia, [11]Kenya Med. Res. Inst., Kilifi, Kenya, [12]Nelson Mandela Univ., Gqebera, South Africa, [13]Rutgers, Piscataway, NJ, [14]Univ. of Cape Town, Cape Town, South Africa, [15]Addis Ababa Univ., Addis Ababa, Ethiopia, [16]Univ California Los Angeles, Los Angeles, CA

Abstract:

Microarrays have been the primary technology used in GWAS because of their low cost and genome-wide coverage. However, these are often biased towards European common variation, which can limit the scope of genetic variants identified, including exclusion of novel variants.
We developed and scaled the Blended Genome-Exome sequencing technology in PUMAS (Populations Underrepresented in Mental Illness Associations Studies), on nearly 50K participants across African, African American, and Latin American populations. We required 90% target coverage of the exome at a read depth of 10x and 1-4x coverage on the genome and found that nearly 99% of samples met these coverage thresholds. We observed minor variation in the performance due to the collection method, with saliva-based samples having slightly higher failure rates than blood-based. We called copy number variants (CNV) and structural (SV) variants using GATK-gCNV and VISTA, respectively and then compared them to high coverage "truth" data. Deletion and duplication CNVs spanning >4 exons each have precision and recall scores of nearly 90%; similarly, SV calls have sensitivity and precision >99% at >1kb but are lower for shorter events. We then imputed common variants across the genomes of participants that met quality control thresholds. We evaluated the imputation accuracy of the imputed genotypes against those from the Illumina Global Screening Array (GSA), specifically using Pearson's aggregated R2. We found that all PUMAS cohorts had ≥90% R2 for SNPs across the allele frequency spectrum; ≥95% for common variants (MAF>1%). Finally, we ran polygenic risk score (PRS) analysis in the African population using the imputed BGE data. Using the Psychiatric Genomics Consortium (PGC) GWAS for schizophrenia (SCZ) and bipolar disorder (BD) to train PRS, we evaluated accuracy for predicting case status. We significantly differentiated cases and controls (P=7.8e-74 for SCZ; P=7.4e-49 for BD), though with low predictive value (liability-adjusted Nagelkerke's R2 = 0.0057 for SCZ and 0.003 for BD), as expected for the mismatch in ancestral background of training and testing datasets.
Thus, BGE offers almost all the advantages of deep whole-genome sequencing at less than one-third its cost. This has the potential to enhance access to genomic testing, empowering genomic discoveries, particularly in underrepresented populations.

Session Title: Omics Technologies Poster Wednesday Session

Board 1183W: The PRIDE Affinity-proteomics Archive (PRIDE-AP): Making affinity proteomics datasets FAIR.

Authors:

E. Schmidt[1], D. J Kundu[2], S. Kamatchinathan[2], J. Bai[2], S. Hewapathirana[2], J. Vizcaíno[2], K. Nevola[1], L. Hultin Rosenberg[1], Y. Perez-Riverol[2]; [1]Olink Proteomics, Uppsala, Sweden, [2]European Molecular Biology Lab., European Bioinformatics Inst. (EMBL-EBI), Hinxton, Cambridge, United Kingdom

Abstract:

Background Affinity proteomic research involves analyzing large data sets, and funding institutions and scientific journals require that such data be made publicly available. The PRIDE database at the European Bioinformatics Institute is a popular resource supporting open science practices in MS-based proteomics. The new PRIDE Affinity Proteomics (PRIDE-AP) Archive aims to revolutionize affinity proteomics by centralizing accessible datasets generated using various affinity proteomics platforms. It seeks to establish an inclusive, technology-agnostic repository representing all companies. The first step has been the data submission of Olink Proteomics Proximity Extension Assay (PEA). By adhering to the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR), PRIDE-AP aims to enhance collaboration, expedite research, and ensure high-quality dissemination of affinity proteomics data. **Methods** The PRIDE-AP Archive sets guidelines for submitting, validating, and disseminating affinity proteomics datasets while adhering to FAIR principles. It employs a robust submission and validation process to ensure datasets meet comprehensive metadata and file requirements. Each dataset receives a unique PAD identifier for citation. Submitters must provide essential metadata, including title, description, sample and data analysis protocols, and information on data submitters and principal investigators. Currently, supported technologies include Olink Proteomics submissions with their specific 'RAW' file type requirements. Additional files, such as experimental design documents and quality control reports, can also be included. The PRIDE-AP Archive infrastructure facilitates data submission, review, and public release, with data remaining unreleased to the public until publication. **Results and Discussion** Since its inception, PRIDE-AP has successfully integrated several high-quality affinity proteomics datasets, making them accessible to the scientific community. Implementing unique PAD identifiers has streamlined dataset citation and potential reuse in the scientific literature. The structured metadata requirements and stringent file validation processes have resulted in well-annotated, interoperable datasets. PRIDE-AP significantly advances the field of affinity proteomics by

adhering to the FAIR principles. By centralizing datasets from technologies such as Olink Proteomics, PRIDE-AP facilitates collaboration, accelerates research, and promotes the widespread dissemination of high-quality data.

Session Title: Omics Technologies Poster Wednesday Session

Board 1184W: Tracking dynamic regulatory changes in rare diseases using single sample networks

Authors:

H. Yuan[1], C. A. Mancuso[2], K. Johnson[2], I. Braasch[1], A. Krishnan[2]; [1]Michigan State Univ., East Lansing, MI, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

In most rare diseases, even if the underlying genetic cause is known, there are significant gaps in our understanding of the impact of the genetic variants on the molecular and cellular mechanisms involved in the disease. To gain insights at such mechanistic levels, researchers frequently use transcriptomic analysis to understand regulatory changes in rare diseases. However, as a snapshot of regulatory change, gene expression itself, cannot effectively capture the rewiring of the underlying gene networks. Hence, instead of inspecting changes at the transcriptomic level, we need to investigate network-level changes to gain a comprehensive view of regulatory dynamics in rare diseases. However, a typical gene network summarizes average interactions among genes from multiple samples, which blurs the regulatory differences among samples collected at distinct disease stages. In this study, we developed an algorithm to reconstruct single sample networks that can accurately infer regulatory relationships among genes of a sample using a single transcriptome as input. We applied this method to investigate the regulatory mechanisms in Facioscapulohumeral muscular dystrophy (FSHD). FSHD is associated with the upregulation of the *DUX4* transcription factor and its target genes in skeletal muscle cells. Although the causal gene is known, the disease pathogenesis of FSHD is highly complex. This complexity arises from epigenetic and post-transcriptional gene regulation, as well as various feedback regulatory loops that control *DUX4* expression. We built single sample networks for time series transcriptomes of an FSHD cell line, which mimics FSHD in patients by inducing the upregulation of *DUX4* in myoblasts. By comparing the single sample networks along the progression, we were able to effectively monitor regulatory changes over time and identify target genes and pathways of *DUX4*. We also identified "dark" genes with nuance fluctuations in expressions, which might be missed by transcriptomic analysis. These "dark" genes include hub genes that could significantly influence regulatory landscapes. Results from

our analysis advance the understanding of complicated regulatory dynamics for FSHD and also offer potential therapeutic targets for FSHD. Beyond FSHD, our single sample network-based approach serves as a systematic framework for understanding rare diseases involving complicated regulatory changes.

Session Title: Omics Technologies Poster Wednesday Session

Board 1185W: Uncovering New Disease Associated Cellular States Within and Across Cell Types Using residPCA in Single-Cell RNA and ATAC Sequencing Data

Authors:

S. Carver[1], A. Gusev[2]; [1]Harvard Univ., Boston, MA, [2]Dana Farber, Cambridge, MA

Abstract:

To enhance our understanding of cellular heterogeneity and disease from single-cell sequencing data, robust methods for cell state inference are essential. While current tools excel at cell type annotation, they fall short in identifying cell states within and across cell types. We introduce residPCA, a novel method for cell state identification in single-cell data that models cell type heterogeneity.
Simulations show that residPCA outperforms state-of-the-art methods, especially in scenarios with numerous cell types, states spanning multiple cell types, and detecting rare states. For instance, in simulations with a state expressed in 1 out of 100 total cell types, residPCA outperforms other methods (residPCA: R2=0.69, PCA: R2=0.22, NMF methods: R2&lt0.15). In 3,000 cells, residPCA identifies states with the same accuracy as NMF does with 14,000 cells. In scRNA-seq data from light-stimulated neurons, residPCA captures stimulus-driven variability better than PCA, NMF, and cNMF (residPCA: R2=0.24, PCA: R2=0.23, NMF methods: R2&lt0.01), while UMAP fails to detect relevant states. residPCA accurately detects light-related states in endothelial cells and smooth muscle, cell types only recently implicated in light stimulation, outperforming other methods (residPCA: R2=0.34, PCA: R2=0.30, NMF methods: R2&lt0.05). Applying residPCA to 191,890 nuclei from an Alzheimer's disease (AD) cohort, including snRNA-seq and snATAC-seq data, revealed 44 ATAC-based and 42 RNA-based states, indicating higher complexity in single-cell data than commonly recognized. Integrated with AD GWAS data, these cell states showed significant heritability enrichment, linking cell state inference to disease risk. residPCA identified more ATAC states enriched for AD heritability than other PCA-based methods (residPCA: 30 states, PCA: 27 states, Excitatory Neuron PCA: 8 states) and identified states with greater significance than previously established enriched cell types like microglia

(most significant residPCA state annotation p-value: 1.89e-8, microglia annotation p-value: 0.14). In complex brain data, residPCA discerned both known and novel cell states, highlighting states contributing to AD-related neuroinflammatory phenotypes, particularly within microglia. Our findings suggest that residPCA can uncover disease-relevant cell states from single-cell data often overlooked by traditional methods. The computational implementation of residPCA provides a user-friendly platform for researchers to apply this method to datasets, facilitating further discoveries in disease-relevant cell states.

Session Title: Omics Technologies Poster Wednesday Session

Board 1186W: Whole Genome Scanning of Deep Variants Existing in Human Tissues by Duplex-seq with Comprehensive Genome Coverage

Authors:

M. Niu[1], Y. Zhang[1], J. Luo[1], H. Doddapaneni[1], D. A. Weitz[2], R. A. Gibbs[1], C. Zong[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Harvard Univ., Cambridge, MA

Abstract:

Somatic mutations continuously accumulate in the human genome, posing vulnerabilities towards increased risk of various diseases along aging. Many of these somatic mutations only exist in a limited number of cells or even single cells, and we denote them as deep variants. It is technically challenging for traditional approaches to profile these deep variants. To solve this problem, recent development of Duplex-seq based methods provide a potential strategy that allows the accurate detection of mutations in single DNA molecules. By preventing the damage induced misincorporation, the development of NanoSeq has greatly improved the detection accuracy of duplex-seq approach from ~10^{-7} errors per base pair to ~10^{-9}. However, the limited genome coverage of NanoSeq (~30%) prevents the genome-wide detection of somatic mutations. To overcome this limitation, we developed a duplex sequencing method with comprehensive genome coverage, which we refer to as CompDuplex-seq. The streamlined chemistry of CompDuplex assay allows efficient generation of libraries that are compatible with standard Illumina paired-end sequencing. We have validated the accuracy of somatic mutation calling and comprehensive genome coverage of CompDuplex by profiling human cord blood cells. And more importantly, we demonstrated successful detection of deep variants existing in postmortem human brain neurons and many of those deep variants were predicted to have functional effects through Combined Annotation Dependent Depletion (CADD) analysis. To summarize, CompDuplex chemistry supports genome-wide coverage while maintaining high

precision with <10^{-9} error rate per base. We believe that CompDuplex will facilitate the whole genome characterization of somatic mosaicism in human tissues.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7001W: A call from the stars in Pharmacogenetics: Accurate genotyping of *CYP2D6* and PGx relevant genes from next-generation sequencing data

Authors:

J. Yu, I. Protopopov, D. Ivanov, X. Bonilla Bustillo, M. Pireyre, **T. Maas**, Z. Xu; SOPHiA GENETICS, Rolle, Switzerland

Abstract:

The genotyping of pharmacogenomic (PGx) genes is crucial for medical decisions in personalized medicine. However, PGx genotyping, or star-allele calling, remains challenging from clinical next-generation sequencing (NGS) data, as some genes (e.g., *CYP2D6*) require accurate phasing and structural variant calling in the presence of homologous regions. Here we present a novel solution, which addresses those challenges and provides the final report from the FASTQ files in an efficient way. Using cross-sample normalization and Hidden-Markov models, we were able to identify structural variants between genes and corresponding pseudogenes. Combining copy number and all known star-alleles, we established a novel framework to accurately phase variants beyond the typical NGS read size. We demonstrated the analytical performance of the solution by processing 135 Coriell reference samples and multiple runs of clinical samples using capture-based NGS. After comparing our performances to the consensus genotypes from GeT-RM and the state-of-the-art methods, such as Aldy4 and PharmCAT, we showed our method reached 99% concordance for CYP2D6 and up to 100% concordance for other genes. We identified that the remaining differences were mostly due to difference in database version for star-allele nomenclature. In conclusion, our one-stop solution addresses numerous challenges encountered in clinical settings, showing at least the same or even superior analytical performance compared to state-of-the-art tools in the field.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7002W: A methodology to assess concordance of last prescribed treatment with CPIC guideline recommendations: A case study application in a diverse All of Us research cohort

Authors:

I. Rattsev[1,2], J. M. Stevenson[3,4], S. Song[5,2], C. Overby Taylor[1,2,5]; [1]Dept. of BioMed. Engineering, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Inst. for Computational Med., Johns Hopkins Univ. Whiting Sch. of Engineering, Baltimore, MD, [3]Div. of Clinical Pharmacology, Dept. of Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [4]Dept. of Pharmacology and Molecular Sci., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [5]Div. of Gen. Internal Med., Dept. of Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Abstract:

Pharmacogenetics (PGx) refers to how variability within a gene contributes to an individual's medication response and may improve the safety and effectiveness of drug therapy. To aid clinicians in using PGx in their practices, Clinical Pharmacogenetics Implementation Consortium (CPIC) has published peer-reviewed and evidence-based clinical practice guidelines. This work proposes a method to evaluate concordance of the last prescribed treatment with a CPIC-recommended medication. First, we calculated *concordance rate* (CR) or the proportion of people with a final medication match with CPIC guidance. We considered CR to be high when it is greater than 80%. A subset of people had genotypes that do not match to CPIC-defined phenotypes, and their phenotype is considered *indeterminate.* To understand the extent of this problem, we reported the frequency this occurs. Chi-squared test for proportions was used to identify statistically significant differences in CR between races that are historically underrepresented in biomedical research (UBR) and non-UBR. As an illustrative case study, we apply this methodology for the CPIC guideline for clopidogrel in a diverse *All of Us* research program (AoU) cohort. This CPIC guideline recommends avoiding clopidogrel for certain *CYP2C19* phenotypes (e.g., intermediate and poor metabolizers), and that an alternative $P2Y_{12}$ inhibitor, such as prasugrel or ticagrelor, be used instead. We selected a diverse cohort of 9,346 *All of Us* research participants with at least one record of a $P2Y_{12}$ inhibitor prescription, including 36% self-reporting as races that are UBR. For each participant, we identified their last prescribed $P2Y_{12}$ inhibitor and their *CYP2C19* phenotype according to the clinical variant calls included in PGx reports provided to AoU participants. A small proportion of participants had an indeterminate *CYP2C19* phenotype (n=100, 1%), the majority of which occurred in UBR races (62/100, 62%). Overall CR was moderate to high (74%). We found that CR was higher (better) for non-UBR participants, when compared to UBR participants (0.75 vs 0.72, *p-value < 0.005*). Given that genotypes were not known at the time of prescribing for most AoU participants, our

findings suggest that UBR participants stand to gain more from PGx-guided care when compared to non-UBR participants. Further work is needed to define functionality of *CYP2C19* variants identified in participants with indeterminate phenotype to ensure they can benefit from PGx-guided antiplatelet selection.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7003W: Association of Chromosome 19 Variants with Nicotine Metabolite Ratio in American Indians: Insights from the Strong Heart Study

Authors:

C. Sherman[1], C. R. Musser[1], K. Ferrier[1], K. M. Aagaard[1], T. H. Wroblewski[2], R. F. Tyndale[3], M. J. Chenoweth[3], N. Franceschini[4], S. A. Cole[5], K. G. Claw[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [3]Univ of Toronto, Toronto, ON, Canada, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [5]Texas BioMed. Res. Inst, San Antonio, TX

Abstract:

Background: Nicotine is the primary addictive substance of commercial tobacco, and it is primarily metabolized by CYP2A6, located on chromosome 19. A measurement of nicotine metabolism rate and CYP2A6 activity is the nicotine metabolite ratio (NMR): 3'hydroxycotinine/cotinine. A higher NMR indicates faster metabolism and influences treatment types and tobacco consumption behaviors. NMR is highly heritable (ranging from 0.49 to 0.81), and previous studies have identified genetic markers, some unique to populations, associated with NMR and other tobacco-related phenotypes. American Indians (AI) have greater prevalence of reported tobacco use compared to other populations within the US, and efforts are needed to investigate the contribution of genetic variation to interindividual variability in NMR to improve tobacco cessation efforts. Methodology: Genotype and NMR data was collected (790 smokers and 30 controls) from phase 2 (1993 to 1995) of the Strong Heart Study (SHS) cohort, a longitudinal study of cardiovascular disease and risk factors in 12 AI tribes in Arizona, North and South Dakota, and Oklahoma. DNA was collected from blood samples and genotyped using the Multi-Ethnic Genotyping Array (MEGA) genome-wide array chip. NMR was measured from blood plasma samples (100 µL; n=820) via LC/MS/MS. The NMR values were log-transformed to normalize distribution. We used linear regression models to examine the association between chromosome 19 variants and log-transformed NMR (N=650). Results: 41 variants on chromosome 19 were found to be significant (p<5.0e-8). There were 5 SNPs within

nicotine-metabolizing genes, CYP2A6 and CYP2B6. The most significant overall association was rs935620232, an intronic SNP in CYP2A6 (P=2.13e-15; beta=-0.16), and the second was an upstream variant in CYP2A6, rs28399433 (P=5.49e-14, beta=-0.16), which is a SNP within the CYP2A6*9 haplotype that is known to affect expression of CYP2A6, reducing its enzymatic activity. We also identified other intronic variants within CYP2A6 rs56267346 (P=1.99e-08, beta=-0.07) and rs2431413 (P= 4.2e-11, beta=-0.14), and one intronic variant within CYP2B6, rs33912321 (P=1.38e-08, beta=-0.09). Conclusion: This targeted analysis with the SHS cohort contributes to a better understanding of inter-population genetic variability and its impacts on NMR. An improved understanding of genetic factors influencing NMR could be used to improve and tailor tobacco cessation efforts.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7004W: Berberine signature and cardiometabolic diseases: a study leveraging randomized controlled trial, cohort study, and Mendelian randomization

Authors:

J. Zhao; The Univ. of Hong Kong, Hong Kong, Hong Kong

Abstract:

Background: Statins are the mainstay for the treatment of hyperlipidemia, the established risk factor for ischemic heart disease (IHD). However, statins are not tolerable to all people and have the side effect of hyperglycemia. Berberine, with benefits on lowering lipids and improving glucose metabolism, is a promising candidate, but its effect on IHD and diabetes has not been clarified.Methods: Leveraging the proteomics and sex hormones data in the completed randomized controlled trial of berberine in 80 men, we built a signature reflecting the response to berberine use (in short “berberine signature”) using elastic net. For validation, we examined the association of the berberine signature with lipids and glucose in the trial and also in UK Biobank, to assess whether the berberine signature had an inverse association with total cholesterol, LDL-cholesterol and fasting glucose, as expected. We then assessed whether this signature was associated with the risk of IHD and diabetes in UK Biobank using observational study and bi-directional Mendelian randomization (MR). We further examined the association of each protein and sex hormone with IHD and diabetes to identify the protein(s) and/or sex hormone(s) underlying the effects of berberine signature.Results: We included 18 proteins and 2 sex hormones in the signature. As expected, berberine signature was related to lower total cholesterol, fasting glucose and LDL-cholesterol in the UK Biobank using observational study and

MR. Berberine signature was also related to lower risk of IHD and diabetes using observational study and MR (odds ratio (OR) for IHD 0.82, 95% confidence interval (CI) 0.76 to 0.88; for diabetes 0.88, 95% CI 0.80 to 0.96). Sex hormone binding globulin (SHBG), CCL5, PRSS2 and PLA2G7 may explain the beneficial effects of berberine on IHD, while testosterone, SHBG, CCL5, PRSS2, CNDP1, F11, LCN2, and THBS4 may explain the beneficial effects of berberine on diabetesConclusion: Our study provides support for the beneficial effects of berberine on IHD and diabetes. The proteins and sex hormones underlying these effects can be considered as potential targets for new drug development or drug repurposing.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7005W: Characterizing drug-induced stuttering in biobank-linked electronic health records enables potential precision medicine approaches

Authors:

D. Pruett[1], C. A. Hunter[2], A. Scartozzi[3], D. Shaw[3], S. Kraft[4], R. Jones[1], M. Shuey[5], J. Below[5]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Lipscomb Univ. Coll. of Pharmacy and Hlth.Sci., Nashville, TN, [3]Vanderbilt Univ., Nashville, TN, [4]Wayne State Univ., Detroit, MI, [5]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

*Purpose*Drug-induced stuttering (DIS) is a phenomenon where the onset of stuttered speech is caused by exposure to pharmaceuticals. DIS features many of the same speech behaviors as developmental stuttering. While DIS is rare, medications implicated in the condition are widely used. Prior research on the topic, however, has been limited to case studies and adverse drug reaction reports. Precision medicine approaches may elucidate possible gene x exposure interactions and help characterize the genetic architecture associated with stuttering at large. As a first step toward this goal, this investigation involved an in-depth electronic health record (EHR) review of DIS cases to classify the drug types and mechanisms of action associated with the condition.*Methods*Suspected cases of DIS were initially identified and flagged in a systematic investigation of developmental stuttering within approximately 3 million de-identified EHRs at Vanderbilt University Medical Center. This study examines the 40 resultant suspected DIS cases to describe: 1) the level of evidence for the implicated drug as a causal agent, 2) name, class, and mechanism of action of suspected drug, 3) other drugs present surrounding the instance of stuttering, 4) therapeutic measures including removal of drug, increase/decrease dosage, or initiation of new drug, and 5) progression or remission of

stuttering.*Results*Eighteen different drugs were linked to possible DIS in 22 individuals. Antiseizure agents (8 cases), CNS stimulants (4 cases), and antidepressants (4 cases) were the most common drug classes implicated in drug-induced stuttering. Topiramate was the most commonly implicated drug across all records reviewed. No clear pattern emerged from the mechanisms of action of implicated drugs; however, effects on GABA, glutamate, norepinephrine, serotonin and dopamine neurotransmitters and receptors were observed across multiple drugs and drug classes.*Discussion*The impact of psychotropic medications on speech fluency has been previously documented. However, the mechanism of action for these implicated drug classes remains unclear. Disfunction in connections between the cerebral cortex, basal ganglia, and thalamus, termed the cortico-basal ganglia-thalamocortical (CBTC) loop, has been studied in the pathogenesis of both developmental stuttering and neurogenic stuttering and may provide a model for DIS. Further studies exploring the epidemiology and pharmacology of DIS, especially pharmacogenetic approaches, are needed to characterize the etiology of the condition and to provide insight into stuttering in general.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7006W: Clinical integration of NGS-based pharmacogenomics: Accurate genotyping and practical applications

Authors:

Y-S. Liu[1], I-H. Chiu[2], C-F. Huang[3], L-J. Shen[4], L-C. Kuo[5], P. Chen[1,2]; [1]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ., Taipei, Taiwan, [2]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [3]Dept. of Pharmacy, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [4]Graduate Inst. of Clinical Pharmacy, Sch. of Pharmacy, Natl. Taiwan Univ., Taipei, Taiwan, [5]Dept. of Internal Med., Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Abstract:

Identifying genetic variants that influence drug response is crucial for precision medicine, yet the full clinical utility of next-generation sequencing (NGS) pharmacogenomics panels has not been fully realized. This study aims to establish a preemptive pharmacogenomics workflow within a hospital setting to integrate pharmacogene data into genotype-based medication prescriptions. Including over 400 participants from the National Taiwan University Hospital (NTUH), we developed a comprehensive NGS panel that targeted 60 key pharmacogenes, selected based on Clinical Pharmacogenetics Implementation Consortium (CPIC), Pharmacogenomics Knowledge Base (PharmGKB), and Pharmacogene Variation Consortium (PharmVar) databases. Fourteen genes were fully sequenced, while 46 genes were sequenced in

their coding regions. Genetic analysis was conducted using the DRAGEN 4.2 star allele caller, with cross-verification by ALDY. Our panel achieved accurate pharmacogenes capture with sufficient read depth for precise genotyping. We have issued genetic test reports for these participants for 20 clinically significant genes, and the observed allele frequencies align with those of East Asian populations. This consistency firmly demonstrates our NGS genotyping as a reliable reference for pharmacogenes in Taiwanese populations. Furthermore, we established a Clinical Decision Support System (CDSS) at NTUH, which offers genotype-based alerts and medication recommendations during prescribing. By integrating NGS panel testing, genetic analysis, and clinical application, this study has enabled the practical application of pharmacogenomics in a real-world hospital setting.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7007W: Design and development of a novel dPCR assay for determination of *CYP2D6* star (*) allele copy number variations.

Authors:

I. Kavanagh[1], A. Coles[1], H. Fielder[1], A. Guardiola[1], S. Agoreyo[1], A. Mihajlovic[1], J. H. McDermott[2,3], W. G. Newman[2,3]; [1]Apis Assay Technologies Ltd, Manchester, United Kingdom, [2]Manchester Ctr. for Genomic Med., St Mary's Hosp., Manchester Univ. Hosp. NHS Fndn. Trust, Manchester, United Kingdom, [3]Div. of Evolution, Infection and Genomics, Sch. of Biological Sci., The Univ. of Manchester, Manchester, United Kingdom

Abstract:

CYP2D6 is important for metabolising up to 25% of drugs, it is encoded by a complex gene with 163 identified star (*) alleles and frequent copy number variations (CNV). Guidelines for prescribing drugs metabolised by CYP2D6 e.g. opioids and tamoxifen are determined by efficacy and safety data from clinical trials, which can be biased by the *CYP2D6* genetic diversity represented within trial populations. Currently there is limited availability of *CYP2D6* genotyping assays that allow CNV quantitation on the * allele level which is important for establishing the correct patient phenotype. Additionally, current methods have high workflow complexity and/or lack the accuracy of digital PCR (dPCR) for copy number determination. Development of a dPCR assay for accurate determination of *CYP2D6* * allele copy number variations has the potential to improve health inequality by enabling tailoring of treatment for those individuals whose *CYP2D6* genotype would pre-dispose them to a worse health outcome. For example, individuals with estrogen receptor positive breast cancer and a poor metaboliser phenotype are at higher risk of breast cancer recurrence when treated with tamoxifen and could be offered

alternative therapies based on the *CYP2D6* dPCR result.
Inclusivity of the dPCR assay was determined by assessing the frequency and function of each of the 163 identified * alleles. Alleles where the function was designated as unknown or uncertain were excluded as detection of these alleles would not impact determination of CYP2D6 phenotype and subsequent clinical decision making. The remaining alleles were prioritised for inclusion if the average frequency across all biogeographical groups was ≥ 1.5%. In cases where the average allele frequency was < 1.5% but the frequency in at least one biogeographical group was ≥ 1% these alleles were also prioritised for inclusion to ensure that any bias resulting from varied genetic diversity within different biogeographical groups was not introduced into the test design. The population coverage of the test is estimated at 93% when alleles with uncertain or unknown function are excluded from the calculation. Of the remaining alleles for 4% the specific * allele is not differentiated but the function would be called correctly giving an effective coverage of 97%.
Primers and probes were designed to detect the pharmacogenomic relevant single nucleotide variations for included * alleles with addition of up to three locked nucleic acids to ensure correct discrimination from *CY2PD6* gene homologues *CYP2D7* and *CYP2D8*.
Herein we describe the *CYP2D6* dPCR assay design, inclusivity and initial proof of concept results.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7008W: Distribution of *CYP2C8*2 and *3* in the South (Velingara Health District) of Senegal: Alleles involved in the metabolism of Artesunate Amodiaquine used in the treatment of malaria ⭐

Authors:

S. Diallo[1], C. Fall[1], J. Diop[2], Y. Dia[1], K. Sylla[1], R. Ndiaye[1], B. Faye[1]; [1]Univ. Cheikh Anta Diop (UCAD) Sénégal, Dakar, Senegal, [2]Univ. Cheikh Anta DIOP (UCAD) Sénégal, Dakar, Senegal

Abstract:

Introduction Malaria remains a major public health problem in developing countries. ACTs are used in the treatment of malaria. Host genetic variation in drug metabolising enzymes influences the selection of drug resistance. Amodiaquine, mainly biotransformed by *CYP2C8*, is known to be associated with adverse events of varying severity. The aim was to assess the distribution of *CYP2C8*2 and *3* alleles involved in ASAQ metabolism.
Materials and methods Our study took place in Kounkané in southern Senegal between September 2022 and July 2023. All patients included in the study were treated with ASAQ. Each

patient received a thick blood sample to diagnose malaria and 3 drops of blood on Whatman paper for PCR and exome sequencing in order to search for the *CYP2C8*2 (c.805A>T, Ile269Phe)* and *CYP2C8*3 (c.416 G>A, Arg139Lys)* alleles, which were located respectively at exon 5 and 3 of the *CYP2C8* gene located on the chromosome *10q23.33* .
Results The prevalence of defective *CYP2C8* alleles was 22% for *CYP2C8*2* and 48% for *CYP2C8*3*. All patients with the *CYP2C8*2* allele were heterozygous, with a high parasite density at D1, and those with the *CYP2C8*3* genotype predominantly had a medium parasite density in the homozygotes and a low parasite density in the heterozygote. A persistence of side effects after 48 hours of treatment was found in patients with a defective allele without significant difference (p value = 0.63). A *CYP2C8*2/*3* double mutation was observed in 3 patients.
Conclusion The high frequency of *CYP2C8*2 and *3* alleles found in our study could have an impact on malaria control. These alleles are associated with a slow metaboliser phenotype which could favour the selection of parasites carrying resistance markers *pfmdr* and *pfcrt* and the accumulation of ASAQ.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7009W: Effect of a novel missense variant in TNF on TNF-alpha binding with receptor proteins

Authors:

Z. Jan, D. Velayutham, B. Mifsud, **P. Jithesh**; Hamad Bin Khalifa Univ., Doha, Qatar

Abstract:

Inhibitors targeting tumor necrosis factor-alpha (TNF-α) have become a common treatment for autoimmune diseases. Despite their widespread use, 40 to 60% of patients do not respond to TNF inhibitors. Numerous genetic variants influencing the response to TNF inhibitors have been identified, but their prevalence in the Qatari population remains unknown. To uncover the genetic variants associated with TNF inhibitors in the Qatari people, we analyzed whole genome sequencing data from 14367 Qatari individuals. We evaluated the prevalence of known variants and identified novel variants in 881 genes associated with TNF inhibitors, TNF signaling pathways and TNF co-expression modules. We identified a novel deleterious missense variant p.(Pro88Ser) in the trimer region of TNF-α. Computational tools predicted that this variant can cause the gain of molecular recognition features (MoRF) binding, loss of methylation at residue Lys87, an increase of phosphorylation, and glycosylation. Moreover, through molecular docking and molecular dynamics simulations, we observed that the variant induces conformational

change near the catalytic site of TNF during its interaction with receptor protein TNFR1. The structural and binding comparison of wild type and mutant TNF reveals interaction plasticity between TNF and TNFR1 in the vicinity of the mutant residue Ser88. Hence, we hypothesize that this interaction plasticity in the TNF-TNFR1 complex induced by the Pro88Ser variant might lead to poor response to TNF blockers. Further studies in *in vitro* and *in vivo* systems will help in validating and characterizing the mutational effect.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7010W: Genetic variants in insulin pathway genes are associated with statin-induced diabetes

Authors:

J. Yee[1], M. Park[2], J. Kim[2], S-A. Choi[2], D. Lee[2], Y-A. Park[2], Y. Chang[3], T-J. Song[4], H. Gwak[2]; [1]Sungkyunkwan Univ., Suwon, Korea, Republic of, [2]Ewha Womans Univ., Seoul, Korea, Republic of, [3]Ewha Womans Univ. Mokdong Hosp., Seoul, Korea, Republic of, [4]Ewha Womans Univ. Seoul Hosp., Seoul, Korea, Republic of

Abstract:

Statins, or 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors, are lipid-lowering agents used for the primary and secondary prevention of cardiovascular diseases. There is growing evidence for the development of new-onset diabetes during statin therapy. One potential explanation is that statins may induce diabetes by impairing the insulin signaling pathway. To explore this further, we aimed to investigate the genetic associations of insulin pathway genes with statin-induced diabetes. Within the insulin pathway, we selected and genotyped candidate single nucleotide polymorphisms within 13 genes: *IGF1, IGF1R, IGF2, INS, INSR, IRS1, IRS2, PIK3CA, PIK3CB, PIK3R1, PIK3R2, AKT1*, and *AKT2*. We utilized dominant and recessive models and chose the most appropriate one based on both the effect size and statistical significance. Unadjusted and adjusted odds ratios were calculated using logistic regression. Among 602 patients receiving statin therapy, 71 (11.8%) developed diabetes. After adjustment, *INSR* rs1799817, *INSR* rs2059807, *IGF1R* rs2715439, and *PIK3R1* rs3730089 were significantly associated with statin-induced diabetes. Patients with the AA genotype of rs1799817 and the G allele of rs2059807 experienced a 2.2-fold and a 2.4-fold higher risk of statin-induced diabetes, respectively. Similarly, those with the CC genotype of rs2715439 and the GG genotype of rs3730089 had a 4.8-fold and 2.2-fold increased risk, respectively. When comparing patients with and without diabetes before statin therapy, none of the associated variants in the main analysis exhibited statistical significance, implying that these associations

may be specific to statin therapy. This study identified the genetic effects of insulin pathway genes on statin-induced diabetes. These findings could help understand statin-induced diabetes and provide optimized therapy to patients receiving statin therapy. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (grant number: NRF-2023R1A2C1007463).

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7011W: Genome-wide association study identifies candidate genetic variants associated with postoperative opioid-related side effects (nausea/vomiting and sedation) in patients undergoing laparoscopic-assisted colectomy (LAC)

Authors:

D. Nishizawa[1], T. Mieda[2], M. Tsujita[3], H. Nakagawa[3], S. Yamaguchi[4], S. Kasai[5], J. Hasegawa[5], Y. Ebata[5], K. Nakayama[5], A. Kitamura[3], M. Hayashida[6], K. Ikeda[5]; [1]Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan, [2]Dept. of Anesthesiology, Saitama Med. Univ. Hosp., Saitama, Japan, [3]Dept. of Anesthesiology, Saitama Med. Univ. Intl. Med. Ctr., Saitama, Japan, [4]Dept. of Gastroenterological Surgery, Saitama Med. Univ. Intl. Med. Ctr., Saitama, Japan, [5]Addictive Substance Project, Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan, [6]Dept. of Anesthesiology & Pain Med., Juntendo Univ. Sch. of Med., Tokyo, Japan

Abstract:

Purpose: Opioid analgesics are commonly required during and after laparoscopic-assisted colectomy (LAC). Whereas postoperative nausea and/or vomiting (PONV) and sedation are known as major opioid-related side effects following LAC, considerable individual differences have been widely observed in incidence of these side effects. We conducted genome-wide association study (GWAS) in patients who underwent LAC to identify potential genetic variants that may significantly contribute to individual differences in PONV and sedation. **Methods**: The subjects recruited were a total of 351 patients with written informed consent who underwent LAC under general anesthesia using sevoflurane and remifentanil and received intravenous fentanyl patient-controlled analgesia for postoperative pain control at Saitama Medical University International Medical Center. The study protocol was approved by the Institutional Review Board at each related institute. Total genomic DNA was extracted from peripheral blood samples and genotyping was conducted using whole-genome genotyping arrays with more than 900,000 markers and

iScan System (Illumina, San Diego, CA, USA). GWAS was conducted in the entire 351 patient subjects to identify single nucleotide polymorphisms (SNPs) associated with of nausea/vomiting and sedation in trend, dominant, and recessive genetic models. **Results**: As a result of GWAS, no SNPs showed significant associations with incidence of nausea/vomiting. However, six SNPs, rs9391859, rs12810375, rs6555261, rs7373242, rs2526909, and rs116325082 (exm520781) SNPs, were significantly associated with presence/absence of sedation in additive model (corrected $P < 0.05$). The rs9391859, rs7373242, rs2526909, and rs116325082 SNPs were located within the *BTNL2*, *LMLN*, *DPF3*, and *SCGN* gene regions, whereas the rs12810375 and rs6555261 SNPs were located within intergenic regions. **Conclusions**: The results indicate that the SNPs identified could serve as markers that predict incidence of sedation. Our findings will provide valuable information for achieving satisfactory control of opioid-related side effects.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7012W: Integration of transcriptome-wide association study of Alzheimer's disease with large-scale perturbation data for drug repositioning and new therapeutic approaches

Authors:

Y. Xie[1], D. Nachun[1], C. Liu[2], F. Morgante[2], The FunGen-AD Consortium, S. B. Montgomery[1]; [1]Stanford Univ., Stanford, CA, [2]Clemson Univ., Greenwood, SC

Abstract:

Two-thirds of FDA-approved drugs in the past decade are supported by genetic evidence, highlighting the capability of genetics to effectively link molecular mechanisms to drug targets for various diseases. Alzheimer's Disease (AD), a leading cause of dementia worldwide, currently has treatments that only alleviate symptoms without addressing the underlying disease. Our objective is to utilize transcriptome-wide association studies (TWAS) and colocalization analyses to estimate the genetically regulated gene expression associated with AD risk from GWAS and eQTL data. By integrating this information with the Connectivity Map (CMap) perturbation dataset, we aim to identify opportunities for drug repositioning and discover new therapeutic approaches for AD.
We performed TWAS analysis using an AD GWAS and eQTL data of 13 brain tissues from GTEx. We focused only on genes showing colocalization between AD risk loci and gene expression traits, which resulted in 67-92 genes for each tissue. From the CMap database of central

nervous system cell lines and monocytes, we identified compounds capable of reversing these genetically-predicted gene expression change patterns based on gene set enrichment analysis (GSEA) and Spearman rank correlation analysis. Some of these compounds are used to treat other diseases but have literature evidence suggesting a correlation with AD symptoms, e.g. Ibrutinib. Enrichment analysis revealed that these compounds have diverse mechanisms of action, implicating multiple pathways for the treatment of AD. Additionally, specific combinations of tissues and cell lines produced distinct results, motivating the use of datasets from cell types most relevant to AD.
As part of the FunGen-AD Project, we plan to incorporate eQTL data from single-cell RNA-seq data from six cell types in the ROS/MAP dataset as well as from four bulk datasets from brain and blood. We will apply advanced TWAS methods developed for FunGen-AD protocol to build a comprehensive pipeline for integrating genetically regulated expression effects associated with AD risk with drug perturbation datasets. We also plan to extend this approach to other diseases to further demonstrate its value in identifying novel therapeutic compounds in many disease contexts.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7013W: Large language models enable accurate and fully automated curation of discrete molecular pathway data

Authors:

C. Deisseroth[1], Z. Liu[1], H. Zoghbi[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. of Med./HHMI/NRI, Houston, TX

Abstract:

Many monogenic disorders are treated at the symptomatic level, with no therapy that directly targets the underlying cause. When no known treatment targets the causative gene directly, finding alternate treatments that correct the imbalance requires a deep understanding of the underlying molecular pathway. We previously developed an automated tool, PARsing ModifiErS via Article aNnotations (PARMESAN), which uses natural language processing to identify discrete claims of regulatory relationships among gene products and small molecules in the biomedical literature. It prioritizes putative relationships based on the amount of supporting and opposing evidence, and its confidence scores have been shown to correlate strongly with its accuracy. PARMESAN currently uses PubTator's determination of genes and drugs mentioned in an article (a task referred to as Named Entity Recognition, or NER), and a simple rule-based formula to identify the reported regulatory relationship between them, if there is one (referred to as

Relation Extraction, or RE). These methods can be limiting, as the former cannot be re-run by end users, and the latter discards everything that does not fit a specific format. Transformers, the current state of the art in natural language processing, can likely perform NER and RE more effectively. We fine-tuned four publicly available transformers (BioMedLM, DistilBERT, BioBERT, and BioMedBERT) for NER, targeting genes and drugs. BioBERT and BioMedBERT have been previously benchmarked for NER and RE, but we sought to assess their practicality for our purposes by comparing them to a larger (BioMedLM) and a smaller (DistilBERT) model. We used a public dataset called DrugProt for training, and a similar one called BioRED for testing. In our hands, BioMedBERT had the highest precision and recall for detecting genes and chemicals. We also fine-tuned these models for RE, using DrugProt (for sentences with relations) and BioREX (for sentences without relations, different from BioRED) as a ground truth. BioMedBERT typically had the strongest performance, although for all of our RE accuracy metrics, the three BERT models performed within 4% of one another, so DistilBERT—the fastest of the three—might be the most practical to use for RE. Due to PARMESAN's dependence on the literature, it is less effective in predicting modulators of less-studied genes. Currently, more than half of the protein coding genes still do not have a genetic regulator prediction at over 90% confidence. We anticipate that integrating BERT models into PARMESAN's prediction pipeline will allow it to predict potential treatments for a larger number of monogenic disorders.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7014W: Mitogen-activated protein kinase kinase 6 (MAP2K6) gene polymorphism is associated with TNF-α inhibitor treatment response in rheumatoid arthritis patients

Authors:

K. Min[1], J. Kim[1], J. Gil[1], J. Park[1], J. Han[1], J. Chung[2], K. Lee[1]; [1]Coll. of Pharmacy, Chungbuk Natl. Univ., Cheongju, Korea, Republic of, [2]Coll. of Pharmacy and Inst. of Pharmaceutical Sci. and Technology, Hanyang Univ., Ansan, Korea, Republic of

Abstract:

In this study, we delved into the potential impact of genetics on treatment response variability in rheumatoid arthritis (RA) patients undergoing therapy that targeted tumor necrosis factor α (TNF-α). In this study, we aimed to investigate the differences in the response to TNF-α inhibitors treatment in patients with Rheumatoid arthritis due to the genetic polymorphism of the MAP2K6 gene. MAP2K6 plays a key role in activating p38 MAP kinase in response to inflammatory cytokines or environmental stress. We enrolled 112 rheumatoid arthritis (RA)

patients from three hospitals, collecting their age, sex, disease history, medication history, and clinical information. We then stratified them into responders and non-responders based on their Disease Activity Score 28 (DAS28) using The European Alliance of Associations for Rheumatology (EULAR) response criteria. Twenty SNPs(rs9905795, rs1858970, rs2440134, rs2159041, rs12172839, rs917802, rs72861370, rs8080760, rs2053597, rs9302938, rs73994816, rs8078890, rs12941198, rs9910398, rs2715834, rs2716208, rs2365494, rs2716205, rs7218416, rs2716213) were selected in consideration of the linkage disequilibrium(LD) relationship and minor allele frequency with the existing literature. After separating the genomic DNA from patients' whole blood, SNP analysis was performed with TaqMan genotyping assay. Following the SNPs selection, we conducted statistical analyses, which included Student's t-test, Chi-square test, and multivariate logistic regression analysis. Statistical analysis was conducted using SAS Enterprise Guide version 9.4. As a result, rs2716205 of the TNF receptor-associated factor gene was significantly associated with responders of RA. Treatment response was less with CC-allele carriers when compared to T allele carriers. Statistically significant variations in MAP2K6 expression according to the genotype of SNP rs2716205 were demonstrated by expression Quantitative Trait Loci(eQTL) analysis (p-value: 5.69e-12). This study identified the MAP2K6 gene polymorphism as a potential biomarker associated with the treatment response of TNF-α inhibitors in patients with RA.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7015W: N-Acetylcysteine ameliorates loss of the electroretinogram b-wave in a Bardet-Biedl Syndrome Type 10 mouse model

Authors:

T. Rankin, S. Mayer, J. Laird, B. Lobeck, E. Kalmanek, A. V. Drack; Univ. of Iowa, IOWA CITY, IA

Abstract:

Bardet-Biedl Syndrome (BBS) is a rare autosomal recessive disorder characterized by retinal degeneration leading to blindness. This study investigates the therapeutic efficacy of N-Acetylcysteine (NAC), an oxygen free radical scavenger, in ameliorating retinal degeneration associated with BBS using a murine model of BBS10. BBS is caused by bi-allelic mutations in one of multiple BBS genes, the protein products of which are involved in ciliary function; mutations in the *BBS10* gene disrupt the assembly of the BBSome protein complex, disrupting ciliary trafficking and leading to photoreceptor cell dysfunction and death. Photoreceptor function can be assessed using the electroretinogram (ERG), optical coherence tomography (OCT), and

histology to demonstrate progressive degeneration over time. This study utilizes *Bbs10*$^{-/-}$ mice to assess the effect of NAC supplementation. Results reveal that NAC supplementation ameliorates the progressive degeneration of the retinal outer nuclear layer (ONL) on OCT and histology and mitigates the no b-wave (nob) ERG phenotype observed in *Bbs10*$^{-/-}$ mice. The ERG b-wave is generated by retinal bipolar cells after synapsing with photoreceptors that have been hyperpolarized by light exposure. Reducing the nob phenotype may indicate improved synaptic function. Synaptic staining demonstrates a correlation between the absence of an electropositive b-wave and mislocalized synapses, highlighting the significance of synaptic integrity for retinal function. Our data raises the question of whether BBS10 may play a direct role in the synapse as well as in photoreceptor health and maintenance. These findings suggest NAC as a promising therapeutic intervention for managing BBS10-related retinal degeneration by mitigating oxidative stress by supporting retinal synaptic function.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7016W: Patterns of New Pharmacogenomic Drug Prescriptions and Socioeconomic Disparities in Glasgow: Insights from Longitudinal Data Analysis

Authors:

Q. Tran[1], S. Lip[1,2], I. Frater[3,2], J. Friday[3], C. Du Toit[3], J. Pell[4], J. Lewsey[4], D. Mackay[4], R. Dundas[4], D. Brown[4], F. Ho[4], C. Hastie[4], M. Fleming[4], C. Geue[4], A. Stevenson[4], L. McCallum[1,2], S. Padmanabhan[1,2]; [1]Sch. of Cardiovascular and Metabolic Hlth., Univ. of Glasgow, Glasgow, United Kingdom, [2]Queen Elizabeth Univ. Hosp., Glasgow, United Kingdom, [3]Univ. of Glasgow, Glasgow, United Kingdom, [4]Sch. of Hlth.and Wellbeing, Univ. of Glasgow, Glasgow, United Kingdom

Abstract:

Background: The prevalence of pharmacogenetic (PGx) interactions in real-world populations, and the influence of socioeconomic factors on PGx implicated medications, remain under-explored. This study aims to understand the patterns of new PGx drug prescriptions and investigate potential socioeconomic disparities in a large, diverse patient population in Glasgow. The social determinants of health are known to impact general healthcare, including prescribing, and their pharmacogenomic implications remain unclear.
Methods: Routinely collected health administrative data of patients across Greater Glasgow and Clyde through NHSGGC SafeHaven was interrogated to form the study cohort. Dispensed prescriptions were available from the Prescribing Information System (PIS). Socioeconomic status was available through the Scottish Index of Multiple Deprivation (SIMD, SIMD1: most

deprived to SIMD10: least deprived). All adults aged ≥51 years on 1st January 2012 were included, with relevant variables: demographics, SIMD decile, number of prescribed drugs, and prevalence of multiple long-term conditions (MLTCs) were extracted. Patients were categorized based on their exposure to PGx drugs before and during 2015: (1) No PGx drug before or in 2015, (2) No PGx drug before 2015, but new PGx drug(s) in 2015, (3) PGx drug(s) before 2015, but none in 2015, (4) PGx drug(s) both before and in 2015.
Results: Screening contained 471,528 patients (age (mean±SD): 59.6±11.7 years; females: 53.3%). At the start of 2015, 439,603 patients were alive and had not migrated, forming the study cohort. 64% of patients were exposed to at least one PGx drug across the study. Of these, 28% received PGx drugs both before and in 2015, 35% had exposure before 2015 but no new PGx drugs in 2015, and 3% received their first PGx drug in 2015. Socioeconomic analysis revealed that 21% of patients were in the most deprived group (SIMD1), while 20% were in the least deprived 2 deciles (SIMD 9 and 10). A higher proportion of the least deprived patients (3.5%) started new PGx drugs in 2015 compared to the most deprived patients (2%). Conversely, among those already prescribed PGx drugs before 2015, 24% of the least deprived patients started new PGx drugs, compared to 30% of the most deprived.
Conclusion: This large-scale study highlights significant exposure to PGx drugs among patients in Glasgow, with notable socioeconomic disparities. The findings suggest a potential for further investigation into disparities in access to PGx-guided therapy and underscore the importance of integrating socioeconomic factors into personalized medicine strategies.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7017W: Retrospective study shows pharmacogenomic testing could reduce adverse events associated with clopidogrel use by 38% ⭐

Authors:

N. Washington[1], N. Telis[2], D. Judge[3], P. Pawloski[4], J. Grzymski[5], A. Chahal[6], C. Hajek[2], T. Kruisselbrink[7], E. Cirulli[2]; [1]Helix, Albany, CA, [2]Helix, San Mateo, CA, [3]Med. Univ. of South Carolina, Charleston, SC, [4]Hlth.Partners, Minneapolis, MN, [5]Univ. of Nevada, Reno Sch. of Med., Reno, NV, [6]Univ. of Pennsylvania, Philadelphia, PA, [7]Helix, San Diego, CA

Abstract:

Introduction: In recent years, genetic interactions with drugs have been demonstrated to alter drug metabolism. Efforts such as PharmGKB and the Clinical Pharmacogenomics Implementation Consortium (CPIC) have developed recommendations to alter the use of several commonly prescribed drugs based on patient genotype. However, most prescriptions are still

given without pharmacogenomic testing.
Objective: Demonstrate the real world implications of the use of pharmacogenomic testing through a retrospective study of clopidogrel prescriptions given without testing.
Methods:We analyzed ~100K individuals with paired medical records and clinical grade exome sequencing data, from population health studies administered at multiple medical centers. We genotyped all individuals for CYP2C19 star alleles. We identified daily dosage of clopidogrel by processing the freetext prescription “sig” field with GPT-4. Then, we identified all instances of individuals with at least one prescription that is not in concordance with the CPIC clopidogrel use guidelines based on their CYP2C19 genotype (Bousman et al. 2023). We identified instances of thrombosis using a comprehensive codeset based on the medical records.
Results:25% of individuals with a calculable daily dose of clopidogrel have a mismatch between the recommended clopidogrel dosage guideline based on their CYP2C19 genotype and their prescribed dose. 12% of these mismatched individuals are poor metabolizers (PM), who should not use clopidogrel at all. The remaining 88% are intermediate metabolizers (IM). PMs and IMs receiving clopidogrel are much more likely to experience thrombosis than other metabolizers. 25% of PMs experienced thrombosis after the initiation of clopidogrel, with 40% of these occurring in the first two months (vs normal metabolizers, binomial p-value = 0.001).
Conclusions:As expected from a lack of testing and a high population frequency of pharmacogenomic variants, many patients are prescribed doses of clopidogrel that are too high given their subsequently derived pharmacogenomic information. There is at least a 38% excess of adverse events (as measured by thrombosis) in this group. At minimum, this testing could prevent 1 thrombosis event per every ~30 people prescribed clopidogrel, demonstrating tangible benefits of population-based pharmacogenomic testing.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7018W: Searching for functional regulatory SNPs that regulate *CYP2C9* expression to improve *CYP2C9* biomarker panels

Authors:

A. Montalvo Lopez de Victoria[1], J. Collins[2], D. Wang[3]; [1]Univ. of Florida Coll. of Pharmacy, Gainesville, FL, [2]Univeristy of Florida, Gainesville, FL, [3]Univ. of Florida, Gainesville, FL

Abstract:

CYP2C9 metabolizes ~25% of clinically administered drugs. Several non-synonymous single nucleotide polymorphisms (SNPs) of *CYP2C9* (e.g., *2, *3, *8) are used as biomarkers to predict CYP2C9 activity, but large variability in *CYP2C9* expression remains unexplained. An additional

distal non-coding SNP upstream of *CYP2C18*, rs12777823, was identified in a GWAS study as being associated with decreased warfarin dose requirements in African Americans. The Clinical Pharmacogenetics Implementation Consortium (CPIC) recommends using rs12777823 as a biomarker for warfarin dose adjustment, specifically in African Americans. However, evidence suggests that rs12777823 is a tag SNP, and its association with warfarin dosage may be caused by its high LD with a currently undiscovered functional SNP. This study aims to improve *CYP2C9* biomarker panels by identifying and characterizing the functional SNP(s) underlying the association between rs12777823 and warfarin. For this, we measured allelic expression imbalance (AEI) of *CYP2C9* in liver samples derived from African American donors and used targeted long-read sequencing to scan the entire *CYP2C locus* in AEI positive (n=7) and AEI negative (n=9) samples. Long-read sequencing revealed >10K SNPs per sample across the ~600 kb *CYP2C locus*, many of which are in high LD. Correlating AEI status with SNP heterozygosity identified 117 candidate SNPs. We identified 25 potential enhancer regions across the *CYP2C locus* using publicly available functional genomics data. We aligned the candidate SNPs with these potential enhancer regions to prioritize SNPs that fall in regulatory regions, resulting in 29 SNPs in 14 regions. One candidate SNP was significantly associated with hepatic CYP2C9 expression. The potential enhancer region is being interrogated with CRISPRi assays in hepatic cell lines to validate its function and the function of the SNP will be tested using reporter gene assays. The clinical impact of identified functional SNP will be tested for association with warfarin doses in patients taking warfarin. Using this approach, we expect to identify the functional regulatory SNP or haplotype driving the association between rs12777823 and warfarin dose. The identified functional SNP is expected to improve *CYP2C9* biomarker panels for warfarin and other medications metabolized by CYP2C9 across populations.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7019W: Strong Heart Study: Serum nicotine metabolite ratio (NMR) biomarker informs nicotine metabolism variation in American Indians who smoke commercial cigarettes

Authors:

C. Musser[1], C. A. Sherman[1], K. R. Ferrier[1], K. Aagaard[1], T. H. Wroblewski[2], R. F. Tyndale[3], M. J. Chenoweth[3], N. Franceschini[4], S. A. Cole[5], J. G. Umans[6], K. G. Claw[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [3]Univ of Toronto, Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [5]Texas BioMed. Res. Inst, San Antonio, TX, [6]Georgetown Univ., Washington, DC

Abstract:

Background. This retrospective study associated individual variation in CYP2A6 haplotypes to a phenotypic biomarker of CYP2A6 activity, the serum nicotine metabolite ratio (NMR), within the Strong Heart Study (SHS) representing American Indians (AI) from 12 Tribal Nations living in Arizona (AZ), Oklahoma (OK), and North and South Dakota (DK). Nicotine is primarily metabolized in hepatocytes by the CYP2A6 enzyme into metabolites that circulate through the bloodstream, and can be extracted from plasma and analyzed to reliably quantify NMR. We hypothesized that NMR in AI differs between CYP2A6 haplotype, biological sexes, body mass index (BMI), and cigarette smoking behavior subgroups.
Methods. Six nicotine metabolites were extracted from 790 SHS phase 2 self-reported "smokers" and 30 "non-smokers" (10 controls per site) plasma via tandem liquid chromatography mass spectrometry (LC-MS/MS). NMR values were determined from each participants' observed serum concentration ratio of 3'-hydroxycotinine to cotinine analytes and log-transformed for normalization. Phenotypic activity scores based on CYP2A6 diplotypes were created and associated with individual NMR, including variables such as tribal groups, BMI, and smoking behavior in AI subpopulations.
Results. After exclusions, 4,386 quantifiable LC-MS/MS nicotine metabolites represented a final sample size of 731 participants, with 6% from AZ, 40% from OK, and 54% from the DK regions. Preliminary analysis revealed an average (+/- SD) cohort NMR value of 0.629+/-0.539 (ranged 0.008-7.231), which is higher than the previously reported serum NMR value for Alaskan Natives (0.50+/-0.33). Linear regression models showed trends similar to previous studies with AIAN populations, such as an increased average NMR in females (0.658+/-0.473) compared to males (0.595+/-0.604), likely due to estrogen interaction decreasing excretion (P=3.65e-05). Chi-square tests in our smoker population reinforced the known association of higher BMI categories with increased NMR, likely due to higher blood volume (P=1.01e-07). A strong association was observed between slight decrease-of-function CYP2A6 diplotypes (particularly CYP2A6 *1/*9) with a decreased serum NMR (P=4.14e-14). Additionally, NMR values were shown to be reliably determined retrospectively from 30+ year-old samples.
Implications. This evidence-based approach showed a relationship between sex and CYP2A6 haplotype with serum NMR in AI smokers. We are currently investigating other predictors of NMR and associations with smoking behaviors. Importantly, this work will inform future studies around personalized smoking cessation.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7020W: Structural Equation Modelling of Polygenic Risk and Environment in Late-Life Depression

Authors:

S. Elsheikh[1], V. Marshe[2], J. Kennedy[3], G. Pare[4], D. Felsky[1], D. Mueller[5]; [1]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [2]Columbia Univ. Irving Med. Ctr., New York, NY, [3]CAMH, Univ of Toronto, Toronto, ON, Canada, [4]McMaster Univ., Hamilton, ON, Canada, [5]Univ of Toronto, Toronto, ON, Canada

Abstract:

Our understanding of the interplay between genetic and environmental factors (Gene x Environment Interaction, or GxE) determining mental health disorders has improved through the proliferation of genome-wide interaction association studies (GWIAS) and targeted GxE analyses.Moreover, multivariate modelling approaches, such as structural equation modelling (SEM) and polygenic risk scores (PRS), offer opportunities for the integration of clinical and genome-wide genotype data in building improved biopsychosocial models of mental illness aetiology and their response to treatment.

We propose to construct a SEM framework to uncover the inter-correlation and directed structure of mental health phenotypes by leveraging the joint predictive capacity of PRS for comorbid traits that share underlying biological and environmental risk pathways. The proposed model will be capable of linking latent constructs to their observed measurements; these will include disease severity, comorbidities and clinical histories, and behaviours and lifestyle factors such as physical and social activity.

Our gene-by-environment SEM (GESEM) will be initially developed and tested using four well-characterized clinical cohorts for older adults diagnosed with late-life depression and treated with antidepressants (CAN-BIND, IRL-GREY, STOP-PD II and IMPACT; n =1,238). The primary outcome will be antidepressant remission. Multiple PRS will be calculated to capture underlying genetic risk across vulnerable pathways which contribute to comorbidities. This selection will be made based on new, largely unpublished work from our group on the impact of PRS and targeted GxE studies on psychiatric outcomes across the lifespan. Each PRS will be calculated using both clumping and thresholding (PRSice-2) and continuous shrinkage (PRS-CS-auto) methods across selected cohorts using well-powered publicly available GWAS summary statistics. The multilevel GESEM model will include interactions between symptoms and comorbidities (i.e., observed measurements), which are caused by unobserved factors (i.e.,latent constructs), and are subject to modification by background PRS. We will compare our GESEM model against existing SEM-based approaches to GxE, including local SEM (LOSEM).

An open-source R package of the analytical code will be created and shared with the research

community. This work has the potential to improve upon existing PRS-based predictive models in a clinical setting.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7021W: TAS-PGx: Targeted Adaptive Sampling-Long Read Sequencing for Enhanced Pharmacogenomics Profiling and Genome-Wide Variant Analysis

Authors:

M. Gonzalez Porta[1], P. H. Gan[1], A. Q. Ng[1], M. Bin Hajis[1], Y. Maulana[1], H. Yeo[1], W. Sim[2], C. Lim[2], A. Irwanto[1], L. Sani[1], E. Lo[3], L. Goh[2]; [1]Nalagenetics Pte Ltd, Singapore, Singapore, [2]Tan Tock Seng Hosp., Singapore, Singapore, [3]Natl. Univ. Hosp., Singapore, Singapore

Abstract:

Background & Objectives:
Genetic variants affecting drug therapy are widespread, impacting over 90% of the population and potentially leading to adverse drug reactions (ADRs). Pre-emptive pharmacogenomics (PGx) testing has demonstrated a 30% reduction in ADRs, resulting in decreased healthcare costs and improved patient outcomes. However, routine genotyping methods such as short-read sequencing and microarray genotyping face challenges in haplotype phasing and adequately resolving complex events. Long-read sequencing holds promise in improving the resolution of fully phased diplotypes, leading to more accurate phenotype prediction. In particular, adaptive sampling from Oxford Nanopore Technologies (ONT) emerges as a flexible and cost-competitive solution, since it offers in-silico driven enrichment of selected regions while generating a low-coverage signal across the entire genome.
Method(s) and Results:
Here, we assess long-read sequencing for PGx testing based on adaptive sampling. Multiple experiments were conducted using reference materials from various sources (GIAB, GeT-RM and 1KGP), as well as clinical samples previously challenging to characterize with genotyping. Our findings demonstrate that our workflow can accurately identify the expected diplotypes for Tier 1 pharmacogenes from PharmVar and elucidate complex haplotypes, including CYP2D6 hybrids. Additionally, we show off-target reads can be used to impute common SNVs genome-wide with high accuracy.
Conclusions:
Our TAS-LRS workflow effectively resolves diplotypes in actionable pharmacogenes, including complex structural and hybrid rearrangements, which were previously uncharacterized with

other platforms. Moreover, it enables research applications by efficiently detecting genome-wide variants and methylation signals. By leveraging multiplexing, it offers great opportunities for accurate PGx testing at a lower cost.

Session Title: Pharmacogenomics Poster Wednesday Session

Board 7022W: Full coverage of Tier 1 Variants for Pharmacogenomics Studies

Authors:

N. Ramalingam[1], J. Qin[2], M. Gonzales[1], M. Bullock[1], R. Kung[1], H. Ren[3], A. Khanna[2], W. Hwang[1], A. Barican[3], G. Harris[3], T. Goralski[3], J. Alipaz[3], C. Park[1], D. King[3]; [1]Standard BioTools Inc, South San Francisco, CA, [2]Standard BioTool Inc, South San Francisco, CA, [3]Standard BioTools, South San Francisco, CA

Abstract:

Background/Objectives: Pharmacogenomics (PGx) testing uses individual genetic variations to enhance drug effectiveness and reduce toxicity. The Association for Molecular Pathology (AMP) PGx Working Group has classified essential genes/alleles for testing as Tier 1. However, many platforms cannot fully test these Tier 1 alleles due to technological limitations. Specifically, the CYP2D6*29 allele is characterized by two missense variants, NM_000106.6:c.1012G>A (p.Val338Met, rs59421388, legacy 3183G>A) and c.406_408delinsATC (p.Val136Ile, rs61736512 + rs1058164, legacy 1659G>A and 1661G>C). Most technologies only test for c.1012G>A, as detecting c.406_408delinsATC is challenging due to its proximity to other variations, which are adjacent and separated by a single base pair. We developed novel qPCR-based assays to comprehensively cover Tier 1 variants, including CYP2D6*29 and six other genes. **Methods:** The PGx Assay Panel includes assays for single-nucleotide polymorphisms (SNPs) and copy number variations (CNVs). DNA samples are subject to a targeted enrichment step through preamplification and analyzed on an integrated fluidic circuit (IFC) using the Biomark™ X9 System. This setup allows simultaneous interrogation of 96 samples against 96 assays in nanoliter volumes. Analysis and interpretation of SNP genotyping and CNV determination is performed using our data analysis software. **Results:** We validated this panel using Coriell DNA samples with known genotypes and copy numbers. The observed average call accuracy was approximately 99.9% for SNPs and approximately 98% for CNVs. **Conclusion:** Our study presents a robust, automated qPCR-based workflow that successfully integrates SNP and CNV analyses, ensuring full coverage of AMP-recommended Tier 1 variants. This comprehensive approach not only streamlines genetic testing but also enhances the accuracy and applicability of PGx testing

in testing settings, potentially improving patient outcomes through more precise and personalized medicine.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7025W: Amplicon-based next generation sequencing reveals low-level germline mosaicism for CNV deletion -discordance between somatic and germline mosaicism

Authors:

F. Chow[1], K. Lam[2], W. Yeung[3], H. Ng[1]; [1]Dept. of Obstetrics and Gynaecology, Sch. of Clinical Med., LKS Faculty of Med., The Univ. of Hong Kong, Hong Kong, Hong Kong, [2]Dept. of Obstetrics and Gynaecology, Queen Mary Hosp., Hong Kong, Hong Kong, [3]Shenzhen Key Lab. of Fertility Regulation, The Univ. of Hong Kong-Shenzhen Hosp., China, China

Abstract:

Introduction: Copy number variants (CNVs) that arises as post-zygotic errors during cell divisions can result in a fraction of cells in the body contain a CNV, leading to mosaicism. Retrospective analysis of asymptomatic parents' DNA showed that a proportion of de novo variants detected in probands are indeed inherited from parents with low-level somatic mosaicism. Here, we reported 2 prospective cases in which low-level germline mosaicism for CNV deletion was detected in sperms of carrier husbands.**Materials & Methods**: In couple 1, husband was found to be heterozygous for a pathogenic NF1 deletion: 17q11.2(29546023_29654857)x1 by whole exome sequencing (WES). In couple 2, husband was found to be a carrier of at least 4 pathogenic gene deletions on chromosome 20, determined by expanded carrier screening on saliva, spanning approximately 9.2 Mb (20q11.23-q13.12). Both families opt to have preimplantation genetic testing for monogenic disease (PGT-M). **Results**: During pre-PGT workup of couple 1, STR markers located inside the deleted region were unexpectedly found to be heterozygous with a minor allele. High resolution mapping of deletion junctions was performed by nanopore sequencing and gap-PCR primers were designed. PCR on sperms revealed a weak band, indicating germline mosaicism for the deletion. To estimate the level of mosaicism, we calculated the proportion of two alleles of a heterozygous SNP in a deletion. By amplicon-based NGS, 2 heterozygous SNPs were analysed, with read depth of >10,000x. The mosaicism level of blood was determined as 40%, whereas low-level germline mosaicism of 12% was determined on sperms. Due to the low-level germline mosaicism, couple decided not to

proceed to PGT-M.In couple 2, chromosome microarray analysis (CMA) of the blood revealed a 14.09 Mb mosaic deletion in 20q11.23-q13.13 (30-40%). Low-pass NGS was then carried out and mosaicism level in the saliva and the blood was determined to be 28% and 21% respectively. No deletion was detected in sperms. To identify low-level mosaicism for deletion in sperms, amplicon-based NGS analysis is in progress.**Discussion**: Mosaicism of CNVs present in different samples of an individual, especially at low levels, can be challenging to detect by standard laboratory techniques such as CMA, WES, low-pass whole genome NGS or Sanger sequencing. Amplicon-based NGS offers a more sensitive assessment of the potential recurrent risk of a pathogenic variant. In this report, discordance between somatic and germline mosaicism was demonstrated, illustrating low-level of mosaicism on sperms, which has significance in genetic counselling of prospective parents.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7026W: Autosomal dominant mutations in the GEMINI male infertility cohort: Detection rates and genetic burden

Authors:

H. Castillo-Madeen[1], A. Riera Escamilla[1], A. Akbari[1], C. Peaslee[1], Genetics of Male Infertility Initiative (GEMINI) consortium, K. Vigh-Conrad[1], K. Aston[2], D. Conrad[1]; [1]Oregon Hlth.and Sci. Univ., Portland, OR, [2]Univ of Utah, Salt Lake City, UT

Abstract:

Non-obstructive azoospermia (NOA) is one of the more severe forms of male infertility, however many cases remain unexplained. In previous GEMINI analyses, we have focused on recessive forms of NOA because identifying autosomal dominant (AD) mutations is challenging due to the high background of deleterious heterozygous mutations in all human genomes. Statistical burden testing of counts within cases and controls is a powerful way to detect AD gene mutations in larger cohorts. Using burden testing, we previously identified an excess of rare AD *DMRT1* variants in our combined male infertility cases (n=14), compared to normozoospermic controls (n=0, Fisher's Exact test $p<0.05$) as well as compared to a larger set of population controls from gnomAD v4.1.0 ($p<0.05$) (manuscript in review). With this result and method success, we performed burden testing for AD genes on the GEMINI cohort first using published AD NOA genes and AD de novo genes linked to male infertility in humans. To identify causative AD gene variants from the list of known AD and AD de novo genes, a computational pipeline was developed to filter variants and ACMG guidelines were followed to determine variant pathogenicity. Statistical burden testing was performed at the GEMINI cohort

level (filtered: n=1474 male cases, n=633 male controls) as well as with the population controls from gnomAD v4.1.0 with similar filtering (n=807162). We found significant burden in 14 AD NOA genes (3.9% of cases, p=0.034) and 22 AD de novo NOA genes (7.1% of cases, p=9.3e-05) in cases compared to controls. *RASAL2* (AD de novo) is the only gene with significant burden (p=0.031) in GEMINI cohort burden testing (n=16 cases, n=1 control), also reaching significance in burden testing with gnomAD controls (n=15091, p=0.026). Several other genes were also significant at the gnomAD burden level but not at the GEMINI cohort level. We also found AD de novo LoF variants to be associated with lower pLI scores (<7) and non-LoF variants to be associated with higher pLI scores (>7) (p=0.031). Continued analysis of the variants within these genes is required to determine the effect of the gene and variants on male infertility. The addition of the AD disease model will increase our detection rate of causative variants in the GEMINI infertility cohort as it includes genes in a disease model we would not have considered in prior analyses. With this filtering and analysis framework set up for AD in our data we can broaden our search of AD genes to those not yet linked to male infertility. Expanding our search of AD genes and mutations to other large infertility cohorts will also further increase our understanding of AD gene mutation burden in male infertility cases.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7027W: Cilia mutation and gene by environment interactions leading to increased phenotypic variance.

Authors:

R. Green[1], C. Waronsky[1], J. Devine[2], N. Young[3], R. S. Marcucio[1], B. Hallgrimsson[4]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]KU Leuven, Leuven, Belgium, [3]Univ. of California San Franscisco, San Fransisco, PA, [4]Univ. of Calgary, Calgary, AB, Canada

Abstract:

Primary cilia are responsible for mediating cell signaling pathways related to embryonic development. Mutations in genes which pattern the primary cilia have been implicated in neural tube defects and genetic syndromes which include craniofacial differences. In general, folic acid supplementation modulates neural tube defects and has been shown to have a strongly protective effect. However, the mechanism for this is unclear. In some animal models of primary cilia defect, increasing folic acid in a multigeneration fashion has led to increased severity of defects and increased fetal resorption rates. However, there has been very little study on cilia mutations which affect craniofacial development and any interactions with folic acid. *B9D1* is a ciliary transition zone protein whose loss leads to a range of craniofacial defects in the mouse

including bilateral cleft lip and palate and narrowing of the midline. Mice and embryos *B9D1*$^{+/-}$ breeding pairsfrom were exposed to a fortified or control folic acid diet for 3 generations. The third generation embryos were examined for morphological difference (microCT and geometric morphometrics), DNA methylation difference and gene expression difference. We observe significant effects of genotype and small but significant effects for both diet and gene x diet interactions. We were further able to identify methylation changes in several genes that have known roles in folate metabolism and cranial development. Together, these results highlight a potential mechanism for gene by environment interactions with folate and how these might relate to highly variable genotype phenotype maps.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7028W: Concomitant Prenatal Diagnoses of Jeune, Malan, and Sotos Syndromes in a Singleton Pregnancy

Authors:

K. Wang[1], Y. Yuk[1], E. Sperry[2], B. Wong[2], J. Berton[2], G. Hijazi[2], Y. Wu[3], A. Shillington[4], R. Hopkin[5]; [1]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [2]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [3]Cincinnati Children s Hosp. Med. Ctr., Cincinnati, OH, [4]Cincinnati Children s Hosp., Cincinnatti, OH, [5]Children's Hosp Med Ctr, Cincinnati, OH

Abstract:

We present a case of a Guatemalan male patient born at term (39w0d) following concurrent prenatal genetic diagnoses of Jeune, Malan, and Sotos syndromes. Prenatal ultrasound imaging showed echogenic bowel and kidneys, shortened long bones, and bilateral talipes equinovarus. Follow-up fetal MRI revealed a bell-shaped thorax with mildly reduced total lung volumes. Prenatal genetic testing revealed biallelic pathogenic variants in *DYNC2H1* (c.4047G>A, p.Trp1349* and c.7442G>A, p.Arg2481Gln) and *de novo* heterozygous pathogenic variants in each of *NFIX* (c.346C>T, p.Arg116Trp) and *NSD1* (c.4072C>T, p.Gln1358*). Post-natal physical exam was significant for clinical features of Jeune syndrome (narrow thoracic cage, respiratory insufficiency), Malan syndrome (macrocephaly, frontal bossing), and Sotos syndrome (macrocephaly). The patient passed away from respiratory failure due to thoracic insufficiency on day 3 of life. As this patient's physical features and clinical course were consistent with Jeune syndrome and contributions from one or both of Sotos and Malan syndromes, this unique case highlights potential embryologic interactions between opposing growth syndromes. Notably, this infant's total lung volumes were higher than typically observed in isolated cases of short rib thoracic dystrophies and he exhibited relative overgrowth of his extremities. Together, these

clinical findings are suggestive of complex and partially temporizing developmental interactions between his pathogenic variants at all three loci. We further discuss the importance of thorough prenatal counseling, including discussion of perinatal survivability, along with complex neurodevelopmental symptoms that would be expected to arise in a patient with these three genetic diagnoses.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7029W: Epigenetic regulator *Tet2* modulates ovarian dysfunction associated with Fragile X premutation

Authors:

K. Shelly[1], J. Merkison[1], S. Call[1], E. Allen[1], P. Jin[2]; [1]Emory Univ., Atlanta, GA, [2]Emory Univ Sch Med, Atlanta, GA

Abstract:

Roughly 20% of women who carry a premutation (PM) allele are diagnosed with Fragile X-associated Primary Ovarian Insufficiency (FXPOI). This is driven by the expansion of a CGG-repeat tract in the 5' untranslated region (UTR) of the *FMR1* gene to 55-200 repeats. Mouse models with ectopically expressed tracts of CGG_{99} repeat-producing expanded repeat RNA and polyglycine peptide (FMRpolyG+RNA) have shown impaired hormonal response and ovulation. Follicular maturation is perturbed through reduced cumulus expansion, decreased meiotic resumption of the oocyte, and, ultimately, reduced numbers of cumulus-oocyte-complexes (COC) released. Aging FMRpolyG+RNA females manifest defects as anovulation and ovarian insufficiency by 4 months. These data show that ovarian function, particularly steroidogenesis and ovulation, is perturbed in PM ovaries. Transcriptome data from human FXPOI PBMCs reveals markers of ovulation are altered in FXPOI women, and untargeted metabolomics of human samples found 2-oxoglutarate, a cofactor for Tet2, as an altered metabolite in women with FXPOI. Given that *Tet2* regulates ovulation via demethylation of key genes *Lhcgr*, *Areg*, and *Ereg* in granulosa cells (GCs) at ovulation, we asked whether *Tet2* haploinsufficiency could modulate ovulation-dependent deficits in our PM mouse model. PM mice were bred with *Tet2* heterozygotes. Trans-heterozygote females do not have significantly decreased fertility from controls. Half of FMRpolyG+RNA; $Tet2^{+/-}$ females remain fertile beyond 6 months of age, whereas all FMRpolyG+RNA females exhibit POI, which suggests partial rescue of ovarian function. Histological analyses support the rescue of ovulation in fertile females, as corpora lutea are present in FMRpolyG+RNA $Tet2^{+/-}$ ovaries but are absent in FMRpolyG+RNA ovaries. The hormone challenge of females enables direct assessment of ovulatory processes. These

data show that expansion of cumulus GCs, as measured by total COC diameter, is rescued in FMRpolyG+RNA; *Tet2*$^{+/-}$ females, while FMRpolyG+RNA females produce significantly smaller COCs. Cumulatively, these data suggest a role of *Tet2* in modulating FXPOI via its role in ovulation. Given that GCs are the only ovarian population to demonstrate dynamic expression, it indicates that *Tet2*-mediated rescue of cumulus expansion will be identified through targeted analyses of the GCs. Identifying the proximal event in the ovulatory cascade that recuses cumulus expansion and release of COCs is a critical first step toward a therapeutic intervention for women with FXPOI.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7030W: Equitable testing and screening for cystic fibrosis in patients of diverse genetic ancestry

Authors:

L. Burnett[1], M. Gruzin[1], J. Knezovic[2], S. Poll[3], A. Chen[1], N. Schonrock[4], S. Aradhya[5]; [1]Garvan Inst. of Med. Res., Darlinghurst, Australia, [2]Blue Jay Consulting Services, Brisbane, Australia, [3]Invitae Corp., San Francisco, CA, [4]Sydney, Sydney, Australia, [5]Invitae, San Francisco, CA

Abstract:

Background: Genetic carrier screening enables informed reproductive decision-making to reduce the likelihood of having a child affected by a severe genetic disorder. It can be performed using targeted panels for specific genes and variants, full gene sequencing, or other molecular techniques. Historically, carrier screening has been implemented in specific ethnic/racial (now "ancestry") groups. Peak bodies now recommend offering ancestry-neutral, universal carrier screening. We have studied whether targeted panels are likely to provide equitable clinical utility and health outcomes when applied to all genetic ancestries, using cystic fibrosis (CF, *CFTR* gene) testing in Australia as an illustrative example. Australia has recently experienced major changes in its population demographics, transitioning from a country of predominantly Western European immigrants to a heterogeneous population with large proportions of Asian, Middle Eastern and local Indigenous ancestries.
Aim: We examined shifts in Australian population demographics and assessed whether current *CFTR* testing methods meet 95% sensitivity thresholds for CF testing and screening.
Methods: Multiple data sources, including ClinVar, Clinical and Functional TRanslation of *CFTR* (CFTR2), CFTR-France, and the Australian Cystic Fibrosis Data Registry (ACFDR) were utilised to determine the prevalence of pathogenic CF variants. Alignment of these variants with Australian population demographics was assessed using gnomAD v4.1 and Australian Bureau of

Statistics (ABS) Census data. Analysis of Cystic Fibrosis European Network (CFEN) External Quality Assessment (EQA) reports identified *CFTR* assays currently in use, with assessment of their capability to detect *CFTR* variants based on in vitro diagnostics (IVD) manufacturer data. *Results*: A decade ago, 95% sensitivity for *CFTR* pathogenic variants was nearly achievable in Australia using only 25 variants. Subsequent population genetic enrichment due to diversification of constituent ancestry groups has altered the frequency of CF-causing variants. Targeted *CFTR* assays of even 100 variants no longer meet 95% sensitivity for any patient population, particularly among African, Asian, Indian or Pasifika genetic ancestries. *Conclusion*: Achieving equitable sensitivity in CF testing and screening across diverse genetic ancestries necessitates large numbers of variants to be assessed and may become unfeasible to define in genetically diverse populations. Full gene sequencing may be more appropriate, providing a more comprehensive and equitable assessment of CF carrier risk, regardless of the ancestry of the tested individual.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7031W: Exploring the genetic architecture of Unexplained Pregnancy Loss and other Adverse Pregnancy Outcomes

Authors:

A. Ikuzwe Sindikubwabo, C. Lannawill, S. Verma, Penn Medicine Biobank, Regeneron Genetics Center; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Pregnancy Loss (PL) is the most common pregnancy complication worldwide affecting 23 million people every year. While ~60% of PL are due to chromosomal abnormalities, 40% remain unexplained. The exact etiology of this Unexplained Pregnancy Loss (UPL) is not well understood, there is still no clear understanding of how UPL and other Adverse Pregnancy Outcomes (APOs) like recurrent PL and preterm birth compares to Live Births. The consequence of PL includes psychological problems like elevated chances of experiencing anxiety, depression which might be even more pronounced in cases of UPL due to a lack of knowledge regarding the cause of the PL. Furthermore, for mostly unknown reasons, Black women have higher rates of PL. In this study we leveraged exome wide sequencing data from genetically diverse women diagnosed with either UPL and APOs in the Penn Medicine BioBank (PMBB), which we then used in SAIGE-GENE to identify genes associated with UPL and APOs. Age at occurrence, first 10 principal components were used as our covariates. In this analysis, we identified missense variance in *VGLL3* gene (P-value: 7.838534e-07, Beta: 0.086719, SE: 0.017557) to be associated

with UPL in women from African ancestry. *VGLL3* (Vestigial-like-family-member 3) is a transcription cofactor, which interacts with TEAD (transcriptional enhanced associate domain) transcription factors to regulate expression of genes involved in cell proliferation, differentiation, and apoptosis. *VGLL3* is mainly expressed in the placenta and plays significant roles in immune response. Its dysregulation has been linked to autoimmune diseases like systemic lupus erythematosus (SLE) which has been shown to increase risk for pregnancy loss and other adverse pregnancy outcomes like preterm birth in previous studies. Auto-immune disorders like systemic lupus erythematous, APS, rheumatoid arthritis have already been shown to be associated with an increased risk of pregnancy loss. Our study identified a missense variant in *VGLL3* which has already been linked to systemic lupus erythematosus (SLE). Our study supports the hypothesis that for a successful embryo development it is crucial to maintain metabolic homeostasis to prevent autoimmunity.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7032W: First cases of Hemogobin J in sub Saharan Africa

Authors:

M. Gadji[1], S. Ndiaye[2], M. Soumboudou[3,2], D. Makalou[4], A. Niang[2], N. Coly[2,3]; [1]Service of Biological Hematology & Oncology-Hematology (HBOH); Faculty of Med., Pharmacy and Odonto-Stomatology (FMPOS); Univ. Cheikh Anta Diop of Dakar (UCAD), Senegal, Dakar, Senegal, [2]Diamniadjio Children Hosp. of Dakar, Senegal, Dakar, Senegal, [3]Univ. Iba Der Thiam of Thies, Senegal, Thies, Senegal, [4]Univ. Gaston Berger of Saint Louis, Senegal, Saint Louis, Senegal

Abstract:

Introduction: Since the first description of Hemoglobin J in an American Negro Family in 1956, no case were described in sub Saharan Africa. We report the first familial hemoglobin J in sub Saharan Africa. **Method:** The proband is a 22-month-old female infant, third child in a family of 3 siblings. She was admitted to the pediatric surgery for a convulsive attack in a febrile context with a temperature of 38°C plus an inflammatory and painful swelling on the anterior surface of the left thigh. A blood count, an ionogram, a blood group typing, a hemostasis assessment, a HIV serology, a renal assessment (urea, creatinin), an Emmel test, a C-reactive protein (CRP) and a cytobacteriological examination of blood and urine were performed meanwhile starting an antibiotherapy with ceftriaxone and vancomycin that has led to a successful clinically recovery of the proband. Additional examinations as iron assessment (serum iron, serum ferritin) and a hemoglobin capillary electrophoresis for suspicion of potential hemoglobinopathy. The results conducted to perform High performance liquid chromatography (HPLC) and isoelectric focusing

for confirmation of hemoglobin J. **Results: T**he proband had a blood count showing a hypochromic, microcytic anemia (Hb=9.5 g/dL; MCV=64.6 fL; MCH= 20.1pg) with an hematocrit (Ht=31.1%). The platelet count was normal (N=270103/mm^3).The leukocyte formula reveals hyperleukocytosis (N=14,000/mm^3) with neutrophils at 12,170/mm^3. Blood ionogram, hemostasis and renal assessment were within the normal reference ranges. HIV retroviral serology, as well as cytobacteriological examination remained negative. The radiological examination of the thigh was normal with an absence of bone lesion along the diaphyseal and a normal appearance of the soft tissues as well as a normal appearance of the above and underlying joints. The CRP was positive at 192 mg/L and an Hemoculture isolated a strain of *Staphylococcus aureus* resistant to methicillin. Hb capillary electrophoresis revealed the presence of HbA=42.7%, HbA2=1.9%; HbF=0.5% and an abnormal fraction migrating into zone Z13 at 54.8%, identified as hemoglobin J characterised by with HPLC a P3 peak with a retention time of 1.77 minutes. A fraction of HbA1c was also detected with a rate of 9.7% for a normal index case blood sugar level of 0.87g/L. A family investigation showed the presence of Hemoglobin J in the 33-year-old mother (52.7%) and a 7-year-old girl (54.7).The sequencing analyses is underway to decipher the exact Hemoglobin J variant in this family. **Conclusion:** Hemoglobin J remains mostly silent with clinical variability presentations between different carriers and carriers in the same family.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7033W: Gene by Prenatal Alcohol Exposure Interaction Effects on Growth and Cognition in Mother-Child Dyads in a South African Birth Cohort

Authors:

V. Ruhela[1], Z. Yang[1], S. W. Jacobson[2], J. L. Jacobson[2], E. M. Meintjes[3], G. Tosto[1], R. C. Carter[1]; [1]Columbia Univ., New York, NY, [2]Wayne State Univ., Detroit, MI, [3]Univ. of Cape Town, Cape Town, South Africa

Abstract:

Background: Fetal Alcohol Spectrum Disorders (FASD), caused by prenatal alcohol exposure (PAE), are a leading, preventable cause of neurodevelopmental delay. Despite phenotypic variability in FASD children, maternal and child genetic influences in FASD remain largely unknown. This study employs comprehensive single-marker and gene-level analyses to identify FASD-associated risk markers and loci utilizing mother-child dyad data from two longitudinal birth cohorts in South Africa, where the prevalence of FASD is among the highest in the

world. **Methods:** Illumina MegaEX genotype data from mother-child dyads were analyzed in single-marker analysis was using Mixed Model Association to study gene-environment interactions (alcohol consumption during pregnancy) on FASD outcomes, including diagnosis, working memory, recognition discrimination, and height. Covariates included maternal age, socioeconomic status, choline supplementation, and prenatal cigarette exposure. Subsequently, gene-based analysis was performed using GCTA based on single-marker summary statistics. **Results:** Separate gxe analyses of mother and child genotype data identified unique genetic biomarkers in both groups. In the single marker analysis for reduced child height, the top SNPs identified were *rs61852292* and *rs10782983* from child and maternal genotype data, respectively (beta = 4.13, p = 1.1e-6; beta = 1.19, p = 2.9e-8). In the gene-based analysis for child height, 150 genes were nominally significant in both mothers and children, including *STX6* (p = 2.06e-6/6.9e-4 (child/mother)) and *MR1* (p = 3.31e-4/9.3e-3). For recognition discrimination, 107 genes were nominally significant in both strata, including *FOXD3* (p = 3.3e-5/0.023) and *IQGAP3* (p = 1.9e-3/1.4e-5). Additionally, several gene markers in candidate methyl donor metabolism pathway genes were significant for child height reductions: *FADS1*, *FADS2* (p = 0.012), *ACTR8* (p = 5.4e-3), and *MAT2A* (p = 0.03) in children; in the mothers, *MTHFR* (p = 0.015) and *MTHFD2* (p = 0.033). **Discussion:** These findings identify novel genetic markers associated with FASD, suggesting mechanistic roles and potential therapeutic targets. Some markers are linked to neurological disorders. Methyl donor metabolism-related genes, notably affected, suggest that choline treatment could mitigate prenatal alcohol exposure effects, highlighting clinical relevance from both animal and human studies. **Conclusion:** Comprehensive single-marker and gene-based analyses of mother and child genotype data revealed statistically significant markers demonstrating interaction effects with prenatal alcohol exposure on FASD outcomes.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7034W: Gene-environment interactions with maternal smoking and vitamin intake identified potential contributors for CL/P

Authors:

Z. Erdogan-Yildirim[1], J. C. Carlson[2,3], N. Mukhopadhyay[1], S. M. Weinberg[1], M. L. Marazita[1,2,4], J. R. Shaffer[2,1]; [1]Ctr. for Craniofacial and Dental Genetics, Sch. of Dental Med., Univ of Pittsburgh, Pittsburgh, PA, [2]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ of Pittsburgh, Pittsburgh, PA, [3]Dept. of Biostatistics, Sch. of Publ. Hlth., Univ of Pittsburgh, Pittsburgh, PA, [4]Clinical and Translational Sci., Sch. of Med., Univ of Pittsburgh, Pittsburgh, PA

Abstract:

Isolated cleft lip with/out cleft palate (CL/P) is a common and highly heritable birth defect with a multifactorial etiology. While maternal exposures such as smoking and vitamin intake as well as numerous CL/P genetic risk loci identified via genome-wide association analyses impact CL/P risk, the role of gene-environment interaction (GEI) in CL/P remains elusive.
To address this gap, we conducted a genome-wide GEI analysis on 540 cases and 260 controls recruited from the Philippines with two environmental exposures: maternal smoking and vitamin use. As GEI analyses are typically low in power and the results can be difficult to interpret, we used multiple testing frameworks to evaluate potential GEI effects. We measured GEI with a gene-by-environment interaction term (GxE) in logistic regression. Then, we conducted the 3 degree-of-freedom (3df) test which jointly tests the marginal genetic effect (G), the GxE effect, and the gene-environment association (GE) with GxEScanR. Lastly, we applied the two-step EDGE approach, to test variants that passed the joint G and/or GE association screening for GxE.
For vitamin use, we detected nine suggestive loci via 3df test and 22 suggestive loci using EDGE method. Two loci, *SV2B* at 15q26.1 (rs12441109: p_{3df} = 2.7e-6, p_{GxE} = 2.7e-6) and *MIER3* at 5q11.2 (rs179464: p_{3df} = 8.3e-6, p_{GxE} = 4.0e-4), showed GEI association using both approaches. Other noteworthy loci include *CECR2*, involved in chromatin remodeling, and *ALCAM*, a folate receptor candidate. The SNP rs12441109 at *SV2B* regulates the expression of *RCCD1*, component of histone demethylase complex, and locates 170kb distal from *FURIN* that encodes a serine protease known to cleave protein precursors including growth factors, their receptors, and extracellular matrix proteins. This broad region harbors transcriptionally active sites for craniofacial tissues. Studies have shown that folic acid inhibits *FURIN* activity and increases the expression of *CECR2*, indicating biologically plausible roles of these genes in CL/P.
For smoking, two loci showed modest GEI association via the 3df test: *TTBK1* at 6p21.1 (rs2756172, p_{3df} = 5.9e-7, p_{GxE} = 1.8e-3) and *PTPRD* at 9p23 (rs10816235, p_{3df} = 9.6e-6, p_{GxE} = 7.5e-4). The EDGE method identified five suggestive loci, two of which, *TWIST2* (p_{GxE} = 2.0e-4) and *NET1* (p_{GxE} = 2.3e-4), regulate epithelial-mesenchymal transition, which is required for lip and palate fusion. Moreover, *TWIST2* expression has been shown to increase upon exposure to cigarette smoke extract.
In summary, our study identified several suggestive GEI loci with maternal vitamin intake and smoking associated with risk of CL/P in Filipinos.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7035W: Genetic Control of Stage-Dependent Alternative Polyadenylation Events in the Developing Human Brain ⭐

Authors:

H. Yang[1], X. Ma[2], T. Zhang[2], C. Han[1], G. Minghui[1], Q. Liang[1], L. Li[2], C. Liu[3], C. Chen[1]; [1]Central South Univ., Changsha, China, [2]Shenzhen Bay Lab., Shenzhen, China, [3]SUNY Upstate Med. Univ., Syracuse, NY

Abstract:

Alternative polyadenylation (APA) events within 3' untranslated regions (3'UTRs) are pivotal modulators of posttranscriptional and translational processes, significantly impacting biological functions and disease risks, particularly in brain development and neuropsychiatric disorders. Over 90% of GWAS variants occur in noncoding regions, with 3'UTRs harboring critical noncoding variants. Quantitative trait loci (QTLs) associated with 3'UTR APA phenotypes (3'aQTLs) elucidate 16.1% of human diseases and traits. However, previous research on the developing human brain has been limited in splicing (sQTLs) and gene expression (eQTLs), left a gap in the dynamic temporal changes of aQTLs. This study explores the genetic control of stage-dependent APA (sdaQTL), with stage defined by fetal stages to old age. We conducted whole genome sequencing (WGS) and RNA sequencing (RNA-seq) on 315 human prefrontal cortex samples from East Asians (EAs). This cohort included 107 fetal brains and 208 postnatal adult brains spanning the fetal period (14 to 38 weeks post-conception, n=107), young and mid-aged (18 to 65 years, n=67), and elderly individuals (66 to 102 years, n=141). Utilizing DaPars2, we calculated the poly(A) site-usage index (PDUI) value. And maSigPro facilitated the identification of stage-dependent APA (sdAPA) events. We further compared developmental trajectories and QTL results between EA and European (EUR) populations. We collected and standardized summary statistics from 13 neuropsychiatric GWAS.
Bayesian co-localization identified GWAS signals sharing genetic effects with 3'aQTLs and eQTLs. We identified 26,359 sdAPA events, classified into four major clustering patterns: lengthening, shortening, lengthen-shortening and shorten-lengthening. Notably, 61.17% of sdAPA events exhibit a "lengthen-shortening" pattern during brain development and aging, exemplified by the gene APP, which underscores the complexity of post-transcriptional regulation.. QTLtools identified cis SNPs within 1 Mb of a 3'UTR associated with differential sdAPA usage, discovering 92,727 lead aQTLs, including the ZNF211, a zinc finger protein associated with schizophrenia. EA-specific aQTLs were identified, with 13% altering motifs crucial for APA , and imply that genetic diversity in 3'UTR might influence disease susceptibility. the co-localization of APA genes with neuropsychiatric disorder risk loci, particularly the identification of aQTLs-specific not

associated with eQTLs, suggests a unique role of APA in disease pathogenesis. Our findings provide novel insights into the genetic regulation of APA events in developing human brain.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7036W: Genomic Indicators of Female Infertility Using Transcriptome-Wide Association Studies in a European Population

Authors:

E. Kolbunova[1], H. Wheeler[2]; [1]Loyola Univ., Chicago, IL, [2]Loyola Univ Chicago, Chicago, IL

Abstract:

Globally, 1 in 6 people are affected by infertility. Cases of infertility can be associated with lifestyle, reproductive system trauma, or cancer treatment. While these factors contribute to the condition, our study focuses on the genetic indicators of infertility in females. We performed transcriptome-wide association studies (TWAS) for female infertility using Summary-PrediXcan (S-PrediXcan) and genome-wide association study (GWAS) summary statistics from the Pan-UK Biobank. The genetically predicted transcriptome may provide insight to the mechanisms and causes underlying female infertility. The GWAS cohort consisted of 1,431 females diagnosed with infertility of a genitourinary origin (phecode 626.8) and 194,287 controls of European ancestry. The top three SNPs associated with infertility were chr18:30326445 (p=1.6E-7), chr7:909277 (p=1.9E-7), and chr3:194191978 (p=2.0E-7, all build GRCh37). We used transcriptome prediction models trained with uterus, ovary, and whole blood data from the Genotype-Tissue Expression Project (GTEx) in our TWAS. Our most significant result from all three tissues was obtained in the whole blood model. Decreased expression of RP11-223C24.1 was associated with infertility (p=1.1E-5, β=-2.5). In the ovary model, increased expression of AC083843.2 was associated with infertility (p=2.4E-5, β=5.5). Highest median expression of AC083843.2 in GTEx occurs in testis tissue, at 4.23 transcripts per million. Our top hit in uterus tissue showed decreased expression of ISG15 in the uterus associated with infertility (p=4.5E-3, β=-2.7). ISG15 plays a vital role in the protein cascade that activates the innate immune response to viral infection and functions to suppress ovulation in mice. Our second top hit in the ovary tissue was CLEC4M (p=2.6E-5, β=0.2), which functions in cell adhesion and pathogen recognition. GIMAP1 is the third most significant gene from ovary tissue (p=4.8E-5, β=2.2) and second most significant in whole blood (p=4.8E-5, β=1.3). GIMAP1 is a GTP-binding protein critical for the development of mature B and T lymphocytes and may upregulate cytotoxic T cells. The number of cytotoxic T cells decreases during early pregnancy, thus, increased cytotoxic T cells caused by increased expression of GIMAP1 could potentially disrupt zygote

development, leading to infertility. These genes are potential biomarkers of female infertility. We plan to perform similar analyses in additional large biobanks like All of Us to confirm and identify additional genes associated with infertility phenotypes.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7037W: Identification of seminal plasma derived predictive biomarker microRNA in idiopathic non-obstructive azoospermia patients

Authors:

N. Sharma[1], A. Halder[2], S. Kaushal[1], M. Kumar[1], M. Jain[1]; [1]All India Inst. of Med. Sci. Delhi, Delhi, India, [2]All India Inst. of Med. Sci., New Delhi, India

Abstract:

Background: Male infertility patients with non-obstructive azoospermia (NOA) due to decrease in number of germ cells within the seminiferous tubule, need to undergo testicular biopsies for careful interpretation, as it may impact the prognosis for successful sperm retrieval and subsequent fertility treatment. The existing body of literature justify the importance of non-invasive molecular biomarker for diagnosis of NOA. This study aimed to determine the feasibility of translating testicular tissue derived miRNAs in to seminal plasma derived predictive biomarker in idiopathic NOA patients. **Materials and methods:** In this study 50 patients reporting with azoospermia due to hypospermatogenesis (HS) were recruited and exclusion of 10 patients was performed using XY fluorescent in-situ hybridization (XY-FISH), Yq microdeletion and whole exome sequencing (WES). Testicular tissue and seminal plasma derived total RNA was used for small RNA sequencing in 17 idiopathic HS patients and 4 normospermatogenesis controls as per the availability of RNA with RIN value >6. **Results:** Genomic analysis for exclusion of causative single nucleotide variants identified 4 HS patients with CFTR (c.1265C>T; p.Ser422Phe), CYP21A2 (c.550C>T; p.Gln184Ter), and AR (c.2179C>T; p.Arg727Cys) variants. Analysis of miRNAs in testicular tissue showed differential expression patterns between HS and normospermatogenic controls, with 92 downregulated and 8 upregulated miRNAs depicting miR-379-5p, miR-449a, miR-181c, miR-34b-3p, and miR-122b-3p being notable biomarkers. Analysis of spermatogenic miRNAs according to profile of each patients showed significant dysregulation in differential expression was linked to 69% of the cases with idiopathic HS. The seminal plasma miRNA profile showed 62 downregulated and 41 upregulated miRNAs. Venn diagram analysis of testicular and seminal plasma miRNAs showed downregulation in miR-132-5p, mir-219a-2-3p to be potential non-invasive candidate biomarker. Gene ontology and KEGG

pathway analysis identified PI3K-Akt and MAPK signalling and molecular functions like apoptosis and cell differentiation to be significantly enriched in HS patients. **Conclusion:** This is the first study of its kind to investigate miRNA differential expression profiling in a cohort of NOA patients exclusively of HS subtype. Comparing the study results to previously published data indicate dysregulated spermatogenic miRNAs are shared in HS, and other NOA subtypes. We identified two seminal plasma derived predictive biomarkers miR-132-5p and mir-219a-2-3p in idiopathic HS patients which needs to be validated further with biological experiments.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7038W: Identifying the causal role of human milk microbiota in childhood asthma using one- and two-sample Mendelian randomization

Authors:

Z. Fang[1], S. Stickley[1], A. Ambalavanan[1], K. Fehr[2,3,4], C. Petersen[5], K. Miliku[2,3,4,6], P. J. Mandhane[7], E. Simons[2,8], T. J. Moraes[9], S. E. Turvey[5], P. Subbarao[9,10,11], M. B. Azad[2,3,4], Q. Duan[1,12]; [1]Dept. of BioMed. and Molecular Sci., Queen's Univ., Kingston, ON, Canada, [2]Dept. of Pediatrics and Child Hlth., Univ. of Manitoba, Winnipeg, MB, Canada, [3]Dept. of Food and Human Nutritional Sci., Univ. of Manitoba, Winnipeg, MB, Canada, [4]Manitoba Interdisciplinary Lactation Ctr. (MILC), Children's Hosp. Res. Inst. of Manitoba, Winnipeg, MB, Canada, [5]Dept. of Pediatrics, British Columbia Children's Hosp., Univ. of British Columbia, Vancouver, BC, Canada, [6]Dept. of Nutritional Sci., Univ. of Toronto, Toronto, ON, Canada, [7]Dept. of Pediatrics, Univ. of Alberta, Edmonton, AB, Canada, [8]Section of Allergy and Immunology, Univ. of Manitoba, Winnipeg, MB, Canada, [9]Dept. of Pediatrics, The Hosp. for Sick Children, Toronto, ON, Canada, [10]Dept. of Med., McMaster Univ., Hamilton, ON, Canada, [11]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada, [12]Sch. of Computing, Queen's Univ., Kingston, ON, Canada

Abstract:

Human milk is recognized as an optimal source of nutrition that provides many long-term benefits to children's health, yet its causal impact on asthma prevention remains poorly understood. Earlier work by our lab identified microorganisms in human milk that are associated with maternal genomics and childhood asthma at age 5 years. In this study, we conducted Mendelian randomization (MR) to investigate putative causal relationships between exposure to human milk microbiota (HMM) and the development of childhood asthma among human milk-fed infants.
HMM composition was determined by 16S rRNA sequencing of milk samples collected between 3-4 months postpartum from a subset of 885 mothers of the Canadian CHILD Cohort Study.

Single nucleotide polymorphisms (SNPs) were genotyped in these mothers and their infants using the Illumina HumanCoreExome BeadChip. Using SNPs identified in our previous genome-wide association studies (GWAS) of HMM as instruments, we performed one-sample intergenerational MR analyses using the two-stage least squares (2SLS) method to identify maternal SNPs that are associated with childhood asthma. To account for genetics shared between mothers and their human milk-fed infants, we adjusted for infant genotypes. Furthermore, we performed two-sample MR using summary statistics from a GWAS of childhood asthma (n=314,633) from the UK Biobank population (Ferreira *et al.*, *AJHG*, 2019). In addition to applying the inverse-variance weighted MR method in the main analysis of two-sample MR, we employed sensitivity analyses, including MR-PRESSO, MR-Egger, and weighted median methods, to determine potential pleiotropic effects of the selected SNPs.
Using both one-sample intergenerational and two-sample MR, we identified genetically determined taxa in HMM that may have causal effects on promoting or preventing children's asthma at age 5 years. For example, our MR results determined that exposure to *Veillonella* (P = 0.02) and *Actinomyces* (P = 0.01) in human milk were protective against childhood asthma, while *Rothia* (P = 0.01) might increase the risk of childhood asthma. To the best of our knowledge, this is the first MR study investigating the potential causal effects of maternal HMM on their children's risk of asthma. A better understanding of the causal impact of HMM on childhood asthma will be instrumental in guiding early-life prevention strategies, including the potential development of milk probiotics.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7039W: Lifespan analysis of repeat expression in the neurotypical human brain reveals age-dependent upregulation of HERV-K

Authors:

A. SantAnna Feltrin[1], T. Evans[2,1], K. J. Benjamin[1,3,4], T. Katipalli[1], T. Hyde[1,5,3], J. Kleinman[1,4], D. Weinberger[1,5,3,4,6], A. Paquola[1,5,3], J. Erwin[1,5,3]; [1]Lieber Inst. for Brain Dev., Baltimore, MD, [2]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [3]Johns Hopkins Univ. Sch. of Med., Dept. of Neurology, Baltimore, MD, [4]Johns Hopkins Univ. Sch. of Med., Dept. of Psychiatry & Behavioral Sci., Baltimore, MD, [5]Johns Hopkins Univ. Sch. of Med., Dept. of NeuroSci., Baltimore, MD, [6]Johns Hopkins Univ. Sch. of Med., McKusick-Nathans Dept. of Genetic Med., Baltimore, MD

Abstract:

DNA repetitive sequences (or repeats) comprise over 50% of the human genome and have a crucial regulatory role in the organism, being one of the key elements in regulating the transcription machinery. The human brain is the tissue with the highest detectable repeat expression. Dysregulation of repeat activity is related to several neurologic and neurodegenerative disorders, as repeat-derived products are capable of stimulating a pro-inflammatory response. The dysregulation of the epigenetic machinery is a hallmark of aging; as such, it is still unclear how repeat expression acts on the neurotypical human brain. Our study leverages a large postmortem transcriptome cohort that spans the human lifespan and examines differentially expressed repeats (DERs) in the neurotypical human brain across various age groups. We also applied Weighted Correlation Network Analysis (WGCNA) to evaluate the repeats' connectivity in the co-expression network of each brain region. Using bulk RNA-sequencing data from post-mortem brain samples, we identified repeat expression patterns across the caudate nucleus (n=271), dorsolateral prefrontal cortex (n=304), and hippocampus (n=310) from male and female neurotypical donors. By analyzing samples across seven age bins (Prenatal; 0-15; 16-29; 30-39; 40-49; 50-59; 60+ years old), we identified 21,696 DERs strand-specific across all three brain regions. Looking into 0-15 years v. 60+ years comparisons, we found that LINEs and LTRs DERs were the most abundant repeat families; from these LTRs, we identified 17 LTRs shared across all brain regions (with HERV-K-int repeat showing an increased expression on older brains). Co-expression analysis using WGCNA also shows that repeats from the HERV subfamily are intramodular hubs in its subnetworks in all brain regions. While we do not observe a strong global relationship between repeat expression and age, we identified HERV-K as a repeat signature associated with the aging neurotypical brain. Our study is the first global assessment of repeat expression in the neurotypical brain.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7040W: Long-read whole genome sequencing for comprehensive genomic profiling of men with infertility ⭐

Authors:

T. Garcia; Baylor Coll. of Med., Houston, TX

Abstract:

Our research leverages long-read sequencing technologies to identify genomic variants linked to male infertility, focusing on complex regions, genotypes, and variants elusive to short-read sequencing. Through the Oxford Nanopore PromethION long-read sequencing platform, we sequenced the whole genomes of over 150 men presenting with various forms of male-factor

infertility—from azoospermia to asthenozoospermia to combinations like oligoasthenoteratozoospermia—at coverages surpassing 40X per male. Our analysis encompassed copy number variations, short tandem repeats, and Y chromosome microdeletions/microduplications, integrating the T2T-CHM13v2.0 assembly for enhanced Y-chromosome mapping. Variant interpretation adhered to ACMG guidelines, using new AI tools such as AlphaMissense for pathogenicity prediction. Our enhanced methodology facilitated the discovery of clinically significant variants that may be associated with the infertility phenotypes observed. Some especially significant findings include a 27-year-old patient exhibiting an absence of LH and puberty, markedly low testosterone levels, and azoospermia. This patient was found to have a novel homozygous p.A37Gfs*45 variant in the *LHB* gene, predicted to lead to protein loss, aligning with his clinical symptoms; a patient with severe asthenozoospermia harboring rare compound heterozygous pathogenic variants (p.K1836R and p.E3097D) in the *DNAH2* gene, essential for sperm motility; a patient with cryptozoospermia possessing rare compound heterozygous pathogenic variants (p.C317G, p.I2078T, and p.P5S) in the *TDRD6* gene, crucial for spermiogenesis; a patient with cryptozoospermia identified with a rare homozygous pathogenic variant (p.L60F) in the *TEX14* gene, required for meiosis; and a patient with normal semen parameters but suspected male-factor infertility, found to have rare compound heterozygous pathogenic variants (p.S269Lfs*33 and p.S271T) in the *PRSS55* gene. This gene, newly identified in mice, is necessary for the proper migration of sperm to the oviduct and for effective sperm-egg binding. In conclusion, our findings underscore the utility of phased long-read sequencing in elucidating genetic causes of sperm dysfunction. This approach enhances genetic diagnosis and treatment of male infertility, offers new insights into its genetic complexities, and advances the field by identifying novel variants. Such discoveries support personalized medicine in reproductive health, aiming to improve patient outcomes.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7041W: Multi-ancestral maternal GWAS meta-analysis of longitudinal trajectory of fetal growth ⭐

Authors:

P. Wijesiriwardhana[1], R. J. Biedrzycki[2], T. D. Habtewold[1], F. Tekola-Ayele[1]; [1]Epidemiology Branch, Div. of Population Hlth.Res., Div. of Intramural Res., Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, Maryland, USA, Bethesda, MD, [2]Glotech, Inc., contractor for Div. of Population Hlth.Res., Div. of Intramural Res., Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, Maryland, USA, Bethesda, MD

Abstract:

Fetal growth sets the foundation for health across lifespan, and abnormalities are associated with increased risk for mortality and cardiometabolic diseases. Previous genome wide association studies (GWAS) in European ancestry populations have identified maternal and fetal genetic variants associated with birthweight. However, the genetic architecture of longitudinal fetal growth may not overlap with that of birthweight, because birthweight does not represent the trajectory of fetal growth throughout pregnancy. We aimed to identify maternal genetic loci associated with longitudinal trajectory of fetal weight throughout pregnancy by combining genome wide datasets of pregnancy cohorts in the United States across seven groups with African, East Asian, Hispanic, and European ancestry (total N=7,918). Fetal weight was calculated from fetal biometry measured by ultrasound at 2-5 gestational time points, and birthweight was measured at delivery. For each of the seven groups, GWAS of the change in fetal weight trajectory was performed using a linear mixed effects-based model adjusted for fetal sex and genotype principal components, with varying slope for gestational week. The results were combined using trans-ancestral meta-regression with genomic control correction on inputs and adjusting for three genetic principal components using the MR-MEGA tool. We identified 57 SNP loci associated with fetal weight trajectory ($p < 5E-8$). For most of the identified loci, allelic effect heterogeneity was correlated with ancestry. Multi-ancestry meta-regression improved fine-resolution mapping for some loci. Genes annotating the associated SNPs are enriched for disease and biological functions such as lipid and carbohydrate metabolism, immune response, neuropsychology, organ development, and perinatal death. A genetic risk score derived from the 57 maternal loci was associated with longitudinal fetal weight trajectory, but one derived from previously known maternal loci for birthweight was not. A genetic risk score derived from a combination of the fetal weight trajectory loci and birthweight loci showed improved association with fetal weight trajectory only when the trajectory included fetal weight and birthweight. In conclusion, our study showed that the maternal genetic architecture of in-utero fetal growth trajectory is largely distinct from that of birthweight. Multi-ancestral cohorts with longitudinal fetal biometry measures advance insights into biological underpinnings of fetal growth and its aberration.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7042W: Placental gene expression signatures based on maternal parity

Authors:

L. Haizler-Cohen[1], G. Wang[2], K. L. Grantz[3], F. Tekola-Ayele[3]; [1]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]The Prospective Group (TPG), contractor to Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, MD, [3]Epidemiology Branch, Div. of Population Hlth.Res., Div. of Intramural Res., Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, MD

Abstract:

Introduction: Nulliparity is associated with adverse pregnancy outcomes including preeclampsia, preterm birth, low birth weight and stillbirth. Differentially expressed genes (DEGs) and gene co-expression networks between nulliparous and multiparous placentas may drive altered placental function, that potentially underly differences in pregnancy complications. Our objective was to identify human placental gene expression differences by maternal parity.
Methods: RNA sequencing was performed on 82 placental samples obtained at delivery as part of the NICHD Fetal Growth Studies. Genes that are expressed differently between nulliparous (n=34) and multiparous (n=41) placentas were determined using the DESeq2 package, adjusting for maternal age, mode of onset of labor, and self-reported race/ethnicity. Analyses were performed in the full cohort and stratified by fetal sex (36 females and 39 males). A total of 16,725 genes that passed quality control criteria were included in analyses. Placental weighted gene co-expression network analysis (WGCNA) was performed to identify co-expression modules correlated with parity. All genes belonging to a significant module were further explored for enrichment in canonical pathways, networks, diseases, and biological function.
Results: In our cohort *DDX5* had higher placental expression in multiparas compared to nulliparas (FDR<0.05) and appeared to be male specific (interaction p=0.03). In females, two genes (*ANKRD33* and *SLITRK6*) had higher expression in multiparas, while two genes (*IL1B* and *MTCO1P40*) had higher expression in nulliparas. WGCNA detected 18 gene co-expression modules in the placenta, of which two modules (namely Grey60 and Tan) were correlated with parity. For the Grey60 module (124 genes) gene co-expression was significantly lower in multiparas for the full cohort (r=-0.26, p=0.023) and female sub-cohort (r=-0.39, p=0.018). For the Tan module (32 genes) gene co-expression was also significantly lower in multiparas in the full cohort (r=-0.25, p=0.029) and male sub-cohort (r=-0.38, p=0.017). Genes in the Grey60 module were enriched in canonical pathways related to immune response and inflammation, hematological and cardiovascular system development, and immune cell trafficking. Genes in the Tan module were enriched for canonical pathways related to endoplasmic reticulum stress response, and reproductive, embryonic and cardiovascular system development.
Conclusion: Differences in human placental gene expression between nulliparous and multiparous pregnancies are related to varied biological processes that may underlie maternal and neonatal outcomes that differ by parity.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7043W: Single nucleus molecular mechanisms associated with Caudal Regression Syndrome in mice

Authors:

C. Zajac[1], R. D. Albanus[2], H. T. Vu[1], N. Manickam[1], E. Curka[1], C. Keegan[3], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Washington Univ., St Louis, MO, [3]Univ Michigan, Ann Arbor, MI

Abstract:

Background Birth defects occur in approximately 3% of all births in the United States including embryonic caudal malformations such as Caudal Regression Syndrome (CRS), which is a condition where the gastrointestinal and genitourinary systems and the lower spinal cord are abnormally developed. The origins and genetic mechanism of this syndrome remain unclear in humans, specifically the affected cell types and regulators. Danforth's short-tail mutant (Sdsd/sd) mouse exhibits a semi-dominant phenotype characterized by cessation of the vertebral column, affecting the proper development of urogenital and gastrointestinal organs, which overlap with human caudal dysgenesis phenotypes. This mutation is caused by an endogenous retrotransposon (ETn) 12 kb upstream of the Ptf1a gene in the promoter region of the lncRNA gene Gm13344 during embryogenesis day E9.5, resulting in overexpression of both genes in the developing tailbud. This region is orthologous to a human pancreatic developmental enhancer making it an attractive model system for deep molecular characterization of the Sd phenotype, and a valuable translational resource to understand the etiology of CRS in humans. **Methods** We conducted transcriptomic profiling of Sd tissues by performing 3' and 5' end single-nucleus RNA sequencing (snRNA-seq) using mouse E9.5 tail buds from wild type (WT) and Sd mice to identify cell-specific transcriptomic changes in the Sd mouse genome. **Results** We obtained 34,416 high-quality nuclei in total. Clustering on 3' RNA data resulted in 15 clusters, and 5' data in 19 clusters, which correspond to different cell types such epithelium and endothelial cells. Comparison of gene expression and regulatory activity between the two conditions showed gene profiles specific to Sd such as Ptf1a and Gm13344 which are associated with transcriptional changes in CRS**Future directions** Preliminary results identified cell types that display early molecular events that may be causal to the caudal phenotype. Integration of 3'-5'snRNA seq will assist to identify Sd-specific transcriptional factor changes and improve our understanding of the causal regulatory networks linked to the Sd phenotype and, consequently, caudal developmental disorders in humans.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7044W: Spatial transcriptomic analysis of the molecular drivers behind bronchopulmonary dysplasia

Authors:

S. Mallapragada[1,2], A. L. Williams[1], S. Shirazi[3], J. M. Sucre[3], N. Banovich[1]; [1]Translational Genomics Res. Inst., Phoenix, AZ, [2]Arizona State Univ., Phoenix, AZ, [3]Div. of Neonatology, Dept. of Pediatrics, Vanderbilt Univ. Med. Ctr., Nashville, TN, USA, Nashville, TN

Abstract:

Bronchopulmonary dysplasia (BPD) results from arrested lung development after preterm birth and is a leading complication associated with preterm birth, affecting nearly 80% of infants born less than 26 weeks gestation. While the environmental associations with BPD are well-described, the molecular drivers of the permanent structural deficits of impaired alveologenesis and angiogenesis are not well-defined. Prior work suggests dysregulated signaling pathways between the mesenchyme, alveolar epithelium, and the endothelium as a key driver of impaired lung development. While there exist many animal models of saccular stage lung injury, studying the evolution of neonatal lung injury and BPD in humans has been difficult. Using the 10X Genomics Xenium platform, we analyzed transcriptomic information in ~1.7 million cells generated from 17 pediatric lung samples across human lung development, neonatal lung injury, evolving, and established BPD that were processed with a panel of 343 genes to study the expression patterns and understand the molecular drivers behind normal lung development, acute neonatal lung injury, and BPD. First, we assessed changes in cell type composition between BPD and control lungs. We observed significant changes in abundance in over half of the cell types (23 of 36; FDR <= 0.1). Cell type composition changes were observed in 7 of the 11 epithelial cell types, 4 of the 5 endothelial cell types, 4 of the 7 mesenchymal cell types, and 7 of the 12 immune cell types. To assess the molecular changes associated with BPD, we performed a differential gene expression analysis. Here, we found that 320 of the 343 genes profiled were significantly different in at least 1 cell type (FDR <= 0.1). Finally, to leverage the full power of spatial transcriptomics, we performed graph-based niche analyses to identify nuanced dysregulation that is associated with alveolar damage during BPD. Understanding evolving human neonatal lung injury at single-cell resolution alongside its spatial context will help us untangle the cellular and molecular drivers of the pathobiology of this disease, with a goal of providing foundational data to develop druggable targets early in the disease process that would promote lung development after preterm birth and lung injury.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7045W: The Australian Genomic Autopsy Study: Summary of outcomes from 414 families, and new diagnoses from systematic reanalysis and reclassification.

Authors:

T. Ha[1,2,3], M. Frank[4], L. Dejong[1], P. Arts[2], A. Byrne[5], M. Babic[1], A. Gardner[3], J. Dearman[6], H. Kovilpillai[6], G. Maxwell[6], S. Montazaribarforoushi[3], J. Feng[7], S. Gao[7], E. Collen[7], S. Marri[7], J. Sanchez Ferrero[7], J. Toubia[8,2], S. Webb[2], D. Lawrence[9,2], J. Andrews[2,9], L. Arriola Martinez[2], L. Eshraghi[10], Genomic Autopsy Study Research Network, Y. Khong[3,11], L. Moore[3,11], A. Bouman[12], M. Jackson[1,9], K. Kassahn[7,3], A. Kulkarni[1,13], H. Scott[1,8,14,3,9], C. Barnett[3,6,9]; [1]SA Pathology, Genetics and Molecular Pathology, Adelaide, Australia, [2]Ctr. of Cancer Biology, alliance between SA Pathology and UniSA, Adelaide, Australia, [3]Univ. of Adelaide, Adelaide Med. Sch., Adelaide, Australia, [4]SAHMRI, Adelaide, Australia, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Women's and Children's Hosp., Paediatric and Reproductive Genetics Unit, Adelaide, Australia, [7]SA Pathology, Technology Advancement Unit, Adelaide, Australia, [8]SA Pathology, ACRF Cancer Genomics Facility, Adelaide, Australia, [9]Australian Genomics, Melbourne, Australia, [10]Garvan Inst. of Med. Res., Adelaide, Australia, [11]Women's and Children's Hosp., Anatomical Pathology, Adelaide, Australia, [12]Erasmus Univ. Med. Ctr., Rotterdam, Netherlands, [13]Monash Hlth.Pathology, Melbourne, Australia, [14]Univ. of South Australia, Clinical and Hlth.Sci., Adelaide, Australia

Abstract:

Background: In Australia, 6 babies are stillborn and 2 die within 28 days of life. Using genomic autopsy as an adjunct for standard autopsy, a definitive diagnosis (LP/P) was identified in 21% of families with pregnancy loss and perinatal death (42/200), while another 31.5% (63/200) had a selected candidate, variant or gene of uncertain significance (VUS or GUS), for follow-up investigations (Byrne et al. 2023). An additional 214 families were analysed, and in total, 61% of families (243/414) remain unresolved following clinical exome (ES) and genome (GS) analysis.
Aim: Develop a systematic framework to increase the number of candidate and definitive diagnosis in families with negative ES and GS clinical findings.
Method: Genotype and phenotype data from 350 families were shared with collaborators to undergo periodic renalysis. 240 GS data was reanalysed in-house using AI-based variant prioritization. 95 families with VUS/GUS were submitted to matchmaking exchange platforms (matchbox/genematcher). 20 families with putative splicing candidates were investigated using RNA studies. 25 families with genes associated with prior episignatures established had methylation arrays performed. 2 additional mouse models were established for novel skeletal

disease genes.
Results: 3/100 families reanalysed from GS data had a new diagnosis from identifying a single nucleotide variant *in trans* with a copy number variant. 15/100 families had a selected candidate (VUS/GUS) for follow-up investigations. 6/95 families have had genes with additional kindreds identified with similar or matching phenotypes from unpublished cohorts. In 4/20 families, RNA studies have validated the suspected aberrant splicing. One of the two families with a novel skeletal disease gene, has a mouse model that recapitulates the fetal phenotype under investigation.
Conclusion: Thus far, the combination of both genematching and adjunct experimental follow-ups have enabled the reclassification of at least 10 families, and ongoing functional studies for another 15 families. Using our reanalysis and adjunct functional framework, we have identified another 25% of additional diagnoses in previously negative ES and GS cases.
***NB:** Scott and Barnett are senior coauthors.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7046W: Unraveling mechanisms of protamine-based packaging in the sperm genome

Authors:

S. Srinivasan, M. Rabbani, R. Agrawal, L. Moritz, S. S. Hammoud; Univ. of Michigan, Ann Arbor, MI

Abstract:

Sexual reproduction hinges on the fusion of sperm and egg to create a zygote. The process of sperm cell generation is known as spermatogenesis, which encompasses three critical steps: spermatogonial stem cell proliferation (mitosis), a reduction division to generate haploid cells (meiosis), and the maturation of haploid round spermatids into elongating spermatids through nuclear and cytoplasmic remodeling (spermiogenesis). A key feature of spermiogenesis is the extensive chromatin remodeling process, during which the round spermatid genome replaces most of its canonical histones with sperm-specific nuclear proteins known as protamines, primarily protamine 1 (PRM1) and protamine 2 (PRM2). This protamine-based packaging system is highly conserved across species, spanning 500 million years of evolution. However, we know little about the chromatin remodeling events and chromatin associated factors required for establishing a sperm-specific landscape. To achieve this protamine-based packaging, the current dogma posits that haploid round spermatids are initially destabilized by acquiring post-translational modifications (e.g., acetylation), which weaken the histone-DNA interactions and

facilitate the sequential replacement of canonical and testis-specific histones with transition proteins, and ultimately, protamines. However preliminary data from our laboratory, as well as others, has shown that PRM1 and PRM2 deposition along the genome is temporally uncoupled, suggesting these proteins bind distinct genomic regions. To explore this further, we have been employing a combination of imaging-based and sequencing-based technologies to temporally map the locations within the accessible spermatid genome. Preliminary chromatin fiber analysis supports the notion that these proteins indeed bind to specialized genomic regions. Resolving the grammar and logic of a dual protamine-based packaging system will address a longstanding mystery: whether these proteins carry instructive information or serve a purely structural role.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7047W: Validating the addition of 39 ACMG-recommended genes to a 235-gene carrier screening panel

Authors:

S. Zhang, L. Eisemann, A. Pirages, B. Obadia, S. Cox, M. Judkins, K. Moyer, E. J. Zmuda, J. J. Alexander; Myriad Genetics, Inc., Salt Lake City, UT

Abstract:

Background: Carrier screening enables prospective parents to assess the risk of transmitting an autosomal recessive or X-linked condition by identifying pathogenic variants across hundreds of genes. The American College of Medical Genetics and Genomics (ACMG) published a practice resource in 2021 that aimed to better define the carrier screening process, while advocating for a more consistent and equitable approach for offering carrier screening to all patients. ACMG proposed a tiered definition of carrier screening to allow patients and health care professionals to communicate with greater precision, specifically recommending tier 3 screening (113 genes) be offered to all patients. Here, we describe the analytical validation of the addition of 39 genes recommended by the ACMG practice resource to a previously validated 235-gene sequencing-based carrier screening panel.
Methods: Following guideline recommendations to assess analytical performance of a carrier screening panel before clinical use, single nucleotide variant (SNV) and small insertion or deletion (INDEL) calling performance was assessed for 39 new genes via concordance with external reference results. 38 Coriell samples were used to evaluate SNV/INDEL concordance. For calculating True Positive (TP) and True Negative (TN), all polymorphic sites (where non-reference bases were observed in any sample) across all samples were considered. Copy number variant (CNV) calling performance was assessed in 37 of the 39 new genes for 861

clinical samples using sensitivity based on simulated CNVs.
Results: Among the 38 Coriell samples utilized to demonstrate SNV and INDEL performance, analytical sensitivity was >99% and analytical specificity was >99% across reportable regions of interest, with 24,025 TP, 93,221 TN, 75 false positives (FP), and 8 false negatives (FN). A subset of FP and FN results were manually examined. The primary cause of discordance were variants located near repetitive or technically challenging regions. Among the 861 clinical samples utilized for CNV calling performance, sensitivity based on panel-wide simulated CNVs was >95.8%. In addition, 100% of known CNVs (n=11) based on orthogonal multiplex ligation dependent probe amplification (MLPA) or junction read evidence were detected in the updated panel.
Conclusions: This validation study demonstrated the high accuracy of detecting SNVs and INDELs in 39 genes, as well as CNVs in 37 genes recommended by the ACMG practice resource. These genes are now validated for clinical testing of patient samples.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7048W: Validation of fetal and maternal recessive disease genotyping with FirstGene: a combined, non-invasive prenatal cfDNA assay for fetal aneuploidy, recessive diseases, and serological screening

Authors:

J. Wang, K. Trettin, C. Battey, R. Patel, A. Srinivasan, J. Saikumar, D. Kushnoor, S. Pierson, K. Johansen Taber, G. Gould, D. Muzzey; Myriad Genetics, Inc., Salt Lake City, UT

Abstract:

Objectives: The success of identifying couples at risk of having fetuses affected by recessive conditions depends on both partners receiving carrier screening. However, carrier screening is often performed only on the pregnant patient, with fewer than 50% of male partners completing screening. The FirstGene assay combines multiple prenatal risk assessments into a single blood draw, predicting fetal and maternal recessive disease status simultaneously, without the need for paternal screening. Here we describe the analytical validation of FirstGene for detecting fetal and maternal single nucleotide variants (SNVs) and small insertions or deletions (INDELs) in genes associated with serious recessive conditions.
Methods: 79 Plasma samples from 59 pregnant patients were run on FirstGene. Concordance analysis was performed by running corresponding fetal DNA—gathered via CVS or amniocentesis—on the previously validated Foresight carrier screen to evaluate SNVs and INDELs across 17 genes. In addition, 93 genomic cell line DNA (gDNA) mixtures, created by

mixing mother-child pairs with rare genotypes of interest at different ratios (3, 5, 10, 15 and 20%) to represent a range of fetal fractions, were run on FirstGene and assessed for concordance with Foresight using the corresponding single source gDNA samples. Samples below 3% fetal fraction are routinely failed in FirstGene.
Results: In the 79 plasma samples, FirstGene accurately detected 5090 SNVs and 122 INDELs in fetuses, resulting in 99.36% sensitivity and 99.81% specificity for SNVs, and 93.13% sensitivity and 99.31% specificity for INDELs. FirstGene-derived maternal carrier screening results had 99.98% sensitivity and 99.998% specificity for SNVs, and 97.71% sensitivity and 99.96% specificity for INDELs. The majority of the maternal miscalls were in repetitive regions. Among the 93 gDNA mixtures, FirstGene achieved 98.36% sensitivity and 99.07% specificity for fetal SNVs (n=5649), and 93.26% sensitivity and 97.56% specificity for fetal INDELs (n=166). 13/15 (86.7%) cystic fibrosis (CF)-affected fetuses were identified correctly; the 2 mis-predicted as carriers had a 3% mixing ratio. Notably, among the correctly identified CF-affected cases, 4 were caused by a homozygous CFTR ΔF508 variant and 9 were caused by a heterozygous state for 2 pathogenic variants.
Conclusions: FirstGene accurately identified fetal SNV and INDELs for recessive conditions. This screening approach streamlines the identification of at-risk pregnancies, especially in cases where carrier status is unknown for the reproductive partner.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Wednesday Session

Board 7049W: Whole-genome sequencing of nonobstructive azoospermia patients identify rare predisposing regulatory variants ⭐

Authors:

Y. Long[1,2], C. Chen[1], L. Yang[1], H. Shen[1], Z. Hu[1], C. Wang[1]; [1]Nanjing Med. Univ., Nanjing, China, [2]Southeast Univ., Nanjing, China

Abstract:

Nonobstructive azoospermia (NOA) represents a rare and severe form of male infertility and remains currently incurable. A number of rare loss of function variants (LoFs) have been implicated in NOA, yet the association between rare non-coding variants and NOA remained largely unknown. In this study, we performed whole-genome sequencing (WGS) on 1338 individuals diagnosed with NOA and 2664 fertile men and standard genotyping protocol were applied. To further decipher the functional significance of regulatory variants, we utilized single-nuclei RNA sequencing and ATAC-seq of human testicular tissues to construct a gene regulatory network. Additionally, we developed a deep learning model to predict the functional impact of

regulatory variants and employed massively parallel reporter assays (MPRAs) to systematically evaluate the function. We found that the burden of LoFs of 355 genes with infertile phenotypes in knock-out mouse model was significantly higher in cases than in controls. In addition, we identified 6 new predisposing genes with elevated LoFs burden. These genes mentioned above were then categorized based on their expression pattern across distinct stages of spermatogenesis, and their regulatory elements (i.e., promoters and enhancers) were subsequently defined. Functional regulatory variants were prioritized using our prediction model and MPRA validation. Notably, we identified 26 genes showed an elevated burden of functional regulatory variants in NOA cases, including crucial genes involved in spermatogenesis such as *RAD51* and *DMRT1*. In total, 3.7% of patients harbored these predisposing regulatory variants. In conclusion, our study confirms the contribution of previously studied genes to male infertility and sheds light on new molecular mechanisms by providing a list of regulatory variants associated with NOA.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4002W: A biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic epistasis

Authors:

B. Fu, A. Pazokitoroudi, A. Xue, A. Anand, P. Anand, R. Border, N. Zaitlen, S. Sankararaman; UCLA, Los Angeles, CA

Abstract:

The contribution of epistasis to human complex trait variation is poorly understood, in part, because attempts to explicitly identify SNPxSNP interactions associated with a trait have low power. An alternate approach involves testing whether a SNP modulates trait variation against a polygenic background. We present a method to test for marginal epistasis (ME) of a SNP on a trait that is applicable to biobank-scale data quantitative traits. We performed extensive simulations to show that our method, FAME, provides calibrated tests of ME while being highly sensitive to true signals of epistasis. Importantly, FAME is the only method that can be applied to biobank-scale datasets. We applied FAME to test for ME at SNPs associated with 53 quantitative traits across ≈ 300 K unrelated white British individuals in the UK Biobank (UKB). Testing 15,601 GWAS significant associations, we identified 16 trait-SNP pairs across 12 traits and 14 SNPs that demonstrate significant evidence of ME (p &lt (5e-8)/53). Spurious signals of epistasis can occur due to imperfect tagging of causal variants and model violation (large

marginal effects, heavy-tailed noise distribution, unmodeled GxE, nonlinear covariate effects). We confirmed that FAME remains calibrated in each of these settings. We validated our findings by showing that significant ME SNPs replicate in imputed genotypes and across independent splits of UKB and by performing association tests of these SNPs against genome-wide SNPs. We use FAME to partition the significant ME signals across the genome identifying 5 trait-SNP pairs evidencing within-chromosome ME while 12 evidencing cross-chromosome ME. 10/14 unique significant ME SNPs are cis-eQTLs in blood e.g. rs964184, a significant ME SNP for ApoB, Cholesterol, and triglycerides, is a cis-eQTL,trans-eQTL, and trans-pQTL. Across the 16 trait-SNP pairs, we document that the proportion of trait variance explained by ME is about 12x as large as that explained by the GWAS effects on average (range: 0.59-43.94) implying that epistasis has a substantial impact on the impact of the variant on the trait. Our results show, for the first time, evidence of interaction effects between individual genetic variants and overall polygenic background modulating complex trait variation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4003W: A Multi-phenotype colocalization framework in*LocusFocus*

Authors:

H. Lu; Univ. of Toronto, Toronto, ON, Canada

Abstract:

Genome-wide association studies (GWAS) have identified genetic loci associated with complex diseases. A next step is to elucidate their mechanism of action. Colocalization analysis, which integrates GWAS data with other GWAS/omics data, can suggest biological mechanisms and evaluate pleiotropic effects. We present a multi-trait extension of our Simple Sum 2 (SS2) frequentist-based colocalization methodology, implemented in our *LocusFocus* software. This extension includes an omnibus test to determine if at least one trait colocalizes with the GWAS summary statistics and a separate test assessing trait-set colocalization. In simulations, our omnibus test demonstrated superior power compared to the minimum p-value approach, matching the performance of the Cauchy combination test (CCT) while maintaining a type one error rate at 5%. This advantage holds as the number of traits increases. For instance, when six out of 10 secondary traits colocalize with the primary trait, both the omnibus test and CCT achieve 80% power, surpassing the 70% power of the minimum p-value approach. From a computational standpoint, our new method outperforms both the minimum p-value and the CCT. The method can test 10 traits within a 500 SNP region in under three seconds, which is 10

times faster than methods. This efficiency increases linearly with the number of traits and cubically with the length of the genomic position. We plan to refine our method to address summary statistics coming from overlapping samples and apply it to our Multisystem Inflammatory Syndrome in Children (MIS-C) cohort to explore colocalization in related phenotypes and omic measurements.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4004W: A novel, scalable, and efficient blockLASSO PGS method in All of Us and the UK Biobank

Authors:

T. Raben[1], L. Lello[2,1], E. Widen[2,1], S. Hsu[1,2]; [1]Michigan State Univ., East Lansing, MI, [2]Genomic Prediction, Inc, North Brunswick Township, NJ

Abstract:

We develop a "block" LASSO (blockLASSO) method for training polygenic scores (PGS) and demonstrate its use in All of Us (AoU) and the UK Biobank (UKB). BlockLASSO utilizes the approximate block diagonal structure (due to chromosomal partition of the genome) of linkage disequilibrium (LD). LASSO optimization is performed chromosome by chromosome basis, which reduces computational complexity by orders of magnitude. The resulting predictors for each chromosome are combined using simple re-weighting techniques.
We demonstrate that blockLASSO is as effective for training PGS as (global) LASSO and other approaches. This is shown for 11 different phenotypes, in two different biobanks, and across 5 different ancestry groupings (African, American, East Asian, European, and South Asian).
The block approach works for a wide variety of phenotypes. In the past, it has been shown that some phenotypes are more/less polygenic than others. Using sparse algorithms, an accurate PGS can be trained for type 1 diabetes (T1D) using ~100 SNVs. On the other extreme, a PGS for body mass index (BMI) would need more than 10k SNVs. blockLasso produces similar PGS for phenotypes while training with just a fraction of the variants per block. For example, within AoU (using only genetic information) block PGS for T1D (1,500 cases/113,297 controls) reaches an AUC of 0.63(0.02) and for BMI (102,949 samples) a correlation of 0.21(0.01). This is compared to a traditional global LASSO approach which finds for T1D an AUC 0.65(0.03) and BMI a correlation 0.19(0.03). Similar results are shown for a total of 11 phenotypes in both AoU and the UKB and applied to all 5 ancestry groupings as defined via an Admixture analysis. In all cases the contribution from common covariates--age, sex assigned at birth, and principal components--are removed before training.

This new block approach is more computationally efficient and scalable than global machine learning approaches. Genetic matrices are typically stored as memory mapped instances, but loading a million SNPs for a million participants can take ~8TB of memory. Running a LASSO algorithm requires holding in memory at least two matrices this size. This requirement is so large that even large high performance computing clusters cannot perform these calculations. To circumvent this issue, most current analyses use subsets: e.g., taking a representative sample of participants and filtering SNPs via pruning and thresholding. High-end LASSO training uses ~500GB of memory (e.g., ~400k samples and ~50k SNPs) and takes 12-24 hours to compute. In contrast, we show that the block approach typically uses 200x (2 orders of magnitude) less memory and runs in 500x less time.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4005W: A structural mean modelling Mendelian randomization approach to investigate the lifecourse effect of adiposity: applied and methodological considerations

Authors:

G. Power[1], T. Palmer[1], N. Warrington[2], J. Heron[1], T. Richardson[1], V. Didelez[3], K. Tilling[1], G. Davey Smith[1], E. Sanderson[1]; [1]MRC Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, [2]The Univ. of Queensland, Brisbane, Australia, [3]Leibniz Inst. for Prevention Res. and Epidemiology - BIPS, Bremen, Germany

Abstract:

The application of a lifecourse approach to genetic epidemiology is key to better understanding causal effects of adversities on health outcomes over time. For some time-varying phenotypes, it has been shown that genetic effects may have differential importance in the development of an exposure at different periods in the lifecourse. Mendelian randomization (MR) is a technique that uses genetic variation to address causal questions about how modifiable exposures influence health. MR studies often employ conventional instrumental variable (IV) methods designed to estimate lifelong effects. Recently, several extensions of MR have been used to investigate time-varying effects, including structural mean models (SMMs). SMMs exploit IVs through g-estimation and circumvent some of the parametric assumptions of other MR methods. In this study, we apply g-estimation of SMMs to MR. We aim to estimate the period effects of adiposity measured at two different life stages on cardiovascular disease (CVD), type 2 diabetes (T2D) and breast cancer in later life. We found persistent period effects of higher adulthood adiposity on increased risk of CVD and T2D. Higher childhood adiposity had a

protective period effect on breast cancer. We compare this method to an inverse variance weighted multivariable MR approach: a technique also using multiple IVs to assess time-varying effects, however, relying on a different set of assumptions and subsequent interpretations. We discuss the strengths and limitations of each approach and emphasise the importance of underlying methodological assumptions in the application of MR to lifecourse research questions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4006W: Abundant parent-of-origin effect eQTL: the Framingham Heart Study ⭐

Authors:

Y. Guan[1], D. Levy[2]; [1]Natl. Heart, Lung, and Blood Inst., Framingham, MA, [2]Framingham Heart Study, Framingham, MA

Abstract:

Parent-of-origin effect (POE) is a phenomenon whereby an allele's effect on a phenotype depends both on its allelic identity and the parent's gender from whom the allele is inherited, as exemplified by the polar overdominance in the ovine callypyge locus and the human obesity DLK1 locus. Systematic studies of POE of expression quantitative trait loci (eQTL) are lacking. In this study we use trios among participants in the Framingham Heart Study to examine to what extend POE exists for gene expression of whole blood using whole genome sequencing and RNA sequencing.

For each gene and the SNPs in cis, we performed eQTL analysis using genotype, paternal, maternal, and joint models, where the genotype model enforces the identical effect sizes on paternal and maternal alleles, and the joint model allows them to have different effect sizes. We compared models using Bayes factors to identify paternal, maternal, and opposing eQTL, where paternal and maternal effects have opposite directions. The resultant variants are collectively called POE eQTL. The highlights of our study include: 1) There are more than 2,000 genes harboring POE eQTL and majority POE eQTL are not in the vicinity of known imprinted genes; 2) Among 180 genes harboring opposing eQTL, 99 harbor exclusively opposing eQTL, and 58 of the 99 are phosphoprotein coding genes, reflecting significant enrichment; 3) Paternal eQTL are enriched with GWAS hits, and genes harboring paternal eQTL are enriched with drug targets. Our study demonstrates the abundance of POE in gene expression, illustrates additional complexity of gene expression regulation, suggest that POE can capture missing heritability, and provides a resource that is complementary to existing resources such as GTEx. We revisited two

previous POE findings in light of our POE results. A SNP residing in KCNQ1 that is maternally associated with diabetes is a maternal eQTL of CDKN1C, not KCNQ1. A SNP residing in DLK1 that showed paternal polar overdominance for human obesity is a maternal eQTL of MEG3, explaining the baseline risk of homozygous samples through association between MEG3 expression and obesity. Finally, we advised caution on conducting Mendelian randomization using gene expression as the exposure.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4007W: Accounting for confounding of different types and in large quantities in causal network inference with Mendelian randomization

Authors:

A. Fu[1,2], J. Kvamme[1], C. Tovissodé[1], B. Karki[1]; [1]Univ. of Idaho, Moscow, ID, [2]Wayne State Univ. Sch. of Med., Detroit, MI

Abstract:

Causal inference methods based on the principle of Mendelian Randomization (MR) have become increasingly popular for learning potentially complex regulatory processes in biology from observational data. Under this principle, genotypes in a natural population can be considered as randomly assigned. We can therefore identify suitable genetic variants and use them as instrumental variables for causal inference, effectively mimicking randomized clinical trials. Reducing the impact of confounding is essential in causal inference. However, limited methods are available to deal with confounding variables of different types and in large quantities. Yang et al (2017, Genome Res.) proposed three types of confounding nodes for two nodes of interest: confounder, which is the parent of the two nodes; common child, which has the two nodes as two parents; and intermediate variable, which acts as a mediator between the two nodes. These different types of confounding nodes have different impact on causal inference.

Here, we present our regression-based MRGN (MR for Genomic Networks) framework for causal network inference with MR on individual-level data that explicitly and efficiently account for confounding. Our MRGNtrio method is designed for genomic trios, each consisting of a genetic variant (as an instrumental variable) and two phenotypes (e.g., gene expression). MRGNtrio identifies confounders from a candidate pool and can account for an arbitrary number of confounders when inferring the causal model for the trio. Our MRGNgeneral method extends to larger networks by iteratively forming trios, while allowing for all three types of

confounding nodes to be accounted for. Both methods are implemented in open-source R packages available on GitHub (https://github.com/Jarred6068/MRGN; https://github.com/audreyqyfu/MRGNgeneral).
We demonstrate the performance of these methods in extensive simulations, on simulated large networks from DREAM Challenges, and on genomic data in the Genotype-Tissue Expression (GTEx) project and The Cancer Genome Atlas (TCGA) consortium. Compared to our previous causal network inference methods (Badsha and Fu, 2019, Front. Genet.; Badsha et al., 2021, Front. Genet.) and other related methods (e.g., those implemented in the popular R package bnlearn), our MRGN methods show substantially improved power and consistently high precision in detecting true edges. Application to the GTEx data identifies long-range cis-acting genetic variants, as well trans-acting ones. Application to the genomic data from the TCGA breast cancer cohort reveals diverse relationships between gene expression and DNA methylation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4009W: An Empirical Simulator of Gene Expression Profiles Capturing Regulatory Effects of Rare Genetic Variants

Authors:

X. Zhang, N. Kannan, W. Bush; Case Western Reserve Univ., Cleveland, OH

Abstract:

The increasing availability of RNA-seq, ChIP-seq, and Hi-C data has provided valuable insights into gene expression regulation, and extensive eQTL analysis further maps the impact of common genetic variation on gene expression. However, it remains challenging to fully understand the regulatory effects of rare genetic variants. Due to limitations in statistical power and data sets with limited sample size, there are few methods for analyzing rare eQTLs. To aid methods development in this area, we developed an empirical simulator of gene expression profiles that captures three key regulatory effects of rare variants: direct eQTL effects, transcription factor (TF) binding consequences, and chromatin interactions in 3D space. We estimated empirical effect sizes by aligning autosomal genes by their transcription start sites (TSS) and analyzing the distribution of these effects from real-world data, including GTEx cis-eQTL associations (at all significance levels), TF-gene regulatory scores and a *DeepLoop*-transformed Hi-C ratio matrix. By incorporating these prior data types, this empirical simulator represents a spectrum of regulatory mechanisms. We applied eQTL mapping and SKAT tests to synthetic data of 1,000 independent genes with randomly sampled rare variants, demonstrating

the robustness and accuracy of the simulated regulatory effects. As expected, we also observe spatial relationships between the tested eQTL effect size and physical distance to TSS, MAF, and Hi-C ratio-defined distance. To further simulate the surrounding and potentially confounding effects of common eQTLs, we have also piloted the use of a generative adversarial network (GAN) framework to model linkage disequilibrium patterns among variants in the cis-regulatory region of a gene. As methods for rare eQTL analysis evolve, there is a critical need for realistic synthetic data based on real genetic effects to benchmark novel analysis approaches.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4010W: An integration of local outlier factor analysis and ensemble machine learning algorithms predicts *MEFV* gene variants: Novel gene-specific voting classifier

Authors:

M. Alay; Etlik City Hosp., Ankara, Turkey

Abstract:

Introduction: Familial Mediterranean Fever (FMF) is a complex hereditary autoinflammatory disorder primarily affecting individuals of Mediterranean ancestry, including Turkish, Syrian, Armenian, and Egyptian populations. In this study, we aimed to provide an accurate classifier for MEFV gene variants. **Material-method:** We extracted 12.017 human MEFV gene variants from the Ensembl database. After the extraction, we detected 6034 missense variants. In the next step, we nominated more than 100 in silico tools for our classification model. In the feature elimination steps, we detected 3-8 features that indicated optimum performance. We selected and compared MEFV gene variants using clinvar and ensembl-based in silico tools. Therefore, we included Revel, MetaLR, and SIFT scores for analysis. During the feature engineering step, we used local outlier factor analysis to determine outliers. For machine learning method implementation, we selected defined variants to train, validate, and test the dataset. We implemented four distinct ensemble machine learning methods, namely Random Forest, Extremely Randomized Tree, XGBoost, and AdaBoost, and adjusted the hyperparameters accordingly. Next, we developed a modified hard voting classifier, a hybrid model that combined elements of a soft and hard margin classifier, to effectively discriminate between even and odd numbers in a hard voting context. After predicting known variants, we deployed our model and compared it with existing algorithms in the final step. **Results:** The XGBoost model attained the highest accuracy of 0.9882, with a standard deviation of ±0.0295, indicating exceptional performance and reliability. Extra Trees achieved an accuracy of 0.9835 (±0.0335), Random

Forest achieved an accuracy of 0.9788 (±0.0158), and AdaBoost achieved an accuracy of 0.9671 (±0.0815) when compared. In addition, we incorporated a voting classifier that employed hard voting to aggregate the predictions generated by these models. In instances where predictions were tied, we employed a custom logic that relied on soft voting probabilities. The modified voting classifier achieved an accuracy of 100% both training, and validation dataset. **Discussion:** While numerous recently developed in silico tools demonstrated improved performance on evaluation metrics, they still required optimization to enhance prediction accuracy. A more precise prediction of MEFV gene variants can be achieved by employing neighborhood-based outlier exclusion and modified hard voting classifier algorithms. This approach provides a more expedient, versatile, and economical solution.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4011W: An interpretable matrix factorization approach for GWAS summary statistics robust to sample overlap

Authors:

A. Omdahl[1], J. Weinstock[1], S. Chhetri[2], A. Battle[1]; [1]Johns Hopkins Sch. of Med., Baltimore, MD, [2]Johns Hopkins Univ., Baltimore, MD

Abstract:

Complex trait-associated genetic variation is highly pleiotropic. Several methods to characterize the latent genetic components shared among GWAS traits have been developed, including GenomicSEM, Pleiotropic Decomposition Regression, and FactorGo. Of these, matrix factorization-based approaches allow for hypothesis-free exploration by analyzing up to thousands of GWAS simultaneously to map traits and SNPs to a common genetic latent space. However, existing factorization approaches have not accounted for the effects of sample overlap between studies, which inflates GWAS-based estimates of genetic covariance. In this study, we interrogate the confounding effects of sample sharing on latent factor discovery across multiple GWAS summary statistics. We hypothesize that accounting for sample sharing in the estimation of latent pleiotropic factors yields a more accurate representation of shared genetic architecture. To test this, we developed a new method for matrix factorization of summary statistics which accounts for sample overlap between studies, “GWAS latent embeddings accounting for noise and regularization” (GLEANER). Using GLEANER, we evaluated 42 GWAS of diverse phenotypes found in both the UK Biobank (UKBB) and Finngen and used cross-trait LD score regression (LDSR) to estimate cohort-specific covariance matrices reflecting confounding from sample sharing. We introduced "shrinkage" to estimated covariance matrices and

iteratively applied decorrelating transformations to the data to evaluate a range of covariance adjustment levels. We observed that this procedure yielded more similar latent factors and estimated effect sizes between Finngen and UKBB GWAS than factorization without adjustment for sample sharing. These findings indicate that accounting for sample overlap in matrix factorization-based genetic factor analysis impacts the generalizability of estimated latent components. Furthermore, by correcting for these effects, GLEANER avoids detecting simulated spurious latent components driven by sample sharing which are detected by other common techniques (SVD and flash). In addition to its covariance adjustment procedure, GLEANER tunes sparsity constraints to automatically determine the optimal number of factors for a dataset, resulting in interpretable genetic factors that reflect pleiotropic relationships adjusted for sample overlap. These factors contain enrichments for relevant trait biology, including tissue-specific chromatin markers and evidence of selection, highlighting GLEANER's utility in characterizing shared genetic variation on complex human traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4012W: Analysis of 3.6 million individuals yields minimal evidence of meaningful pairwise genetic interactions for height

Authors:

R. Jabal-Ameli[1], M. V. Holmes[1], D. Hinds[2], 23andMe Research Team, A. Auton[1], P. Fontanillas[2]; [1]23andMe Inc., Sunnyvale, CA, [2]23andMe, Sunnyvale, CA

Abstract:

Adult height is a highly heritable polygenic trait with heritability attributable to thousands of independent SNPs. Non-additive genetic interactions may, in part, account for the difference between broad-sense and narrow-sense heritability estimates. However, prior studies have failed to identify non-additive effects, possibly due to the lack of statistical power. Leveraging 3.6M individuals of European genetic ancestry in the 23andMe research cohort, we selected 1,092 independent (separated by > 250 kb) common SNPs associated with height (GWAS p-value < 5e-8) and screened for evidence of non-additive effects. Among pairs of the 1092 SNPs, we selected those with potential evidence for non-additivity (lenient p-value < 1e-4) and, for each SNP pair, constructed a fully saturated model including additive, dominant, and epistatic (additive-by-additive, additive-by-dominance and dominance-by-dominance) terms. We tested for the presence of epistatic interactions by comparing models with and without epistatic terms using a likelihood ratio test. We adopted a Bonferroni-corrected threshold of 8.4e-8 (0.05/595,686) as the strict p-value for calling significant pairwise interactions. Among 595,686

pairwise SNPs evaluated, 72 showed evidence of pairwise GxG interactions (at a lenient p-value < 1e-4) in the initial screening, of which none surpassed a strict p-value threshold of 8.4e-8 in the fully saturated model. There was sufficient statistical power (> 80%) to detect pairwise genetic interactions at a mean effect size of 1.9mm(s) for SNP pairs with MAF≥1%. Our large-scale analysis provides further evidence of the minimal contribution of non-additivity in the genetic architecture of adult human height.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4013W: Analysis of ancestral impact on disease across Mendelian diseases

Authors:

L. Smith, J. A. Cahill, K. Graim; Univ. of Florida, Gainesville, FL

Abstract:

Lack of ancestral diversity in genomic datasets has limited our understanding of disease and it's relationship to ancestry-specific genetic variation. Since modern humans expanded beyond Africa, populations have undergone unique bottlenecking events and local adaptations for variations beneficial to them in their environment. As a result, human ancestral populations differ in ways that can affect disease progression and treatment. Due to the under-representation of non-European populations in disease data, these differences are often unaccounted for. Here we present a survey of ancestral impacts on disease via a genome-wide analysis of existing disease signatures and local adaptations. We perform this by identifying statistical enrichment in ancestry-specific alleles within existing disease signatures. Additionally, we demonstrate how machine learning methods can help mitigate ancestral bias in genomic data by targeting these enriched alleles. Using OMIM along with global population variation data, we identify ancestry-specific variants associated with genes related to a Mendelian disorder. We do this by calculating variant frequency enrichment for each ancestral population in both gnomAD and the 1000 genomes project. Subsequently we show that prioritizing genes with identified ancestry-specific variants in machine learning methods can help to overcome bias currently existing in genomic data. Our results show how using currently available population level variation data from healthy individuals can be informative of how ancestry may affect disease and can be used in disease modeling methods to counteract existing bias. With this survey, along with our presented machine learning method, we provide means to bridge the gap in sequencing data availability between different ancestral populations and disease

data. Our methods offer an alternative to the mass sequencing and ancestry calling often needed for current ancestry inclusion methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4014W: Ancient selective pressures on metabolic traits contribute to obesity and weight-related diseases in modern populations

Authors:

A. Vaughn, R. Nielsen; UC Berkeley, Berkeley, CA

Abstract:

One of the biggest global health issues of the 21st century is the stark increase in the prevalence of obesity and weight-related diseases, such as type 2 diabetes and coronary artery disease. While diet, physical activity level, and other lifestyle factors are known to play a role in the etiology of these diseases, there is growing awareness of the contribution of genetics to these interrelated phenotypes. However, a central question still remains: why are variants that contribute to such adverse phenotypes found at such appreciable frequencies in modern populations, particularly when their associated phenotypes would appear to be under negative selection? To answer this question, we here develop a method to infer selection on polygenic traits, rigorously test for changing selective pressures through time, and concurrently analyze genetically correlated traits to account for the pleiotropic effects of certain alleles. We apply our method to a set of modern and ancient genomes from western Eurasia and find two explanations for the high prevalence of variants that contribute to weight-related diseases. The first is that the selective pressures on weight-related illness phenotypes appear to have changed significantly through time, which we hypothesize was due to the introduction of agriculture to Europe and the huge dietary shifts that accompanied this major lifestyle transition. The second is the pleiotropic effect of these variants on several interrelated metabolic-associated traits, including triglyceride levels, HDL cholesterol, and BMI. We hypothesize that positive selection acting on one of several metabolic traits at ancient times, but not at more modern times, could explain the high prevalence of disease-causing alleles at the present through pleiotropy. This analysis highlights the continued contribution of ancient selective pressures on our genomes and on modern global health. We release the method we developed for this analysis, called datePALM, as a well-documented Python package that can be readily used by human genetics researchers to identify and characterize changing selection coefficients on genetically correlated traits using any combination of modern and ancient DNA.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4015W: Applications and Evaluations of Polygenic Risk Scores in Cancer Outcomes: Context Matters

Authors:

J. Shen[1], J. Baurley[2], G. King[3], D. Bogumil[3], C. Haiman[3], D. Conti[4]; [1]Univ. of Southern California, Los Angeles, MA, [2]BioRealm LLC, Mammoth Lakes, CA, [3]Univ. of Southern California, Los Angeles, CA, [4]Univ of Southern California, Los Angeles, CA

Abstract:

The performance of multi-population polygenic risk score (PRS) models is often assessed using summary measures such R^2 or AUC among out-of-sample individuals. However, PRS performance should be interpreted with caution, as multiple factors can influence validation outcomes. Here, we constructed various multi-population PRS models for four cancer outcomes (prostate, breast, lung, colorectal) using their latest multi-population GWAS summary statistics and evaluated their out-of-sample performance in the Multiethnic Cohort Study (MEC, including individuals of European, African, East Asian, Hawaiian, and Hispanic ancestry). We compared two types of PRS methodology: 1) independent SNPs that are significantly associated with the outcome and selected via pruning and thresholding or through fine-mapping and 2) approaches that incorporates all measured variants across the genome (such as “Lassosum” and “PRS-CSx”). AUC, as well as the OR of 1SD increase in standardized PRS by ancestral groups, are used as the main metrics to evaluate the out-of-sample performance of different PRS models. We found that PRS performance varies not only by construction methods and ancestral population but also by the genetic distance of an individual from the overall diversity of the discovery GWAS. In addition, PRS effects are generally better among younger individuals. For example, the OR of 1SD increase in prostate cancer PRS among the Hispanic cases and controls is 1.82 (95% CI: 1.60 to 2.07) marginally in the MEC but ranges among individuals aged 60 from 2.13 to 1.72 as the genetic distance increases. For individuals aged 70, the 1SD OR ranges from 2.03 to 1.64 for the same increase in genetic distance. These general trends hold for all four cancer types investigated, suggesting the interpretation of PRS performance should carefully consider the specific context of evaluation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4016W: ASDecoder: a semi-supervised deep learning model for decoding autism risk genes

Authors:

W. Li, Y. Shen; Columbia Univ. Irving Med. Ctr., New York, NY

Abstract:

Autism is a condition with strong but heterogeneous genetic contributions. Recent exome and genome sequencing studies have uncovered many new risk genes through de novo variants. However, a large fraction of enrichment of de novo variants observed in cases are not accounted for by known or candidate risk genes, suggesting that most risk genes are still unknown. Additionally, the contribution of rare inherited variants to autism genetic risk is not well understood due to limited statistical power. We hypothesize that autism risk genes share cell-type specific expression patterns in the developing brains of neurotypical individuals, and such information can be quantified to improve the statistical power of detecting new risk genes. To this end, we developed a generative model, ASDecoder, that captures gene regulatory network information to inform the prior probability of bottleneck genes during the brain developmental process. We obtained large-scale single-cell RNA-seq data from neurotypical human brains collected through a range of developmental stages, collected known high-confidence autism risk genes from the previous reports, as well as the information regarding transcriptional regulatory networks, signal transduction networks, and brain-specific gene-gene interaction networks. We calculated the cosine similarity score of each gene with the known autism risk genes based on gene embeddings learned from the generative model. Using data from recent exome sequencing studies of autism, the ASDecoder score achieves better performance in prioritizing de novo and rare inherited variants than other methods, especially for genes that are less intolerant of loss of function variants. With the unprecedented growth of single-cell transcriptomics data and expanding autism cohorts with exome or genome sequencing, ASDecoder will facilitate the discovery of novel risk genes and understanding of the etiology of autism.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4017W: Assessing the impact of social exposomes on genetic liability for bipolar disorder

Authors:

R. Sharp[1], L. Glover[2], L. Raffield[1], M. Love[1]; [1]UNC-Chapel Hill, Chapel Hill, NC, [2]Duke Univ., Durham, NC

Abstract:

Polygenic risk scores (PRS) estimate genetic liability for complex disorders and have been used across medical disciplines to dissect clinical heterogeneity and inform early risk prevention. However, polygenicity and environmental effects limit PRS utility in psychiatry. Our understanding of how genetic risk interacts with social factors to influence disorder manifestation remains limited. Bipolar disorder (BP), a highly heritable and complex disorder, has shown SNP-based heritability estimates of only 18.6% from the current largest BP genome-wide association study (GWAS), despite twin studies showing estimates of 60-90%. This discrepancy suggests the presence of factors not captured by common variation, implicating a role of environmental factors yet to be elucidated.

Recent studies estimate that up to 40% of BP patients are misdiagnosed, with rates varying due to cultural context and implicit biases. For example, Black Americans are more likely to be misdiagnosed with schizophrenia compared to BP. Such misdiagnoses prevent patients from receiving appropriate treatment, exacerbate health inequities, and make assignment of case status in research studies difficult, hindering result interpretability.

This gap in our ability to accurately estimate risk for complex neuropsychiatric disorders could be ameliorated using risk-estimation methods that account for the effects of complex social factors such as healthcare utilization and perceived stress scores (PSS). Here, we leverage data from the All of Us (AoU) Research Program, proposing analyses that combine epidemiological approaches and new statistical genetics tools such as PheMED, which allows for the comparison of study power between GWAS and phenotyping strategies. First, we quantify multicollinearity and missingness of social variables in AoU, revealing correlations between social risk factors such as PSS and discrimination in medical settings. Additionally, we observe significant shifts in social risk scores by BP case status, with BP cases (N = 3,962) having a mean PSS of 17.48 (out of 40), while controls (N = 28,989) have a mean of 12.68 ($p = 2.2e-16$). Next, we apply PheMED to quantify phenotypic misclassification dependent on case status definition, ancestry, and degree of social risk. Finally, we fit a model quantifying the joint effect of social risk factors and PRS for BP on disorder manifestation. Completion of these analyses could reveal the impact of unaccounted-for social factors on genetic liability for BP, underscore the need to account for social factors in quantifying disorder risk, and contribute to the refinement of PRS applications in clinical settings.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4018W: Association analysis of affected cancers with subtype of *TP53* gene pathogenic variants in Li-Fraumeni syndrome

Authors:

T. Sanomachi[1], M. Ushiama[2], M. Goto[3], H. Cho[1], N. Tanabe[1], T. Watanabe[1], M. Matsukawa[1], S. Oda[1], Y. Odaka[3], T. Koyama[1], H. Sakamoto[2], K. Sugano[4], T. Yoshida[2], M. Hirata[2]; [1]Natl. Cancer Ctr. Hosp., Tokyo, Japan, [2]Natl. Cancer Ctr. Hosp./Natl. Cancer Ctr. Res. Inst., Tokyo, Japan, [3]Natl. Cancer Ctr. Res. Inst., Tokyo, Japan, [4]Natl. Cancer Ctr. Hosp./Kyoundo Hosp., Sasaki Fndn., Tokyo, Japan

Abstract:

PURPOSE: Li-Fraumeni syndrome (LFS) is a rare tumor-predisposition syndrome caused by a germline pathogenic variant of the tumor suppressor gene *TP53*. Genotype-phenotype correlation analysis of *TP53* variants has been conducted considering the types of variants, such as loss-of-function and dominant-negative effects, however the correlation between phenotypes and subtypes of *TP53* pathogenic variants has not been sufficiently verified. **METHODS:** Forty-five patients were selected, including 25 patients with *TP53* pathogenic variants detected by genetic testing, and 20 of their first-degree relatives with a history of LFS-related cancer, young-onset cancer, or multiple cancers at the National Cancer Center Hospital (Tokyo, Japan) from November 2014 to January 2024. Of these, 3 cases where pathogenic variants were suspected of postzygotic mosaicism (Variant allele frequency<30%) and 1 case of comorbidity with other hereditary tumor syndromes were excluded. Finally, 41 were included for retrospective review of the medical record data. **RESULTS:** Of the 41 patients from 18 families, 25 (61.0%) were female, and the most common age at first onset was 30-39 years (34.1%). The most common types of cancer were breast cancer (15 cases, 36.6%), soft-tissue sarcoma (12 cases, 29.3%), and lung cancer (12 cases, 29.3%). *TP53* pathogenic variants were different in all 18 families. 26 (63.4%) had missense variants (MV), and 15 (36.6%) had protein-truncating variants (PTV), including nonsense, frameshift, and splice-site variants. In the comparative analysis of cancer-type frequency between MV and PTV, the incidence rate of cancer during child and adolescence tended to be higher in MV than in PTV, while age at first onset was comparable. There was no statistically significant difference in the frequency of any cancer-types between MV and PTV. Correlation analysis of cancer types revealed relatively high correlation between cancer types, such as soft-tissue sarcoma and osteosarcoma (r=0.59) in male. Analysis of cancer, which caused mortality, revealed that lung cancer was the most frequent in both MV (5 cases, 33.3%) and PTV (4 cases, 50.0%). **CONCLUSION:** We analyzed the cancer incidence status of LFS patients at our hospital, through stratifying them by variant type and sex. We continue further accumulation of LFS cases, aiming at establishment of more precise risk evaluation algorithm.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4019W: Association analysis of rare genetic variants in circulatory and respiratory phenotypes among participants from the *All of Us* Research Program ⭐

Authors:

J. Zhang[1], S-h. Hong[2], X. Wang[3], S. J. Jurgens[3,4], C-T. Liu[1], J. Dupuis[1,5], P. L. Ellinor[3], G. T. O'Connor[6,7], Q. Mei[1], S. Choi[1]; [1]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [2]Div. of Endocrinology and Metabolism, Dept. of Internal Med., Ewha Womans Univ. Coll. of Med., Seoul, Korea, Republic of, [3]Cardiovascular Disease Initiative, The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Dept. of Experimental Cardiology, Amsterdam UMC, Univ. of Amsterdam, Amsterdam, Netherlands, [5]Dept. of Epidemiology, Biostatistics and Occupational Hlth., McGill Univ., Montreal, QC, Canada, [6]Pulmonary Ctr., Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, [7]The Natl. Heart Lung and Blood Inst.'s Framingham Heart Study, Framingham, MA

Abstract:

Background: Gene-based rare variant analysis through large-scale whole-genome sequencing and whole-exome sequencing has been widely used to improve our understanding of the complex mechanisms underlying human diseases. However, the statistical power of detecting causal genes that are associated with diseases of interest remains limited. Alternative splicing and tissue-specific expression further complicate the identification of relevant genes.
Methods: We proposed a novel approach to conduct gene-based rare variant tests using whole genome sequencing and phenotypic data based on approximately 250,000 participants from the All of Us Research Program. This approach integrated multiple known transcripts which might vary in expression by tissue or condition. Rare high-confidence loss-of-function and predicted deleterious missense variants with a minor allele frequency less than 0.1% were obtained by annotating variants in exon regions with the LOFTEE and dbNSFP database, using the Variant Effect Predictor. Association analyses were conducted for 118 binary phenotypes (42 respiratory and 76 circulatory) defined by matching the Phecode map with International Classification of Diseases (ICD) codes. A single p-value was obtained for each gene by aggregating test results from all transcripts (canonical, non-canonical, and pseudo transcripts) for a gene. All p-values were adjusted using the False Discovery Rate (FDR) method, and a q-value less than 0.05 was considered significant. For phenotypes from significant associations, the proposed approach was applied with rare synonymous variants to assess the type I error.

Results: Among 118 binary phenotypes, 22 significant associations were identified, including one novel association. Notable increases in statistical significance were observed for associations related to gene *TTN* and cardiomyopathy when using all transcripts ($p=1.1\times10^{-27}$), compared to tests using only the canonical transcript ($p=2.4\times10^{-20}$). A smaller p-value for the association between deep vein thrombosis and gene *PROC* was obtained ($p=5.6\times10^{-7}$), compared to the result from the UK Biobank, which had a sample size twice as large ($p=5.0\times10^{-6}$). This association would not achieve the statistical significance threshold if only the canonical transcript was used ($p= 3.5\times10^{-6}$). The analyses of rare synonymous variants confirmed effective control over type I error rates.
Conclusion: The proposed approach successfully improved the statistical significance of the rare variant tests by effectively navigating and aggregating information from phenotype-relevant transcripts when available, while controlling the type I error rate.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4020W: Associations of combined genetic and lifestyle risks with incident type 2 diabetes in the UK Biobank

Authors:

C. Zhao[1], K. Hatzikotoulas[2], R. Balasubramanian[1], E. Bertone-Johnson[1], N. Cai[3], L. Huang[3], A. Huerta[4], M. Janiczek[1], C. Ma[1], R. Mandla[5], A. Paluch[1], N. W. Rayner[2], L. Southam[6], S. Sturgeon[1], K. Suzuki[7], H. J. Taylor[8], N. VanKim[1], X. Yin[9], C. Lee[1], C. Spracklen[10]; [1]Univ. of Massachusetts Amherst, Amherst, MA, [2]Inst. of Translational Genomics, Helmholtz Zentrum München, German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [3]Helmholtz Zentrum München, Neuherberg, Germany, [4]Broad Inst., Cambridge, MA, [5]The Broad Inst., Cambridge, MA, [6]Inst. of Translational Genomics, Neuherberg, Germany, [7]Ctr. for Genetics and Genomics Versus Arthritis, Ctr. for Musculoskeletal Res., Div. of Musculoskeletal and Dermatological Sci., Univ. of Manchester, Manchester, United Kingdom, [8]Ctr. for Precision Hlth.Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [9]Nanjing Med. Univ., Nanjing, China, [10]Univ. of Massachusetts, Amherst, MA

Abstract:

Type 2 diabetes (T2D) is predominantly caused by an interplay between genetic and lifestyle factors. Both genetic susceptibility and unhealthy lifestyle were found to be associated with elevated T2D risk. However, the combined effect of genetic risk and behavioral lifestyle on T2D is less well studied. To determine whether poor modifiable health behaviors were associated with similar increases in the risk of incident T2D among individuals with low, moderate, and

high genetic risk, we performed a genetic risk score (GRS) by lifestyle interaction analysis within 332,436 individuals from the UK Biobank.
Multi-ancestry GRS were calculated by summing the effects of 1,286 T2D-associated variants (number of risk alleles multiplied by the reported effect estimates); low, moderate, and high GRS were defined by tertiles of GRS. We generated lifestyle factor variables using baseline body mass index (BMI) and self-reported data on smoking status, physical activity level, and diet quality, each categorized as ideal, intermediate (not applicable for diet quality), or poor. An overall lifestyle classification was then created combining all individual lifestyle factors, and defined as an ideal, intermediate, or poor level.
Multivariable Cox proportional hazards models were used to generate adjusted hazards ratios (HR) and their associated 95% confidence intervals (CI). Age, sex, annual income, years in education, Townsend deprivation index (TDI: an area-based measure of socioeconomic status), and 16 principal components (population structure) were considered as covariates in the time-to-event analysis. During follow-up (median: 13.5 years), 13,148 (4.0%) participants developed T2D, with a median of 8.0 years to incidence. Genetic risk (moderate: HR=1.69, 95% CI 1.61 - 1.78; high: HR=3.02, 95% CI 2.88 - 3.16) and lifestyle classification (intermediate: HR=2.38, 95% CI 2.24 - 2.53; poor: HR=6.89, 95% CI 6.38 - 7.45) were independently associated with increased risk for T2D. Compared with ideal lifestyle, poor lifestyle was associated with a substantially increased risk within and across all genetic risk strata, with HR ranging from 7.69 to 21.48. Overall, poor lifestyle and high genetic risk were the strongest risk factors for incident T2D in our study. Our study results suggested that individuals at all levels of genetic susceptibility of T2D can greatly mitigate their risk through their behavioral lifestyle.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4021W: Bayesian network imputation methods applied to multi-omics data identify putative causal relationships in a type 2 diabetes dataset containing incomplete data: An IMI DIRECT Study

Authors:

R. Howey, A. Viñuela, The IMI DIRECT consortium, H. J. Cordell; Newcastle Univ., Newcastle Upon Tyne, United Kingdom

Abstract:

We report the results from Bayesian network analysis of data from a large pan-European study of type 2 diabetes (T2D) conducted by the DIRECT consortium. Data were collected for 3029 individuals (795 with T2D and 2234 without) within 7 different study centres, with the data

comprising genotypes, proteins, metabolites, gene expression measurements and many different clinical variables. The data was analysed with Bayesian network methods implemented in the BayesNetty software package, which can handle mixed discrete/continuous data with missing values. The original dataset consisted of over 16,000 variables, which were then filtered down to 261 variables for analysis. Even with this reduction, not one individual had complete data for all variables, making it impossible to analyse using standard Bayesian network methodology. However, using the recently proposed novel imputation method implemented in BayesNetty, we were able to compute a large average Bayesian network from which we could infer possible associations and causal relationships between variables of interest. Our results confirmed many previous findings in connection with T2D, including some possible mediating proteins and genes, some of which have not been widely reported. We also confirmed potential causal relationships with liver fat that were identified in an earlier study that also used the DIRECT dataset.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4022W: Benchmarking gAIRR-Call in identifying *T Cell Receptor* alleles from whole genome sequencing data

Authors:

K-T. Huang[1], Y-H. Yang[1], M-J. Lin[2], C-Y. Yao[3,4], S-K. Lai[5,6], C-L. Hsu[7,1,8], Y-C. Yang[9,3], C-Y. Chen[10], P-L. Chen[1,6,5,11,4], S-J. Hsu[1]; [1]Graduate Inst. of Med. Genomics and Proteomics, Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [2]Dept. of Computer Sci., Johns Hopkins Univ., Baltimore, MD, [3]Dept. of Lab. Med., Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [4]Graduate Inst. of Clinical Med., Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [5]Genome and Systems Biology Degree Program, Academia Sinica and Natl. Taiwan Univ., Taipei, Taiwan, [6]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [7]Dept. of Med. Res., Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [8]Graduate Inst. of Oncology, Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [9]Dept. of Clinical Lab. Sci. and Med. Biotechnology, Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [10]Dept. of Biomechatronics Engineering, Coll. of Bioresources and Agriculture, Natl. Taiwan Univ., Taipei, Taiwan, [11]Div. of Endocrinology and Metabolism, Dept. of Internal Med., Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Abstract:

Immunoglobulins (IG) and T-cell receptors (TR) constitute adaptive immune receptor repertoire (AIRR). AIRR is encoded by somatically recombined *IG* or *TR* from a pool of gene segments: variable (V), diversity (D), and joining (J) gene segments, each of which encompasses numerous

alleles. The gAIRR-call tool was developed to characterize germline *IG* and *TR* (gAIRR) alleles from the short-read targeted panel sequencing, while the gAIRR-annotate was utilized to accurately annotate gAIRR on whole genome assembly data, released from the Human Pangenome Reference Consortium (HPRC). With the increasing prevalence of whole genome sequencing (WGS), it is necessary to benchmark gAIRR-call for the short-read WGS data with 30x, the typical read depth in many biobanks. In this study, we aim to evaluate the performance of gAIRR-call in identifying germline *TR* alleles from 30x WGS data. We generated 132 callsets by calling 44 subjects from the HPRC across three data types: WGS data (PE 150 bp) and targeted panel data (PE 150 bp and 300 bp). We compared each callset with the annotation results. Callsets for WGS data had a lower average recall rate: 72.5±17.6% in *TR* D gene segments (TR_D), 98.5±1.6 % in TR_J, and 87.1±4.1% in TR_V, compared to those for panel data. On the other hand, WGS data had higher average precision in each gene segment: 100% (TR_D), 100% (TR_J), and 98.6±0.9% (TR_V). After integrating flanking sequences and fine-tuning gAIRR calling process, the recall rate for WGS data was improved to 100% in TR_D, 99.7±0.6% in TR_J, and 91.2±3.1% in TR_V, without largely depreciating the precision (100%, 100% and 96.5±1.4%, respectively). Additionally, we conducted a trio analysis of two GIAB families. The mendelian violation rates for HG002 were 0% in TR_D, 2.2% (2/91) in TR_J, and 1.55% (3/193) in TR_V. For HG005, the rates were 0% in TR_D, 1.14% (1/88) in TR_J, and 1.19% (2/168) in TR_V. Both trio-analytical results were consistent with previous studies using manual inspection. In conclusion, we have demonstrated that the modified gAIRR-call provides enough comprehensive and accurate identification of *TR* alleles from 30x WGS short-read data. This improvement not only expands the applicability of gAIRR-call but also helps pave the way for further advancements in immunogenomic research and personalized medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4023W: Benchmarking of missense variant effect predictors using known gene-trait associations

Authors:

R. Hoffing[1], A. Holleman[1], A. Deaton[1], L. Krohn[2], P. LoGerfo[1], M. Plekan[1], S. Akle Serrano[1], L. Ward[1]; [1]Alnylam Pharmaceuticals, Cambridge, MA, [2]Alnylam Pharmaceuticals, Montreal, QC, Canada

Abstract:

Predicting the effects of variants on disease is a challenging task but remains a critical need. Variant effect predictors (VEP) provide an *in silico* method for assessing the impact and

pathogenicity of rare missense variants, and thoroughly assessing these VEPs is crucial in determining the appropriate predictive threshold at which a missense variant is considered deleterious. This is necessary for performing rare-variant association studies (RVAS) that inform drug discovery efforts, and for interpretation of variants in known disease-causing genes in a clinical setting. Some, but not all, pathogenic missense variants will lead to destabilization and degradation of their *cis*-protein. In large-scale RVAS, deleterious missense and predicted protein-truncating variants (PTV) can often be analyzed together as they are expected to have shared effects on phenotypes. In this work, we identify a set of "gold standard" associations caused by loss-of-function variants that can be used to determine the appropriate VEP threshold at which to include missense variants alongside PTVs in burden testing. To do so, we performed a large-scale RVAS in the UK Biobank in genes and a broad set of quantitative traits, including physical measures, blood and urine biochemistry, cognitive measures, metabolomics, telomere length, lifestyle data, and family history. We retain the most significantly associated phenotype per gene, identifying 377 phenotypes surpassing $p < 1 \times 10^{-7}$. We then manually annotated these associations based on whether there was a known molecular mechanism through which biochemical activity of the encoded protein affects the measured trait. This was true for 172 gene-trait associations which became our gold standard. Many of these included genes that have been pursued as current or historic drug targets, such as *PCSK9* and LDL cholesterol, *CETP* and HDL cholesterol, and *APOC3* and triglycerides. We compared these gold standard associations with missense variants scored from four VEPs (CADD, REVEL, Alpha Missense, and ESM1b). We benchmarked current VEP thresholds by their ability to recapitulate the loss-of-function phenotypes among approximately 400,000 UK Biobank participants. For ESM1b, we use this method to identify a conservative (LLR < -11.519) and lenient (LLR < -9.382) threshold for use in RVAS. We show that our ESM1b suggestive thresholds return a similar number of significant RVAS results as those previously recommended for Alpha Missense, CADD, and REVEL, demonstrating the framework and utility of benchmarking across a known standard of gene-trait associations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4024W: Beyond the Liability Threshold: A Theoretical Framework for Genetic Disease Risk Based Upon Multiple Pathways ⭐

Authors:

M. Costantino[1,2], L. Wang[1,2], R. Fonseca[1,3], A. Dahl[1]; [1]Dept. of Med., Section of Genetic Med., Univ. of Chicago, Chicago, IL, [2]The Committee on Genetics, Genomics and Systems Biology, Univ. of Chicago, Chicago, IL, [3]Dept. of Human Genetics, Univ. of Chicago, Chicago, IL

Abstract:

Genome-wide association studies (GWAS) have associated complex diseases with thousands of loci across the genome, but how they relate to causal disease biology often remains unclear. The standard framework for disease risk is the liability threshold model, where an individual's risk depends on the sum of individual genetic effects. However, genetic effects on disease likely operate through diverse pathways which are likely to interact non-additively. Type 2 diabetes, for instance, could arise from complex interactions between genetic effects on insulin production and insulin sensing. To address this issue, we introduce the Latent Liability Pathway (LLP) model, in which multiple axes of liability exist. Under this model, the genetic architecture of a disease is characterized by a set of pathways, where an individual has a disorder if their liability on any pathway exceeds a healthy threshold. We developed an expectation-maximization algorithm (EM-LLP) to fit this model, which takes a case/control label and GWAS hits as input and partitions the GWAS hits into distinct disease pathways. In realistic simulations with ~90K individuals, we find that EM-LLP can reliably recover the latent liability pathways with high heritability. To validate our model in read data, we analyzed data from ~90K unrelated White British individuals from the UK Biobank. Specifically, we created a positive control pseudo-phenotype that was attributed to individuals in the top deciles of BMI and eosinophil count and input this variable, along with the top GWAS hits for BMI and eosinophil count, into EM-LLP to obtain estimates for the two pathways. First, we used top GWAS SNPs as predictors and found that EM-LLP was unable to recover the genetic effect sizes for either pathway (correlation between true and predicted values $p > 0.05$). However, this is expected from our simulations, given that GWAS SNPs explain very little variance in these phenotypes (~0.01). To evaluate EM-LLP in a high-power scenario, we replaced the SNP predictors with non-genetic variables such as sex, age, blood biomarkers, and top principal components to explain more variance. We found that across 13 of the 176 "pseudo-phenotypes," EM-LLP could predict genetic effects significantly associated with the true values. Next, we will replace the non-genetic variables with polygenic scores and different types of stratified polygenic scores. Our results show the potential of this novel method to endotype disease, as well as the challenges of studying traits with weak SNP effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4025W: BinMap: A novel Method to Finemap Genetic Associations with Binary Traits with an Application to T2D

Authors:

L. Marsh, W. Gan, J. Howson, N. Koelling; Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom

Abstract:

Finemapping genetic associations can identify potential causal variants and thereby help to pinpoint the biological mechanism underlying the association. Finemapping methods currently established in the literature, such as SuSiE and FINEMAP, usually assume a linear genotype-phenotype relationship and normally distributed additive noise. While these assumptions may be justifiable for many quantitative traits, they break down for binary traits such as disease states in studies containing cases and controls. This issue becomes particularly apparent when finemapping with summary association statistics, where established finemapping methods assume z-scores are calculated from a linear model while they are, in fact, derived from a logistic model. In this work, we propose a new finemapping method, BinMap, which models a logistic relationship between genotype and phenotype.
BinMap combines a stochastic shotgun search with variational inference methods to finemap under the logistic model assumptions. We highlight BinMap's utility on simulated binary phenotypes and real-world examples given by three known Type 2 Diabetes (T2D) loci in the UK Biobank (UKBB). The modelling assumptions of SuSiE and FINEMAP are violated by these data and the loci thus represent a challenge to these methods. We show that on both simulated and real-world data, even after parameter tuning, SuSiE does not converge and FINEMAP returns very large credible sets (containing more than a hundred variants). By contrast, BinMap accurately finemaps the simulated data and returns small credible sets, containing only a handful of variants in strong LD. On the UKBB data, our method managed to disentangle potential genetic signals associated with T2D by returning small, plausible credible sets, highlighting candidate variants for further investigation. BinMap is thus able to give insights into relevant disease associations which are out of reach for established finemapping methods that assume a linear model.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4026W: CADET: Enhanced transcriptome-wide association analyses in admixed samples using eQTL summary data

Authors:

T. Head, Q. Dai, J. Schildkraut, D. Cutler, J. Yang, M. Epstein; Emory Univ., Atlanta, GA

Abstract:

A transcriptome-wide association study (TWAS) is a popular statistical method for identifying genes whose genetically-regulated expression (GREx) component is associated with a trait of interest. Most TWAS approaches fundamentally assume that the training dataset (used to train the gene expression prediction model) and testing GWAS dataset are from the same ancestrally homogenous population. If this assumption is violated, studies have shown a marked negative impact on expression prediction accuracy and power of the downstream gene-trait association test. These issues pose a particular problem for admixed individuals, whose genomes represent a mosaic of multiple continental ancestral segments. To resolve these issues, we present CADET, which enables powerful TWAS of admixed cohorts leveraging the local-ancestry (LA) information of the cohort along with summary-level eQTL data from reference panels of different ancestral groups. CADET combines multiple polygenic risk score models based on the summary-level eQTL reference data to predict LA-aware GREx components in admixed test samples. Using simulated data, we compare the imputation accuracy, power, and type I error rate of our proposed LA-aware approach to LA-unaware methods for performing TWAS. We show that CADET performs optimally in nearly all settings regardless of whether the genetic architecture of gene expression is dependent or independent of ancestry. We further illustrate CADET by performing a TWAS of 29 common blood biochemistry phenotypes within an admixed cohort from the UK Biobank and identify 15 hits unique to our LA-aware strategy. The majority of hits uniquely identified by CADET have consistent evidence in the GWAS literature and harbor one or more significant GWAS variants for the relevant traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4027W: Causal relevance of adiposity for infectious diseases in diverse populations

Authors:

R. Huang, Z. Chen, T. Iain, C. Kartsonaki, B. Fiona; Univ. of Oxford, Oxford, United Kingdom

Abstract:

Background: On a background of escalating rates of excess adiposity globally, the causal relevance of adiposity for infectious disease risks remains unclear. **Methods**: We analysed data from ~512,000 China Kadoorie Biobank (CKB) participants in observational analyses and ~76,000 in genetic analyses, and from ~502,000 UK Biobank (UKB) participants. A transethnic BMI genetic score was developed using dosages of 851 bi-allelic SNPs extracted from BMI GWAS in CKB and UKB. Cox regression yielded adjusted hazard ratios (HRs) associating measured body mass index (BMI) with infectious disease incidence and mortality. Linear and non-linear

Mendelian randomization (MR) assessed the associations of genetically-predicted BMI. **Results**: Overall, the mean (SD) BMI was 23.8 (3.2) kg/m^2 in CKB and 27.4 (4.8) kg/m^2 in UKB. During ~12 years of follow-up, 102,866 and 125,598 participants experienced at least one infectious disease hospitalisation or death in CKB and UKB, respectively. In observational analyses in both CKB and UKB, there was a positive log-linear association between measured BMI and risk of both respiratory and non-respiratory infections at BMI &gt 22.2 kg/m^2, and weaker inverse associations at lower BMI levels. Excess risks at lower BMI levels were more extreme for mortality due to both respiratory and non-respiratory infections. Non-linear MR analyses showed positive log-linear associations of genetically-predicted BMI with overall risks of incident respiratory and non-respiratory tract infections. These associations were modestly stronger in UKB (respiratory: 1.13 [95% CI 1.11-1.14] per 1 SD higher genetically-predicted BMI; non-respiratory: 1.11 [1.10-1.13]) than in CKB (respiratory: 1.04 [1.00-1.17]; non-respiratory: 1.04 [1.00-1.08]). Among non-respiratory infections, there were notably strong associations of genetically-predicted BMI with hepatobiliary infections (UKB: 1.19 [1.14-1.25]; CKB: 1.12 [1.05-1.21]) and with skin, soft tissue and bone infections (UKB: 1.18 [1.16-1.20]; CKB 1.09 [0.97-1.22]). Although limited by small sample size, there was an apparent positive log-linear association between genetically-predicted BMI and risk of death due to infectious diseases, stronger for non-respiratory (UKB: 1.29 [1.15-1.44]; CKB: 1.54 [1.07-2.20]) than for respiratory (UKB: 1.06 [0.98-1.14]; CKB: 1.10 [0.88-1.38]) tract infections. **Conclusions**: Higher BMI was causally associated with higher incidence and mortality of infectious diseases, similar in European and East Asian ancestry populations. Excess risks of respiratory infections at lower measured BMI levels may reflect residual confounding and reverse causality.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4028W: Comparison of Bayesian methods for rare variant association using whole exome sequencing and medical record data in the UK Biobank

Authors:

J. Shen[1], Y. Xie[1], H. Zhao[2]; [1]Yale Univ., New Haven, CT, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract:

Rare variants explain a substantial portion of phenotypic variance and play an important role in complex disease etiology. However, the commonly used methods for rare variant tests often suffer from low power and inflated type I error rates in imbalanced case-control studies.

Incorporating additional information, such as phenotype structure and patient similarity, can potentially boost the power of rare variant tests. In this study, we compared the performance of three Bayesian methods, including PheWAS, TreeWAS, and Similarity Regression, to identify associations between ICD-coded patient phenotypes and genotypes through both simulations and applications to 177,504 UK Biobank participants with whole exome sequencing data. Simulations showed the better performance of TreeWAS than the other two methods. When these methods were applied to investigate the associations between 101 genes and 18,904 clinical terms, we found that TreeWAS was able to identify many more association signals than PheWAS and Similarity Regression. Notably, 232 significant associations not detected by PheWAS were identified by both TreeWAS and Similarity Regression. Our results demonstrate the robustness and effectiveness of methods that leverage the tree structure of phenotypes to identify genetic associations, particularly for rare diseases and small sample size studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4029W: Coupling foundation models with chromatin QTL analysis reveals unique genetic-epigenetic associations that emerge in opioid overdose

Authors:

F. Qureshi[1], T. Lambert[2], V. Evans[2], A. Hoang[1], R. Sallari[3], M. Chakraborty[3], B. Quach[4], K. Lovrenert[5], C. Willis[4], C. Bartels[5], P. Scacheri[5], S. Akbarian[2], E. Johnson[4], O. Corradin[6,1]; [1]Whitehead Inst., Cambridge, MA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Axiotl Inc, Cleveland, OH, [4]RTI Intl., Durham, NC, [5]Case Western Reserve Univ., Cleveland, OH, [6]MIT, Cambridge, MA

Abstract:

Opioid use disorder, a major public health crisis, is associated with epigenetic modifications that influence gene expression in the brain. The extent to which these alterations are driven through genetic versus environmental factors remains unclear. Employing statistical and machine learning (ML) techniques to analyze chromatin data offers a promising approach to determine whether the epigenetic alterations associated with opioid exposure are primarily driven by underlying genetic variation.
To investigate this, postmortem brain samples from the nucleus accumbens (NAc) of a mixed ancestry cohort were genotyped and analyzed using ChIP-seq for H3K27 acetylation. After standard quality control filtering, the cohort included 47 opioid overdose cases and 44 accidental death controls. Genotype information was combined with ChIP data and analyzed

using an allele-specific statistical framework (RASQUAL) to identify H3K27ac quantitative trait loci (hmQTLs), regions where differences in H3K27ac occupancy are associated with genetic variants. In total, 3,412 unique chromatin regions were observed to have statistically significant hmQTL-associated variants across both cases and controls (q-value < 0.05). Performing RASQUAL analysis on the case and control groups separately revealed a subset of hmQTLs (n=692) that were specific to opioid overdose cases. Interestingly, case-specific hmQTLs were observed proximal to genes previously associated with substance use disorders, such as SULT1A1, PNPLA2 and DTNA.
We next sought to validate the context-dependent relationships of genotype and regulatory element activity by leveraging pretrained DNABERT2 ML models. Focusing on regions with significant hmQTLs, we generated custom models to predict ChIP-seq activity from sequences for individual genetic regions ranging from 400-2500 bp. These models were fine-tuned and validated using control NAc data on a peak-to-peak basis to ensure consistent accuracy and performance. The models' accuracy was then tested on held-out controls. DNABERT2 models attained an average accuracy of 86.8% in control samples. However, when the same models were applied to opioid cases, prediction accuracy dropped significantly ($\bar{x}$ = 76.9%, p-value <0.001) with 23% of models having an accuracy less than 60% (compared to 0% for controls). These results suggest that sequence data alone may not sufficiently account for differences in regulatory element activity associated with opioid overdose. Ultimately, furthering our understanding of the interaction between genetic variants and environmental exposures will be critical for understanding complex traits like opioid addiction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4030W: Deciphering the genetic structure underlying childhood asthma heterogeneity

Authors:

R. Vernet[1], C. Linhard[1], A. Estermann[1], Y. Suzuki[2], F. Demenais[1], H. Aschard[2], H. Julienne[2,3], E. Bouzigon[1]; [1]INSERM, Université Paris Cité, UMR 1124, Group of Genomic Epidemiology of Multifactorial Diseases, Paris, France, [2]Inst. Pasteur, Université Paris Cité, Dept. of Computational Biology, Paris, France, [3]Inst. Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris, France

Abstract:

Asthma is the most common chronic disease in children, with various clinical expressions that reflect the existence of distinct asthma endotypes. These endotypes likely involve different

genetic pathways.
We aimed to characterize the genetic structure underlying endotypes of childhood-onset asthma (COA) using genetic information from multiple biological and physiological traits involved in the disease.
For that purpose, we built a harmonized, comprehensive database of GWAS summary statistics containing (i) 34 studies of asthma and asthma subtypes (childhood-onset, adult-onset, etc.) and (ii) 217 studies of 58 different asthma-related traits (blood cells counts, various cytokines, lung function phenotypes, etc.). We identified 228 independent SNPs ($r^2 < 0.1$) associated with COA at $P < 5x10^{-8}$ and we built the matrix of 228 SNPs by 58 traits association results, with Z-scores oriented to COA risk-increasing alleles. Then we applied a partition around medoids (based on the cosine distance between the 228 Z-score vectors) to cluster SNPs with similar patterns of associations across the 58 asthma-related traits. Finally, we mapped SNPs to the closest gene (+/-10kb) and assessed enrichment of each cluster for gene sets belonging to biological pathways from GO biological process and Reactome databases and for gene expression in specific tissues and cell types.
We identified five clusters of COA-associated SNPs that showed distinct patterns of associations with asthma-related traits and were all enriched for multiple immune-related pathways at FDR < 5%. In particular, one cluster was associated with decreased levels of many cytokines (including CCL24, CXCL12 and IL12) but increased levels of CCL5, and was primarily enriched for PD1 signaling pathway ($q = 7.3x10^{-5}$) and for gene expression in innate lymphoid cells (ILCs) in fetal spleen, lung and adrenal gland ($P < 7.6x10^{-4}$). Another cluster was associated with increased levels of Lymphotoxin-α and enriched for virus-related pathways (viral entry into host cell, type I interferon production, toll-like receptor 3 cascade; $q < 5\%$) and for gene expression in dendritic cells (DCs) in the lung ($P = 7.3x10^{-4}$).
This study allowed identification of five genetic clusters underlying COA. One cluster was related to the PD1 signaling pathway, whose activation by PD1 agonist was previously shown to suppress lung inflammation in mice (PMID: 32778730), and another cluster was related to response to virus, supporting the contribution of viral infections in childhood-onset asthma. Thus, these clusters point to distinct pathways involved in COA and might help identify new potential therapeutic targets.
Funding: ANR-20-CE36-0009

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4031W: Deep learning estimation of lumbar vertebral and inter-vertebral morphology from the UK Biobank dual energy X-ray dataset

Authors:

K. Landheer[1], A. Kwon[2], J. G. Raya[2], B. Geraghty[1], J. Herman[1], J. L. Marchini[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron Pharmaceuticals, Inc, Tarrytown, NY

Abstract:

Lower back pain is a prevalent condition, with approximately 25% of people experiencing an episode of one month or longer across their lifespan. Chronic back pain is a multifactorial disorder, with most low back pain sufferers having some degree of intervertebral disc degeneration (IDD), typically in the lumbar. IDD is characterized by the loss of proteoglycans and dehydration of the disc, macroscopically resulting in a reduction in disc height. Despite its prevalence, limited effective treatments are available for IDD and uncovering its genetic basis may help in identifying targets for therapeutics. To date, the largest IDD GWAS performed (58,854 cases, 922,958 controls) identified 41 variants at 33 unique loci. Notably, the largest effect on IDD was from a rare loss of function variant in SLC13A1.

We assessed the intervertebral and vertebral morphology in the lumbar spine from 41,628 sagittal dual energy X-ray (DXA) images in the UK Biobank (UKB). We trained a UNet model to segment six vertebrae (L12-L5) with training data of 270 manually labeled examples. From these segmentations we used custom image-processing routines to measure intervertebral height from (T12-L1 through L4-L5, inclusive), vertebral thickness and area (T12 through L5, inclusive) and lumbar spinal curvature.

We performed GWAS and ExWAS on individual disc or vertebrae phenotypes as well as the mean disc or vertebrae morphology, with age, age^2, sex, height and body-mass index as covariates. For the mean disc thickness we found 77 index SNPs across 39 unique loci. Of these 9 were a recapitulation of the largest GWAS of IDD. The mean disc thickness GWAS p-values were compared to the p-values from the large case-control ICD10 GWAS, with excellent concordance being observed, thereby providing strong validation that these phenotypes are accurate. One gene with non-synonymous burden mask had a significant association (from a unified test strategy, GENE_P): SLC13A1 (P = 6.6E-7, previously found by the large IDD GWAS). For mean vertebral height (T12-L4) we found 67 index SNPs across 45 loci. For mean vertebral area (T12-L4) we found 104 index SNPs across 54 loci. For spinal curvature we found 6 index SNPs across 6 loci.

As the UKB expands their longitudinal cohort to 60,000 participants, this will enable the measurement of progression of disc height, which is hypothesized to be a better indicator of disc health than height alone in a sample size amendable to GWAS and ExWAS studies. Overall, this study clearly demonstrates the improved power of image-derived quantitative traits over binary phenotypes, despite the much smaller sample size

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4032W: Deep learning prediction of RNA binding proteins reveal variant-effect scores for Alzheimer's disease association studies

Authors:

G. Bartels-Quansah[1,2], A. Das[2,3], C. Lakhani[2], D. Knowles[2,3]; [1]Barnard Coll., New York City, NY, [2]New York Genome Ctr., New York City, NY, [3]Columbia Univ., New York City, NY

Abstract:

RNA binding proteins (RBPs) are principal regulators of splicing, a key component of the transcription process known to play an important role in neurodegenerative disorders such as Alzheimer's disease (AD). While several RBPs, including TIA-1 and FUS, have been found to play a significant role in AD pathogenesis (Rybak-Wolf and Plass, 2021), there is limited understanding of RBP-specific variant-level effects. Integrating disease-relevant functional annotations into methods such as fine-mapping and gene-based rare variant tests has become increasingly common, but existing applications have not yet included RBP-specific information in AD association tests. Traditional prediction models are primarily restricted to regressions and are unable to capture the complex interactions of RBPs. This has motivated the development of several machine learning models such as RNAProt (Uhl et al., 2021) and Seqweaver (Park et al., 2020), which have shown impressive results in binding predictions. We aim to extend these methods by designing a model specifically trained and optimized for AD-relevant RBPs. Specifically, we develop a 2-layer fully-connected convolutional neural network (CNN) that utilizes observed peaks from RBPs to predict the binding affinity of RBPs on AD variants by generating binding effect scores. The model takes in reference and alternative RNA sequences (generated by mutating a central base according to the variant allele) from RBPs from the ENCODE eCLIP data consortium in FASTA format and outputs variant annotations. We use 180 bp sequences and perform a 5-fold cross-validation, partitioning the dataset by chromosome. Each fold is trained separately, and we compute the average AUC and standard error for each RBP. Preliminary results show that our CNN outperforms traditional regression methods. Specifically, K562_FUS and K562_TIA1 models achieve an average AUC score of above .90 In comparison, our baseline logistic regression model yields AUC scores of 0.8765 and 0.8140 respectively. These results are very high and require further analysis. We use our trained model for variant effect prediction on the latest release of ADSP (Alzheimer's Disease Sequencing Project) data, which consists of over 300 million variants. These functional predictions will provide crucial insights into the genetic landscape of AD and can be applied to other complex neurological diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4033W: Detecting and correcting discrepancies in summary statistics for genetic association studies

Authors:

H. Sun[1], R. Dong[1], X. Cao[1], R. Feng[1], K. Dey[2], Y. Zou[3], G. Wang[1]; [1]Columbia Univ. in the City of New York, New York, NY, [2]Mem. Sloan Kettering Cancer Ctr., New York, NY, [3]Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

Summary statistics are commonly used to present discoveries from genetic association studies. Approaches such as fine-mapping and TWAS incorporate information from linkage disequilibrium (LD) to identify potential causal variants and genes using these summary statistics. However, it is widely acknowledged that discrepancies between summary statistics and LD can significantly influence the results of many summary statistics-based methods, particularly leading to severe false discoveries. Methods such as DENTIST and SLALOM have been used in practice to address this issue. However, in our experience implementing these methods for real-world datasets, we recognized that they still have limited capacity to fully mitigate the problem, particularly in the presence of complicated LD structures and multiple causal variants.

We propose a new method, RSS QC, to correct summary statistics and LD mismatch for association studies. As an extension to the SuSiE RSS model for fine-mapping, RSS QC is shown to be effective in reducing false positives and increasing power in fine-mapping applications, although the approach can also help other applications involving summary statistics.

We conducted a systematic comparison of RSS QC with DENTIST and SLALOM, simulating z-scores derived from meta-analyses of different underlying LD structures. On average, the ROC curve for assessing sensitivity and specificity for RSS QC has an AUC of 0.84, compared to 0.70 for DENTIST and 0.74 for SLALOM.

We applied RSS QC to multiple Alzheimer's disease GWAS studies (Bellenguez 2022, Jansen 2019, Kunkle 2019, Wightman 2021, Beecham 2014, and Strickland 2020) with different cohorts using the Alzheimer's Disease Sequencing Project LD reference developed in-house. A total of 128 unique loci (defined by 95% credible sets, CS) were identified, with the Bellenguez dataset containing the most loci (79). We also analyzed the Metabrain cis-eQTL dataset, which comprises summary statistics from five different brain tissues. After RSS QC and fine-mapping, we identified 27,861 95% CS for the cortex, 8,540 for the cerebellum, 1,124 for basal ganglia, 779 for the hippocampus, and 497 for the spinal cord.

In conclusion, RSS QC can enhance the consistency between summary statistics and LD reference. Therefore, it is recommended for summary statistics-based analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4034W: Detecting interpretable, genetically-driven patterns of brain structure using a data-driven deep learning approach

Authors:

M. De Pian[1], J. Wen[2], C. Davatzikos[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Southern California, Los Angeles, CA

Abstract:

Background: Advances in brain imaging genetics have propelled the field of precision medicine by generating imaging-derived phenotypes linked to an array of genetic variants. This is particularly relevant in polygenic neurodegenerative diseases like Alzheimer's Disease (AD), in which genetic variants exert pleiotropic effects along the whole genome. Here, we introduce a deep learning model that provides interpretable summarizations of brain structural networks jointly guided by variations from brain MRI and genetics.
Methods: Our approach integrates two generative models: an orthonormal projective Nonnegative Matrix Factorization (opNMF) autoencoder and a Variational Autoencoder (VAE). This combination is designed to identify genetically influenced patterns of brain structure from MRI data, partitioning the brain into interpretable anatomical brain elements (ABEs) that correlate with specific genetic variations, including disease-associated ones, in a data driven manner. We evaluated our model on a set of simulated data (N = 1500) to validate patterns of brain structure influenced by genetic variations and applied it to clinical data from CN, AD, and MCI patients (N = 1564) from the ADNI dataset.
Results: The model robustly detected specific imaging and genetic patterns and their associations in both simulated and real voxel-wise volumetric measurements and SNPs. In clinical settings, it detected AD-informed ABEs and established a comprehensive brain anatomy. When applied to the ADNI data, the model effectively grouped brain regions based on structural and genetic covariance, providing a parcellation distinct from those based solely on structural data. Notably, in extracting 5 ABEs, the model differentiated subcortical regions and the posterior superior lobe of the cerebellar vermis and hemisphere from the rest of the cerebellum and the frontal lobe was divided into genetically and non-genetically influenced areas. These examples are showcasing two differentiations that were not observed when genetic information was excluded. We also pinpointed specific SNPs linked to each ABE. For instance, rs429358 (APOE) and rs10948363 (CD2AP) were associated with the genetically informed ABE of the temporal and hippocampus regions, aligning with existing literature, and several CD2AP polymorphisms (rs9381563, rs9473117, rs10948363) were grouped together as highly

associated variants with a specific ABE.
Conclusion: These insights highlight the model's ability to design a brain atlas that captures the complexity of genetic and neuroanatomical heterogeneity, refining our understanding of how genetic factors influence brain anatomy.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4035W: Dissecting the biological mechanisms behind type 2 diabetes causal effects on non-cardiometabolic comorbidities.

Authors:

A. Ana Arrudas[1], O. Bocher[2], D. Cammann[3], S. Yoshiji[4], X. Yin[5], C. Zhao[6], H. Taylor[7], J. Chen[8], A. Wood[9], K. Suzuki[10], J. Mercader[11], C. Spracklen[12], J. Meigs[13], **J. Rotter**[14], M. Vujkovic[15], B. Voight[15], A. Morris[16], E. Zeggini[17]; [1]Helmholtz Zentrum München, Neuherberg, Germany, [2]Helmholtz Zentrum Muenchen, Neuherberg, Germany, [3]Nevada Inst. of Personalized Med., Las Vegas, MD, [4]Broad Inst., Cambridge, MA, [5]Nanjing Med. Univ., Nanjing, China, [6]Univ. of Massachusetts Amherst, Amherst, MA, [7]Natl. Human Genome Res. Inst., Bethesda, MD, [8]UNLV-NIPM, Las Vegas, NV, [9]Univ Alabama at Birmingham, Birmingham, IN, [10]Tokyo Univ, Tokyo, Japan, [11]Broad Inst Harvard & MIT, Cambridge, MA, [12]Univ. of Massachusetts, Amherst, MA, [13]Massachusetts Gen Hosp, Boston, MA, [14]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [15]Univ. of Pennsylvania, Philadelphia, PA, [16]Univ. of Manchester, Manchester, United Kingdom, [17]Helmholtz Munich /Helmholtz Zentrum Muenchen GmbH, Neuherberg / Munich, Germany

Abstract:

Type 2 diabetes (T2D) is a heterogeneous disease characterized by the interplay of different cell-specific biological mechanisms. In addition to cardiometabolic complications, T2D is epidemiologically associated with other diseases including musculoskeletal, respiratory, neuropsychiatric, and eye conditions.
The T2D Global Genomics Initiative (T2DGGI) has defined eight non-overlapping mechanistic clusters of index variants from multi-ancestry T2D GWAS meta-analysis (2,535,601 individuals including 428,452 cases) that represent distinct pathways to disease (including beta-cell dysfunction and obesity). To assess putative causal relationships between T2D liability and non-cardiometabolic comorbidities, we performed bidirectional two-sample Mendelian randomization (MR) analyses. In addition, we conducted cluster-specific MR analyses to assess the effect of distinct biological mechanisms driving T2D liability on the different comorbidities. We used matching ancestry data for both diseases, meta-analyzing across ancestries when

possible.
We find evidence of a causal effect of overall T2D liability on several comorbidities, including glaucoma, anorexia, cataracts, and osteoporosis. In some cases, the causal effect appears to be driven by specific clusters. For instance, we find evidence that the causal effect of T2D liability on cataracts (OR=1.06 [1.04, 1.08], FDR adjusted P=4.33x10-9) is driven by the obesity cluster (OR=1.12 [1.07, 1.17], FDR adjusted P=3.24x10-6). We also detect cluster-specific effects in the absence of an overall causal effect. For example, although no overall causal effect of T2D liability on osteoarthritis is observed, we recapitulate the well-established link between T2D, osteoarthritis, and obesity by showing evidence of a causal effect of the obesity cluster on osteoarthritis (OR=1.18 [1.1.4, 1.22], FDR adjusted P=2.65x10-17). Conversely, we find evidence that the T2D liability linked to beta-cell dysfunction positively associated with proinsulin has a protective effect against osteoarthritis (OR=0.96 [0.94,0.99], FDR adjusted P=1.97x10-2).
Our results shed light on specific biological mechanisms, represented by the different mechanistic clusters of T2D risk variants, that are shared between T2D and non-cardiometabolic comorbidities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4036W: ENPACT predicts personalized DNA methylation patterns from DNA sequences using a combined linear and deep-learning approach

Authors:

E. Wu[1], T. Adeluwa[1], S. Gona[2], E. Thompson[1], K. Aracena[2], L. Barreiro[2], M. Chen[2], C. Ober[2], H. Im[2]; [1]The Univ. of Chicago, Chicago, IL, [2]Univ. of Chicago, Chicago, IL

Abstract:

DNA methylation is an important epigenetic regulatory mechanism that may play a role in the unsolved etiology of many complex diseases, but its heterogeneity across specific cell-type, tissue, and disease contexts limits the generalizability of available methylation data. To address this, the ENPACT approach uses the reference genome outputs of Enformer, a deep-learning model that generates epigenomic and transcriptomic features from input DNA sequences, as inputs to train cell-type specific elastic net models of methylation. Models trained across 77 primary cell types achieved on average prediction correlations of 0.55 and AUROC of 0.83. We found that applying matched cell-type ENPACT models to real genotype data led to predictions in concordance with both in vivo measurements and in silico predictions from the current state-of-the-art linear models. We achieved significant correlations of on average 0.14 in terms of

variation within individual genomes, as well as absolute average correlations of 0.27 across different individuals when comparing with real data, and higher when comparing to other model predictions. Because reference-trained models are agnostic as to which allele causes the observed variance in methylation, the sign of the correlations across individuals is negative approximately half the time, which will necessitate new methodological adaptations when applying ENPACT predictions to downstream analyses. The ability to train ENPACT with only existing methylation data and produce personalized, context-specific methylation levels with only DNA sequence data may allow for broader examination of the role of DNA methylation in complex traits in a convenient, generalizable, and cost-efficient manner.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4037W: Estimating the effect of opioid prescribing on addiction with Mendelian randomization

Authors:

J. Rich; Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Abstract:

The opioid crisis continues to worsen amid steady increases in overdose mortality among most populations over the past two decades in the USA. Although white affliction originally brought attention to the crisis, increasing law enforcement involvement in prescribing surveillance has exacerbated police interactions and, possibly, mortality among disenfranchised communities. Black opioid overdose rates surpassed white mortality for the first time in 2020, despite significantly higher rates of opioid use disorder (OUD) among whites, which signals a major disparity. Moreover, this analysis is the first attempt to measure the probability of addiction following opioid prescription for various populations to determine how medical exposure to opioids contributed to the current crisis. Evaluating genomic samples linked to electronic health records (EHR) among All of Us volunteers, genes that increase the probability of receiving an opioid prescription, such as those linked to Denisovan heritage that would cause a 3-rooted third molar, are set as instrumental variables to isolate the effect of opioid prescribing on developing a substance use disorder (SUD) within 10 years initial exposure, controlling for age of first exposure. The population includes 1,027 All of Us participants from Version 7 who received an opioid prescription and had available genomic information. Preliminary estimates find that the probability of developing an addiction after an initial opioid prescription was not statistically significant to the rest of the population ($p > 0.05$). The estimates also align with population-level extrapolations from the National Survey on Drug Use and Health (NSDUH) based on the

prevalence of initial medical exposure for one year and subsequent new diagnoses of SUD. Moreover, differences in opioid overdose mortality may not be explained by variations in opioid prescribing and might be motivated by exposures outside the medical setting.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4038W: Estimating the length distribution of non-crossovers from gene conversions

Authors:

M. Hardarson[1,2], G. Palsson[1], B. Halldorsson[1,2]; [1]deCODE genetecs, Reykjavik, Iceland, [2]Reykjavik Univ., Reykjavik, Iceland

Abstract:

Genetic diversity is governed by mutations and meiotic recombinations. Meiotic recombinations split into crossovers and non-crossovers. Both have been studied in the past but non-crossovers are difficult to detect and thus less studied. Most non-crossovers do not leave any trace in the offspring's genome but those which do are observed as gene conversions. We have previously developed a method (NCOurd) to estimate length distributions of non-crossovers from gene conversions. We have now extended the NCOurd framework to better model the prevalence of gene conversions within non-crossovers. Our new method obtains better estimates for this prevalence using a maximum likelihood estimate of hyperparameters, computed within the NCOurd framework. We show that the extended NCOurd framework is accurate and can be efficiently used on large datasets of gene conversions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4039W: Evaluating the role of imputing gene expression in the brain for psychiatric phenotypes.

Authors:

J. Drake[1], Z. Taylor[1], A. Denham[2], S-A. Bacanu[3], J. Shin[4], T. Hyde[4], V. Vladimirov[1]; [1]Univ. of Arizona, Phoenix, AZ, [2]Texas A&M Univ., College Station, TX, [3]Virginia Commonwealth Univ, Richmond, VA, [4]Lieber Inst. for Brain Devel, Baltimore, MD

Abstract:

Background: Due to the inherent difficulties and challenges associated with measuring gene expression (GE) in the brain, GE imputation algorithms have gained traction. Additionally, methods that capture expression quantitative trait loci (eQTLs) can leverage much larger genome-wide association studies (GWAS) to impute the expression of millions of subjects. However, only a limited number of genes can be imputed with accuracy and overfitting remains a large concern, which limits the broad adoption of these methods. In this study, we've (1) replicated PrediXcan's elastic net methodology using a large transcriptome dataset, (2) expanded on the methodology by incorporating cell proportions (CP), and (3) evaluated the utility of these newly created models. **Methods:** Genotype and amygdala expression data was obtained for 187 controls, 231 with major depressive disorder (MDD), and 123 with bipolar disorder (BP). In total, 25,212 mRNA models with 10-fold cross-validation were generated. To assess overfitting, we reserved 30% of the dataset for testing. Previously, we observed that incorporating CP estimates increases performance but is impacted by the software used to generate those estimates. Hence, we integrated CATD software into our pipeline, which evaluated 28 mRNA deconvolution methods for estimating cell fractions from pseudo-bulk amygdala tissue. After generating genotype-only (GT) and CP models, differential expression (DE) was conducted. To understand if the CP models are capturing psychiatric-related eQTLs, matrixEQTL was used to conduct an eQTL analysis. **Results:** Our GT models obtained 45% more quality gene models (N = 4,298) then the off-the-shelf models from predictdb.org. Our CP models imputed a much higher number of genes (N = 16,659) and with higher accuracy; the median Pearson correlation between the measured and imputed expression of the testing dataset was 0.422. When comparing the percent of DE imputed genes (nominal p-value ≤ 0.10) that overlap with the measured results (FDR p-value ≤ 0.10), the percent is similar for BP and MDD. However, the Pearson correlation of t-statistics is much stronger with the CP models than GT models for both BP (0.778 vs 0.122) and MDD (0.927 vs 0.285). Literature review of the eQTL analysis revealed top genes are implicated in BP (i.e., *ROR2*) and neurons (i.e., *EMP3*). **Conclusion**s: Our GT models outperformed publicly available models for the amygdala. Additionally, we showed the inclusion of CP estimates can drastically increase the number and accuracy of quality gene models. However, current methods to estimate CP utilize mRNA GE, which limits its utility when researchers are interested in imputing mRNA GE.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4040W: Exploring the age-related cognitive impairment with data from Taiwan Biobank : A genome-wide association study

Authors:

M-C. Tsai[1], C-W. Fang[2], C-Y. Yang[2]; [1]Natl. Cheng Kung Univ. Hosp., Tainan City, Taiwan, [2]Natl. Cheng Kung Univ., Tainan City, Taiwan

Abstract:

Alzheimer's disease (AD), the leading etiology of dementia, is considered to have a significant genetic component. Genome-wide association studies (GWAS) from several large cohorts revealed multiple loci relative to AD. However, few studies focus on mild cognitive impairment (MCI), the transitional stage of AD. We conducted a GWAS on a cohort retrieved from the Taiwan Biobank (TWB). The Mini-Mental State Examination (MMSE) was used as the proxy for aging-related cognitive impairment. A score lower than 24 on the MMSE was defined as cognitive impairment. A final sample of 21,212 individuals were retrieved, and 462,992 genetic markers passed quality control. Results showed that only two signals surrounding GFRA2 (rs1383582, β = -0.22±0.047, P = 2.30 x 10E-6) and C10orf11 (rs4746390, β = 0.095±0.047, P = 8.23 x 10E-6) were suggestively (P < 10E-5) associated with continuous MMSE scores in the linear model, while three signals surrounding ZMYM2 (rs180836599, odds ratio = 1.444±0.145, P = 4.49 x 10E-6), DHX9 (rs2282264, odds ratio = 1.651±0.11, P = 5.33 x 10E-6), and FTCD (rs141888659, odds ratio = 1.909±0.11, P = 7.70 x 10E-6) were suggestively associated with low MMSE scores the logistic model. To our knowledge, our present study is the first GWAS on cognitive aging in the Taiwanese population. However, we failed to observe any signal reaching a genome-wide significant correlation with meaningful MCI determined by the MMSE score in a nationally representative cohort. Further studies with larger cohorts and more definite diagnoses are indicated.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4041W: Exploring the impacts of multi-generational environment sharing on phenotypic covariance

Authors:

M. Ianni-Ravn[1], J. Novembre[2]; [1]Univ. of Chicago, Chicago, IL, [2]Univ Chicago, Chicago, IL

Abstract:

Traditional complex trait genetics models often assume that non-genetic (environmental) factors are experienced independently of genetic relatedness. However, this assumption is violated for traits with multi-generation shared environments (MGSEs) — such as exposure to air pollution or dietary composition — that impact phenotypic variation. Studies of phenotypic covariance in modern biobanks, which contain hundreds of thousands of cryptic distant

relatives (e.g., 3rd to 5th-degree cousins), may be significantly affected by complex environment sharing. In this work, we carry out simulations and empirical analyses to investigate the impact of MGSEs on phenotypic covariance.
Using simulation tools, we explore various models to characterize phenotypic covariance patterns induced by MGSEs and assess the bias introduced by MGSEs on genetic association statistics in each model. Preliminary results indicate that phenotypic covariance can be inflated due to MGSEs, particularly among closely related individuals, and this may drive observed differences in heritability as function of genetic relatedness. We also examine how variation in genetic relatedness among relatives of the same degree can be used to demonstrate the contribution of MGSEs to phenotypic variation and to distinguish these effects from direct and indirect genetic influences.
Our findings underscore the importance of accounting for MGSEs in genetic studies to refine heritability estimates, improve the accuracy of genetic association studies, and gain a complete understanding of complex trait variation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4042W: Exploring the potential of protein domain-specific burden tests to unveil insights into protein function and association with disease risk

Authors:

J. Rodriguez-Flores, B. Ye, K. Sun, N. Parikshak, Regeneron Genetics Center, S. Gelfman, T. De, S. Balasubramanian, J. Backman, A. R. Shuldiner; Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

Domains are the functional building blocks of proteins, and methods designed to identify deleterious and pathogenic variants are not sensitive to the domain-specific roles of specific amino acid substitutions in protein function. The aim of this project is to broadly assess the value of a domain-specific burden test, as opposed to burden tests focused on predicted loss of function (pLoF) or predicted deleterious missense (pDM) variants anywhere in a protein sequence. The long-term objective of the project is to develop domain-specific burden tests for all known protein domains, although the current focus of this study is variants that add or remove Cysteine (Cys-altering) residues in Epidermal Growth Factor-like repeat (EGFr) domains. These domains are highly conserved across species and always contain six Cys residues that form three disulfide bonds. In a recent preprint it was demonstrated that a burden test limited to Cys-altering variants in *NOTCH3* EGFr domains identified a genome-wide-significant

association between *NOTCH3* and stroke in UKB 450 K data (p value = $6.3x10^{-10}$), compared to a nominal association for LoF + pDM burden test (p value = 0.033). Motivated by this discovery, we sought to test the same hypothesis on a broader set of 211 human genes containing at least one EGFr domain, 1197 EGFr domains in total (up to 37 per gene). In these EGFr-containing genes, we identified, 15,137 Cys-altering missense variants in 450 K UKB exomes, including 4,268 within EGFr domains. PheWAS was conducted for 10,924 UKB phenotypes across 874 EGFr domain Cys-altering variants with >5 carriers to identify 228,786 weak (p value < $5.0x10^{-2}$); 4,442 nominal (p value < $5.0x10^{-4}$); 186 suggestive (p value < $5.0x10^{-6}$) and 75 significant (p value < $5.0x10^{-8}$) associations for EGFr domain Cys-altering variants. These included previously reported associations with stroke and white matter hyperintensity for *NOTCH3* p.Arg1231Cys (Rodriguez-Flores et al., 2023, Medrxiv). In this study, we re-analyzed UKB phenotypes to compare statistical association signal (p value and effect size) between burden tests focused on EGFr domain Cys-altering, pDM and pLoF variants. An initial assessment focused on *NOTCH* family genes (*NOTCH1*, *NOTCH2*, *NOTCH3*, *NOTCH4*), followed by a broader assessment of all 211 EGFr domain-containing genes. This study has the potential to advance the field of genetics by discovering novel associations missed by state-of-the-art analysis methods, and to accelerate therapeutic development by providing clues into the pathological mechanism of missense variants associated with disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4043W: Extending REGENIE's whole genome regression framework for increased power in large-scale association analysis

Authors:

J. Mbatchou, J. Marchini, Regeneron Genetics Center; Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

Genetic association analyses have now been routinely performed on large-scale data sets through the formation of biobanks that derived extensive phenotypes from electronic health care records and self-reported information. They have been instrumental in target discovery and finding new uses for existing therapeutics. The REGENIE method, which we previously proposed, has been widely used in the field of genetics. It is a computationally efficient method for analyzing thousands of quantitative and binary phenotypes in the presence of population structure, relatedness and polygenicity. It has been extensively applied in the field of genetics and has supported the finding of several new targets using imputed and whole exome sequenced data with extended functionalities such as gene-based testing and interaction

testing. REGENIE's whole genome regression (WGR) model, which computes polygenic score predictions, is closely tied to the infinitesimal model often used in linear mixed model approaches. It assumes similar effect sizes across a set of variants (respective of the minor allele frequency) and thus can lose power when violations to the assumption are present in the data. To address this, we propose a more flexible WGR approach that better captures the effects of strongly associated variants when building the polygenic scores. We demonstrate the benefits of this enhanced model through simulation studies and real data applications on several phenotypes in the UK Biobank, where we observe up to 5% increase in the mean $\chi 2$ statistics for highly polygenic traits. We have implemented the model in the publicly available REGENIE software to facilitate the identification of new genetic associations in large-scale analyses and refine our understanding of the genetic basis of health-related outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4044W: Flexibly encoded genome-wide association study identifies novel nonadditive genetic risk variants for cardiometabolic traits

Authors:

J. Zhou[1,2], A. Garao Rico[3], L. Guare[3], Million Veteran Program, P. Tsao[1,2], T. Assimes[1,2], S. S. Verma[3], M. A. Hall[3]; [1]VA Palo Alto Hlth.care System, Palo Alto, CA, [2]Stanford Univ. Sch. of Med., Palo Alto, CA, [3]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Most genome-wide association studies (GWAS) assume an additive inheritance model, where heterozygous genotypes (HET) are coded with half the risk of homozygous alternate genotypes (HA), leading to a focus on additive genetic effects for complex diseases research. Yet, growing evidence indicates that with flexible modeling, many single-nucleotide polymorphisms (SNPs) show nonadditive effects, including dominant and recessive, which will be missed using only the additive model. We developed Elastic Data-Driven Encoding (EDGE) to determine the inheritance model each SNP demonstrates for a given trait, which can be subsequently used to uniquely and flexibly encode SNPs for GWAS. Simulation results demonstrated that EDGE outperformed additive and other genetic encoding models across a wide range simulated inheritance models for power while maintaining a conserved false positive rate. EDGE GWAS were performed in UK Biobank and Million Veteran Program (MVP) data, comprising over half a million participants, to study coronary artery disease (CAD) and body mass index (BMI). Results

replicated many previously known associations. For instance, variants, near the 9p21 locus and *FTO*, were identified by EDGE for CAD and BMI, respectively. Of these replicating results, less than 48% (47.6% for BMI and 42.44% for CAD) were determined by EDGE to demonstrate an additive inheritance model. Novel variants associated with CAD (15) and BMI (4) were additionally identified using EDGE, with 67% (BMI) and 25 (CAD) exhibiting nonadditive effects. For instance, we identified a novel SNP at *FTO* (rs35420030, additive, $p=4.18 \times 10^{-8}$) significantly associated with BMI but was only linked to lung function previously. We also identified two new SNPs in the 9p21 region (*CDKN2B-AS1*) for CAD, rs28557075 (sub-additive [between recessive and additive], $p=1.29 \times 10^{-9}$) and rs62560775 (sub-additive, $p=1.97 \times 10^{-8}$), previously associated with multiple cancers. Additional nonadditive SNP that associated with CAD were identified including rs17881561 (*FCSK*, sub-additive, $p=3.49 \times 10^{-9}$) related to red blood cell features, blood pressure, and lipoprotein cholesterol levels in previous GWAS. This research lays the necessary groundwork for integrating nonadditive genetic effects into GWAS workflows to identify novel disease-risk SNPs, which may ultimately improve polygenic risk prediction in diverse populations and springBoard future applications to thousands of disease phenotypes to improve disease-prediction capability.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4045W: From cortex to cortex: the shared genetic components of kidney function and Alzheimer's Disease

Authors:

D. Yang, Y. Yang, X. Zhu, W. S. Bush; Case Western Reserve Univ., Cleveland, OH

Abstract:

Alzheimer's disease (AD), a leading cause of death among older Americans, has a long preclinical phase with a complex, multifactorial etiology and progression, but lacks effective treatments beyond targeting brain proteinopathy. Prior research has suggested an inconsistent link between kidney function and dementia risk, including AD and vascular dementia, with some studies reporting no significant correlation. Despite the crucial role of the kidneys in blood filtration for the clearance of proteins like β-amyloid and blood pressure regulation to maintain vascular integrity, the contribution of kidney function to AD pathogenesis has been relatively underexplored. Here we aimed to compute genetic correlations, identify overlapping genetic loci, and delineate shared biological pathways between AD and kidney function traits to gain a better understanding of their shared genetic etiology. We applied the bivariate causal mixture

model (MiXeR) to assess the number of shared genetic variants and explore the overlapping genetic architectures between AD and estimated glomerular filtration rate (eGFR). We analyzed recent genome-wide association study (GWAS) results from non-Hispanic White populations, including AD (n=7,264) and eGFR (n=1,014,789) datasets. Based on MiXeR models, eGFR is considerably more polygenic (1.2 K to 14.3 K causal variants) than AD (0.1 K to 8.8 K causal variants). Our analysis reveals that the overall genetic correlation between eGFR and AD is not strong (r_g = 0.007, SE = 0.09), indicating no global genetic association. However, we observed strong genetic correlations localized to chromosomes 12 (r_g = -0.15), 16 (r_g = -0.10), and 19 (r_g = -0.08). The genetic overlap with mixed effect directions between AD and kidney function demonstrates a complex genetic relationship between these conditions, in line with their established comorbidity association. Moreover, there is extensive phenotype heterogeneity within AD, and individuals that have a more vascular component may have a stronger genetic link to kidney function compared to other AD subtypes. Despite a small genome-wide correlation, the identification of specific regions with strong correlation signals suggests that any influence eGFR may have on AD progression is likely through specific biological pathways. Ongoing analyses include data from individuals of African descent and additional kidney function indicators like blood urea nitrogen (BUN) and Cystatin C. Subsequently, we plan to validate these findings in independent cohorts and functionally characterize the identified loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4047W: Genetic and non-genetic factors impacting MRI-estimated liver iron in the iron overload genetic disease *HFE* hemochromatosis

Authors:

M. Lucas, L. Pilling, J. Correa Delgado, J. Atkins; Univ. of Exeter, Exeter, United Kingdom

Abstract:

Hemochromatosis, a common genetic iron overload condition, increases systemic iron levels which can result in liver disease and cancer. 1 in 150 persons of Northern European descent and 1 in 300 non-Hispanic whites in the United States are homozygous for the high-risk *HFE* C282Y genotype (C282Y+/+). Penetrance, however, is variable, and the factors influencing the levels of iron deposition are not fully understood. We compared MRI-estimated liver iron across *HFE* genotypes and investigated factors associated with increased liver iron concentration.
37,287 UK Biobank participants of European genetic ancestry had *HFE* genotyping and magnetic

resonance liver iron concentration (MRLIC) measurements (2014-2020); the mean male and female age was 64.8 (SD7.7) and 63.5 years (SD7.5). Linear regression models assessed associations between environmental factors, blood biomarkers, and iron-related polygenic scores with MRLIC, adjusted for age, sex, and genetic principal components.
Within *HFE* genotype groups, liver iron was markedly higher in those undiagnosed with hemochromatosis compared to those diagnosed, and presumably treated. In undiagnosed participants, C282Y+/+ had significantly higher MRLIC compared to those without mutations (males: 2.56mg/g vs 1.27, $p<0.001$; females: 2.31mg/g vs 1.22, $p<0.001$). Increases in MRLIC in other *HFE* C282Y/H63D genotype groups were very minor compared to those with no mutations.
In a combined group of *HFE* genotypes (excluding C282Y+/+), liver iron levels were positively associated with a high polygenic score for transferrin saturation and serum ferritin ($\beta=0.33$, $p=1.5*10^{-30}$ & $\beta=0.19$, $p=2.3*10^{-10}$). In the same group, lifestyle factors associated with MRLIC included alcohol consumption >30 units per week ($\beta=0.13$, $p=7.1*10^{-61}$), smoking ($\beta=0.06$, $p=7.8*10^{-6}$), red and processed meat intake ≥3 times per week ($\beta=0.04$, $p=1.4*10^{-15}$), and proton pump inhibitor medication ($\beta=-0.05$, $p=1.5*10^{-12}$). An underweight body mass index (18.5kg/m^2) was negatively associated with liver iron ($\beta=-0.08$, $p=3.1*10^{-3}$) and in females, waist-hip-ratio showed a positive relationship ($\beta=0.11$, $p=7.2*10^{-3}$). In C282Y+/+'s only, drinking 4 or more green/black teas per day was negatively associated with liver iron concentration ($\beta=-0.96$, $p=3.8*10^{-2}$); a similar relationship was found for higher HbA1c levels ($\beta=-0.34$, $p=3.3*10^{-3}$). There was excess liver iron in *HFE* C282Y homozygotes undiagnosed with hemochromatosis, but low penetrance of iron loading in other *HFE* genotype groups. Genetic and environmental factors may exacerbate iron loading among *HFE* genotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4048W: Genetic association meta-analysis is susceptible to confounding by between-study cryptic relatedness

Authors:

T. Tu, A. Ochoa; Duke Univ., Durham, NC

Abstract:

Meta-analysis is recommended and used by many researchers to enhance power, improve runtime, and to better account for heterogeneity in Genome-Wide Association Studies (GWAS). The independence of samples is an important assumption to consider, yet there is unclear guidance for how much between study relatedness is permissible in meta-analyses. As a

motivating example, recent work that suggested applying sex-stratified meta-analysis to correct for participation bias did not consider the potential relatedness between individuals of opposite sex from the same population. Confounding due to cryptic relatedness across meta-analyzed studies has been ignored by many researchers and cannot be corrected by modeling relatedness solely within studies, which can be done through kinship matrices in standard GWAS Linear Mixed Models. The presence of cryptic relatedness between studies violates the independence assumption, leading to correlated residuals and inflated test statistics, which contributes to increased type I error.
We characterize the effects of different between-study relatedness scenarios, particularly population structure and recent family relatedness, on meta-analysis type I error control and power. We consider both sex and subpopulation stratified analyses by testing simulated replicates with (1) no family relatedness between subpopulations, (2) family relatedness within subpopulations, (3) family relatedness across subpopulations, and (4) single population with family relatedness. Additionally, we compare the performance of joint vs meta-analysis of 1000 Genomes Project genotypes with simulated traits (n = 2504 with 1,111,266 loci from 5 major ancestry groups), and plan to extend our analysis to real datasets with larger sample size and more complex family pedigree. We perform GWAS using a logistic mixed-effects model implemented in SAIGE R-software package and meta-analysis with command-line tool METAL. We run joint and meta-analyses on simulations using both binary and quantitative traits for comparison. Results from simulated data show that for scenarios with family relatedness (cases 2-4), we observe severe inflation and relatively lower AUC for sex-stratified meta-analysis compared to joint and subpopulation meta-analysis. We recommend avoiding meta-analysis of studies that share the same populations, especially sex-stratified analysis, because male and female populations are highly likely to contain substantial cryptic relatedness.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4050W: Genetic Basis of Partner Choice

Authors:

Q. Zheng[1], S. V. Alten[2], T. H. Lyngstad[3], E. Ciscato[4], Z. Sun[1], J. Miao[5], Y. Wu[1], B. Zheng[1], M. Nivard[6], T. Galama[7], P. Turley[8], P-A. Chiappori[9], A. Havdahl[3], J. Fletcher[10], Q. Lu[1]; [1]Univ. of Wisconsin-Madison, Madison, WI, [2]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [3]Univ. of Oslo, Oslo, Norway, [4]Katholieke Univ.it te Leuven, Leuven, Belgium, [5]Univ. of Wisconsin–Madison, Madison, WI, [6]VU Amsterdam, Amsterdam, Noord Holland, Netherlands, [7]Univ. of Southern California, Los Angeles, CA, [8]Univ. of Southern California, Mansfield, MA, [9]Univ. of Columbia, New York, NY, [10]Univ. of Wisconsin, Madison, WI

Abstract:

Marriage partner choice is not random in human societies. Individuals often opt for partners similar to themselves, a phenomenon known as assortative mating (AM). Previous studies estimating the genomic footprint of AM primarily focused on providing evidence for AM but revealed little insights into mechanisms behind AM. It is also known that ignoring AM could lead to substantial biases in genome-wide association studies (GWAS) and post-GWAS analysis such as heritability and genetic correlation estimation, but no existing approach has been proposed to remedy these issues, partly due to our incomplete understanding of the social processes behind AM. In this study, we couple econometric matching models for partner choice with molecular genetic and phenotypic data in four global biobanks to quantitatively investigate the genetic basis of partner choice. Adopting an econometric framework assuming AM being driven by a latent index quantifying partner choice preference, we illustrate that genetic associations with this index can be empirically identified through GWAS of spousal traits. Based on this key insight, we assembled 206,617 individuals with both own and spousal genomic data in four large biobanks: UK Biobank (UKB), the Norwegian Mother and Child Cohort Study (MoBa), the Health and Retirement Study (HRS), and the Lifelines cohort study. We found many significant and robust genetic correlations with partner choice index, with educational attainment showing the highest genetic correlation (cor = 0.88 and p = 4.3e-460). We also found that while specific traits can show different levels of assortment, there is a broadly consistent genetic component of the partner choice index across countries, environments, and sexes. Leveraging data from UKB and All of Us, we show that polygenic score (PGS) of partner choice index is highly significantly correlated between odd and even chromosomes (p = 4.0e-14 and 1.1e-24), demonstrating strong evidence for index-based AM in both UK and US. We also introduce a statistical framework to adjust for AM-driven biases in genetic correlation estimation. Conditioning on AM, genetic correlation of educational attainment and cognition remains substantial, while the highly significant population-based genetic correlation of height and education (cor = 0.14, p = 8e-40) reduces to null (cor = -0.05, p = 0.21). Taken together, through several key methodological innovations, this study performs the first GWAS on human partner choice, uncovers phenotypic characteristics and social processes underlying partner choice, and provides solutions to reduce the AM-driven biases in complex trait human genetic studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4052W: Genetic studies of liver fat and volume using whole-body MRI and extensive phenotyping provide metabolic insights

Authors:

G. Carrasquilla[1,2], S. Ahmad[3,4], T. Langner[5,6], U. Menzel[3], N. Ahmad[5], S. Sayols-Baixeras[3,7,8], K. Dekkers[3], B. Kennedy[3], F. Malmberg[5,9], U. Hammar[3], M. Romero-Lado[1], J. Censin[10,11], D. Nguyen[3], A. Mora[5], T. Kilpeläinen[1,2], L. Lind[12], J. Eriksson[13], R. Strand[5,9], J. Kullberg[5,6], H. Ahlström[5,6], T. Fall[3]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]Novo Nordisk Fndn. Ctr. for Genomic Mechanisms of Disease, Broad Inst. of MIT and Harvard, Boston Cambridge, MA, [3]Molecular Epidemiology, Dept. of Med. Sci., Uppsala Univ., Uppsala, Sweden, [4]Preventive Med. Div., Harvard Med. Sch., Brigham and Women's Hosp., Boston Cambridge, MA, [5]Dept. of Surgical Sci., Uppsala Univ., Uppsala, Sweden, [6]Antaros Med. AB, Mölndal, Sweden, [7]Med. epidemiology, Dept. of Surgical Sci., Uppsala Univ., Uppsala, Sweden, [8]CIBER Cardiovascular Diseases (CIBERCV), Inst. de Salud Carlos III, Madrid, Spain, [9]Image Analysis, Dept. of Information Technology, Uppsala Univ., Uppsala, Sweden, [10]Big Data Inst. at the Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom, [11]Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [12]Clinical Epidemiology, Dept. of Med. Sci., Uppsala Univ., Uppsala, Sweden, [13]Clinical Diabetology and Metabolism, Dept. of Med. Sci., Uppsala Univ., Uppsala, Sweden

Abstract:

Background and Aims: Metabolic dysfunction-associated steatotic liver disease affects about a quarter of the global adult population. However, the genetic factors and their impact on body composition and metabolic health remain unclear. We aimed to investigate the associations between liver trait-associated genetic variants and detailed fat content and tissue volume across various body compartments and organs. Additionally, we explored how these genetic variants relate to cardiometabolic biomarkers.
Methods: Genome-wide association analyses were performed for liver fat and liver volume assessed with magnetic resonance imaging in up to 27,243 unrelated participants of European ancestry in the UK Biobank. Each independent variant was assessed for body-wide associations with fat content and volume of >2 million 3D “Imiomics” image elements in 22,261 individuals. Additionally, these variants were assessed for association with a large set of biomarkers in 310,239 individuals.
Results: We confirmed four liver fat content and nine liver volume genetic variants and identified one novel liver volume loci *ADH4*. The Imiomics maps showed that all four liver fat variants increased liver fat fraction, with *TM6SF2* positively associated with visceral adipose (VAT) tissue fat fraction in women and *PNPLA3* positively with VAT in men but negatively in women. Liver fat variants showed diverse tissue volume associations. Most liver volume loci were positively associated with liver volume, except *ARID1A*, *TNKS2*, and *LPAR2* in women. *PPP1R3B* decreased liver fat fraction in women, while *LPAR2* increased it in both sexes. *PDIA3* was positively associated with VAT and lower body skeletal muscle volumes in both

sexes but negatively with lower body subcutaneous adipose tissue volumes. *CENPW* and *GCKR* showed, in general, whole-body pleiotropic effects. Liver fat variants were linked to higher levels of biomarkers for inflammation, liver and renal disease, and lower lipid levels.

Conclusions: Liver fat-associated variants are specifically linked to fat in the liver but not in other organs, with two exceptions showing an association with VAT. Additionally, these liver fat-associated variants are linked to markers of inflammation and liver injury, and negatively associated with lipid biomarkers. Meanwhile, liver volume-associated loci exhibit broader effects on organs and tissues, including sex-specific effects on organ volume and fat, along with diverse biomarker associations. These findings enhance our understanding of the sex-specific and heterogeneous effects of liver variants on organ and body composition, with diverse metabolic implications.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4053W: Genetic Variants Associated with Lactase Persistence in the United Arab Emirates: Insights from the Emirati Genome Project

Authors:

S. AlMarzooqi[1], W. Mohamed[1], I. Iqbal[1], A. M. Yousif[1], N. Marzouka[2], A. A. AlSuwaidi[2], A. AlAamri[2], R. M. Olbrich[2], M. Mousa[2], H. A. AlNaqbi[2], J. Quilez[3], T. Magalhaes[3], M. S. AlAmeri[1], S. M. Ibrahim[2], A. I. AlMannaei[1], H. Alsafar[2]; [1]Dept. of Hlth.-Abu Dhabi, Abu Dhabi, United Arab Emirates, [2]Khalifa Univ. for Sci. and Technology, Abu Dhabi, United Arab Emirates, [3]G42 Hlth.care, Abu Dhabi, United Arab Emirates

Abstract:

Lactase persistence (LP), the ability to digest lactose into adulthood, varies significantly across populations due to genetic differences in the LCT gene and its regulatory regions. While LP is common in Northern Europeans, it is less prevalent in many Middle Eastern populations and almost absent in Asia. To date, no study has investigated the presence and frequency of LP genetic variants in the United Arab Emirates (UAE). This study aims to address the gap in LP research within the UAE, a region with a unique genetic composition shaped by its geographical crossroads. Utilizing data from the Emirati Genome Project, we investigate the frequency and distribution of LP-associated genetic variants in over 43,000 Emiratis, providing a comprehensive overview of the genetic underpinnings of lactose intolerance (LI) in this diverse population. Whole genome sequencing (WGS) data from the Emirati Genome Project were analyzed to identify genetic variants in the LCT gene and its enhancer regions. Variants were

annotated using Annovar, and linkage disequilibrium (LD) analysis was performed using Plink. Allele frequencies (AF) of LP-associated variants were compared with global populations. Population stratification was analyzed to determine the distribution of LP variants across different genetic subgroups within the UAE. We identified several known LP-associated variants within the Emirati population, including LCT-13907 C>G, LCT-13910 C>T, LCT-13913 T>C, LCT-13915 T>G, LCT-14009 T>G, LCT-14010 G>C, and LCT-22018 G>A. The most common variant was LCT-13915 T>G, followed by LCT-22018 G>A and LCT-13910 C>T. Notably, variants such as LCT-13908 G>A, previously reported in other populations, were absent in our UAE cohort, while LCT-13914 G>A was found in a single individual. The other variants were found at very low frequencies. Overall, 46% of the individuals had at least one LP variant. Our genetic subpopulation analysis showed an uneven distribution of variants and haplotypes, with some subpopulations having a higher cumulative frequency of all LP SNPs than others. This large-scale study of the Emirati population is the first to provide new insights into the genetic basis of LP in the UAE, emphasizing the need for region-specific research to address public health and nutritional challenges. Our findings highlight the genetic diversity underlying LP in the UAE, with implications for dietary recommendations and healthcare strategies tailored to this population.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4054W: Genetics of cardiovascular outcomes in individuals with chronic kidney disease: the Chronic Renal Insufficiency Cohort (CRIC) study

Authors:

J. Wen[1], B. Lin[1], Q. Sun[1], M. Jiang[1], G. Li[2], R. Li[3], A. S. Go[4], T. Miller-Fleming[5], M. M. Shuey[5], D. L. Cohen[6], P. S. Rao[7], M. Rahman[8], N. Cox[5], J. P. Lash[9], W. Guan[1], H. Zhou[1], Y. Li[1], N. Franceschini[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Dept. of Statistics and Operations Res., Chapel Hill, NC, [3]Dept. of Biostatistics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Div. of Res., Kaiser Permanente Northern California, Oakland, CA, [5]Vanderbilt Univ Med Ctr., Nashville, TN, [6]Dept. of Med., Univ. of Pennsylvania, Philadelphia, PA, [7]Div. of Nephrology, Univ. of Michigan, Ann Arbor, MI, [8]Div. of Nephrology and Hypertension, Univ. Hosp. Cleveland Med. Ctr., Louis Stokes Cleveland VA Med. Ctr., Case Western Reserve Univ., Cleveland, OH, [9]Dept. of Med., Univ. of Illinois, Chicago, IL

Abstract:

Individuals with chronic kidney disease (CKD) have accelerated atherosclerosis and are at higher risk of cardiovascular disease (CVD). Despite the observable elevated risk in patients with CKD

for CVD there remains a substantial gap in our understanding of the shared genetic liability for these conditions. In this study, we performed genome-wide association studies (GWAS) analyses of coronary heart disease (CHD) and all-cause stroke using participants with CKD in the Chronic Renal Insufficiency Cohort (CRIC). We used mixed-effect logistic regression models, adjusting for age, sex, site of recruitment, estimated glomerular filtration rate (eGFR), and genotype principal components. We also assessed genetic correlation between our CRIC results and published GWAS of cardiovascular disease that assess differences in genetic susceptibility. Results showed that, among the 3,588 CRIC participants, 1,203 (33.53%) had CHD and 535 (14.91%) had all-cause stroke. Our GWAS analyses identified six SNPs across three loci (*LINC02744*, *AZIN1-AS1*, *ATP6V0A4*) associated with all-cause stroke, and two intronic SNPs at the *PPARG* locus associated with CHD. None of the published SNPs were associated with CHD or stroke in CRIC. We observed significant genetic correlations between published GWAS and CRIC results (genetic correlations: 0.39 to 0.51, *p*-values: 0.0009 to 0.0068) for CHD but not for stroke. Our findings suggest that there are some differences in the genetic architecture of CHD and stroke among individuals with CKD and those from the general population. Whether these differences exist due to non-traditional risk factors in people with CKD is not fully understood and future work seeks to further disentangle this differential risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4055W: Genome-wide machine learning analysis on Alzheimer's disease

Authors:

M. Bracher-Smith[1], F. Melograna[2], B. Ulm[3], C. Bellenguez[4], B. Grenier-Boley[4], R. Frikke-Schmidt[5], N. Amin[3], G. Roshchupkin[6], J-C. Lambert[4], K. Van Steen[7], C. Van Duijn[3], V. Escott-Price[1], The European Alzheimer & Dementia Biobank (EADB) Consortium; [1]Cardiff Univ., Cardiff, United Kingdom, [2]KU Leuven, Leuven, Belgium, [3]Oxford Univ., Oxford, United Kingdom, [4]Inst. Pasteur de Lille, Lille, France, [5]Copenhagen Univ. Hosp. - RigsHosp.et, Copenhagen, Denmark, [6]Erasmus MC, Rotterdam, Netherlands, [7]Univ. of Liège / KULeuven, Liège, Belgium

Abstract:

Traditional statistical approaches have greatly advanced our understanding of risk genes for Alzheimer's disease (AD); however, they typically rely on linear additive models. Here, we use gradient boosting machines (GBMs), a flexible machine learning (ML) approach, to identify novel genes and assess prediction accuracy in the largest genome-wide ML study of AD to-date, trained on 14,006 AD cases and 15,174 controls derived from the EADB consortium. GBMs

confirmed previously identified genome-wide significant variants and using Shapley values identified the coding variants in *APOE* and *ABCA7* as the most important drivers in their region. Further, GBMs uncovered novel loci including *HS3ST2* and *DTNB* and achieved a predictive performance (AUC=0.692) that was comparable to that of GWAS based polygenic risk scores (AUC=0.689), but with a significantly reduced set of variants. Our findings demonstrate that ML methods, applied to a large sample size, have reached a point where they can compete with classical GWAS in detecting novel variants in AD pathogenesis and enhancing the predictive modelling of AD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4056W: Genome-wide Search of Exonic Variants Affecting Translational Efficiency

Authors:

A. Vanpoperinghe[1], E. Wheeler[1], L. Marchand[2], C. Polychronakos[2]; [1]McGill Univ., Montreal, QC, Canada, [2]McGill Univ. Hlth.Ctr., Montreal, QC, Canada

Abstract:

Background: The elucidation of the functional consequences of genetic variation in humans is essential in understanding phenotypic diversity and personalizing medicine. Transcriptional effects, as expression quantitative trait loci (eQTLs), have been extensively studied, but gene expression can also be affected by genetic variation at the translational level, on which only fragmentary published information exists for individual genes. We addressed this question at a transcriptome-wide level. **Experimental approach:** We used ribosomal association to specific mRNAs from 200 individuals' lymphoblastoid cell lines as a proxy for translational efficiency, to detect cis-acting translational quantitative trait loci (tQTLs) genome-wide. Our phenotype was the ratio of heavy ribosomal RNA over total RNA, obtained by gradient ultracentrifugation of cell lysates after treatment with cycloheximide.
Results: By QTLtools, we identify 120 protein-coding genes whose translation is modulated by genetic variation. We observe an 8-fold enrichment of association with exonic polymorphisms (p < 1e-16), and a 6-fold enrichment in 5' untranslated regions (p < 6.94e-3), as expected for direct effects on the mature mRNA's translation. We have validated 5 of these associations at the protein level using large-scale antibody-based proteomics, despite their methodological limitations in genome-wide quantification and our limited sample size. **Conclusion:** Our results validate an important addition to eQTLs in mapping genetic associations to function, the indispensable first step towards targeted therapeutics. Being constrained to an exonic location,

tQTLs can be mapped much more precisely than eQTLs and can explain many pQTLs. **Future Directions:** With the help of *in-silico* predictors of mRNA features, we aim to uncover potential mechanisms through which tQTLs impact translational efficiency, starting with alterations in secondary structures and miRNA binding sites. Additionally, we will assess the likelihood of a shared functional effect between our tQTLs and GWAS variants by Regulatory Trait Concordance score analysis in hopes of increasing the explanatory power of GWAS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4057W: Haplotype analysis reveals pleiotropic disease associations in the HLA region

Authors:

C. Smith[1], S. Strausz[2], F. Finngen[2], J. Spence[1], H. Ollila[2], J. Pritchard[1]; [1]Stanford Univ. Sch. of Med., Stanford, CA, [2]Inst. for Molecular Med. Finland, Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland

Abstract:

The human leukocyte antigen (HLA) region plays an important role in human health and disease through its involvement in immune cell recognition and maturation. While genetic variation in the HLA region is associated with many diseases, the pleiotropic patterns of these associations have not been systematically investigated. Many studies investigating disease associations in the region focus on associations of protein-coding variation of classical HLA genes, a small number of traits at a time, or treat the region as another locus in the genome. Here, we developed a haplotype regression approach that enables evaluation of the full genetic complexity of region and exploration of pleiotropic disease associations at the phenome-wide scale for 412,181 Finnish individuals and 2,405 traits. We performed disease association testing with single nucleotide polymorphisms (SNP), HLA alleles, and haplotypes to cover disease associations throughout the entire HLA region, identifying a total of 7,649 significant trait associations and 647 HLA-associated traits. The SNP associations in the HLA region were enriched compared to the rest of the genome for a broad variety of disease groups, including a 393-fold enrichment for infectious traits, 53-fold enrichment for rheumatic traits, and 18-fold enrichment for neurologic traits. Haplotype analysis uncovered 469 significant (|Z|>4) associations, indicating a complex pleiotropic landscape within the region that affects a wide range of diseases. This study highlights the extensive impact of HLA variation on disease risk, underscoring the importance of both classical and non-classical genes, as well as SNP and haplotype variation within the region.

Our findings emphasize the role of HLA in influencing protection and risk for a broad spectrum of diseases, offering insights into its pleiotropic effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4058W: Haplotype association using LOCATER improves trait association power at loci with allelic heterogeneity

Authors:

X. Wang[1,2], R. Christ[1,2], E. Young[3,4], C. Kang[4], I. Das[4,5], E. A. Belter, Jr.[4,6], L. Aslett[7], D. Steinsaltz[8], N. Stitziel[3,9], I. Hall[1,2]; [1]Dept. of Genetics, Yale Univ., New Haven, CT, [2]Ctr. for Genomic Hlth., Yale Univ., New Haven, CT, [3]Ctr. for Cardiovascular Res., Div. of Cardiology, Dept. of Med., Washington Univ. Sch. of Med., St. Louis, MO, [4]McDonnell Genome Inst., Washington Univ. Sch. of Med., St. Louis, MO, [5]Current affiliation: Inst. for Informatics, Data Sci. & Biostatistics, Washington Univ. in St. Louis, St. Louis, MO, [6]Current affiliation: Collaborative and Integrative Genomics Lab at the McDonnell Genome Inst. at Washington Univ. Sch. of Med., St. Louis, MO, [7]Durham Univ., Mathematical Sci., Durham, United Kingdom, [8]Oxford Univ., Statistics, Oxford, United Kingdom, [9]Dept. of Genetics, Washington Univ. Sch. of Med., St. Louis, MO

Abstract:

Genome-wide association studies have left many causal loci undiscovered. Combining common and rare variant signals at loci with allelic heterogeneity, across both coding and non-coding regions, remains a key methodological challenge. In parallel work, we developed a haplotype association method (LOCATER) that tests inferred local genealogies for association with traits and shown via simulations that this approach improves statistical power at loci with multiple causal variants. Here we present the first application of LOCATER to a real-world dataset. This study includes 6,795 Finnish individuals (METSIM) with 101 cardiometabolic quantitative traits, and 18.9 million autosomal variants derived from whole genome sequencing data. We found a total of 351 significant trait associations using both LOCATER and standard single marker test (SMT). Most associations were found by both methods at a similar level of significance; however, LOCATER discovered 7 significant trait associations that were not found by SMT, and LOCATER had a more significant p-value than SMT at 23 additional trait associations. We investigated the source of LOCATER's improved performance at these loci and found allelic heterogeneity that yielded boosted signals. These include two independent trait associations at LIPG that were not found by SMT, both of which appear to be caused by multiple causal variants, as well as a novel trait association that appears to be caused by the combined effect of five rare haplotypes. In each case, the power boost provided by LOCATER was due to its ability

to combine association signals from distinct haplotypes found in disparate clades of the local ancestral tree. LOCATER also recovered known allelic heterogeneity at the APOE/C1/C4/C2 gene cluster. We found that confounders have a more pronounced effect on genealogy-based methods relative to SMT, and propose a more general method for genomic control to successfully eliminate the impact of these confounders in large-scale population based studies. This study demonstrates that haplotype-based association methods such as LOCATER are highly effective when multiple causal variants are present, and suggests that application of these methods to larger and more diverse cohorts will be especially fruitful for novel gene discovery.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4059W: High polygenic risk assignment instability and mitigation

Authors:

A. Misra[1,2], S. Urbut[2,1,3], B. Truong[1,2,4], J. Smoller[2,3], A. Patel[2,1,3], P. Natarajan[2,1,3]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Background: Polygenic risk scores (PRS) continue to improve with advances in statistical methods and increases in size of genome-wide association studies. However, numerous PRS exist for several complex traits. It is unknown how the classification of high polygenic risk changes across PRS for given traits.
Methods: We determined the associations of published PRS for complex traits from the Polygenic Score Catalog in the All of Us cohort (AoU), examining 57 scores for coronary artery disease (CAD), 18 scores for type 2 diabetes mellitus (T2DM), 129 scores for major depressive disorder (MDD). PRSMix is a tool that incorporates information across several PRS for a given trait to improve prediction accuracy for a target population. We generated sequential add-one-in (AOI) PRSMix_AOI$_n$ scores, wherein we combined PRS$_n$ with all prior published scores for each PRS$_n$. PRSmix$_{all}$ was defined as the PRSMix_AOI$_n$ for the most recently published PRS. Phenotypes were ascertained through hospitalization ICD9 and ICD10 diagnosis and procedural codes. Congruence of risk estimates was assessed with the Jaccard similarity index. Logistic regression was performed, adjusting for age, sex, and the first 10 principal components of genetic ancestry.
Results: Among the 129,370 AoU participants of European ancestry with srWGS data, 60% are female, mean enrollment age is 56 ± 17 years. While each PRS$_n$ demonstrated consistent strength of associations, classification of individuals in the top 10% of each PRS distribution

varied widely. Chronologically adding score PRS_n in order of publication into $PRSMix_AOI_n$ led to stronger associations with the outcome and more consistent high risk classification compared to the PRS_n itself. The OR associated with top 10% risk classification ranged from 1.6 - 2.8 for PRS_n published over the last two years vs. 2.8 - 3.0 for $PRSMix_AOI_n$. The median Jaccard index for top 10% classification by CAD PRS_n was 0.17 [0.13-0.22], indicating poor correlation between scores. Using $PRSMix_AOI_n$ led to increasing stability of high PRS assignment over time, with Jaccard index improvement to median 0.76 [0.74-0.95] over the past two years. Participants classified in top 10% risk by the $PRSMix_{all}$ for CAD fall into a lower and wider range of percentiles for PRS_n published over the past two years (median % [IQR] of 76 [52-91] % vs. for paired $PRSMix_AOI_n$ scores (95 [90-98] %). Similar results were observed for T2DM and MDD. **Conclusion**: In a comprehensive assessment of published PRS for 3 complex diseases, high PRS classification was highly inconsistent. Use of PRSMix_AOI to sequentially incorporate new PRS enables harmonization toward more stable classification of high risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4060W: iCLOUD: Utilizing family history for sex-difference analysis in genetic association study of Alzheimer's disease

Authors:

X. Wu[1], B. Jiang[1], L. Carrel[2], D. J. Liu[1]; [1]Penn State Univ., Hershey, PA, [2]Penn State Univ Col Med, Hershey, PA

Abstract:

Alzheimer's disease (AD) is a late-onset neurodegenerative disorder with notable sex differences: females often have earlier onset, faster progression, and more severe symptoms than males. Genome-wide association studies (GWAS) have identified over 100 loci for non-familial AD, suggesting various disease mechanisms and potential therapies. However, all largest AD GWAS primarily focus on main effects, leaving sex-specific genetic differences underexplored. The GWAS-by-proxy method (GWAX) increases sample size by labeling controls as proxy cases if they have a relative with AD. GWAX is popular for boosting sample sizes and is easy to use with existing GWAS tools. However, it struggles with analyzing gene x sex interactions because the sex of proxy cases and their affected relatives may differ. Additionally, GWAX is prone to participation bias (biobank individuals may not complete family health surveys) and survival bias (relatives may not live long enough to develop AD despite carrying the risk allele). To overcome the limitation of existing approaches, we propose iCLOUD (**i**n**C**omplete **L**ikelihood **O**f **U**ngenotyped **D**isease status), a likelihood-based method, for testing

gene-by-sex effects on AD. iCLOUD directly models parents' phenotypes utilizing the conditional genotype distribution in the likelihood and accounting for the missing covariates in ungenotyped AD relatives. In order to account for survival bias and participation bias, we also developed iCLOUD-mixture, an extension of iCloud which employs mixture modeling to account for the possibility of participants not filling out family health surveys. We perform extensive simulation to compare iCOUD with existing GWAX methods. In the scenarios we examined, iCLOUD on average has 1081% and 15% higher power in testing the gene-by-sex interaction effect than GWAX methods. It also yields considerably smaller estimation biases (e.g. with true gene x sex interaction effect 0.1, GWAS bias 0.0005, iCLOUD bias -0.0093, GWAX biases -0.0979 and -0.0515) and mean squared errors (iCLOUD 0.0017, GWAS 0.0241, GWAX 0.0101 and 0.0032). iCLOUD models also exhibited comparable powers and smaller biases in estimating the main genetic effect than GWAX. Application of the proposed method to UK Biobank and All of Us studies is ongoing. In summary, we present a new approach iCLOUD to model family history in genetic association studies. It not only improves the power of detecting gene x sex interaction effects, but also yield unbiased genetic effect estimates, and hence improves all downstream analysis, including Mendelian randomization and genetic correlation analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4061W: Identification of *trans* regulatory networks from multi-condition multi-omics quantitative trait loci with applications in Alzheimer's disease

Authors:

X. Cao[1,2], Y. Wang[1], T. Fabiha[2], Y. Bai[1], Z. Qi[1], E. Tsai[3], H-J. Westra[4,5], L. Franke[4,5], The Alzheimer's Disease Functional Genomics Consortium, G. Wang[1], K. Dey[2]; [1]Columbia Univ., New York City, NY, [2]Mem. Sloan Kettering Cancer Ctr., New York City, NY, [3]Biogen, Cambridge, MA, [4]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [5]Oncode Inst., Groningen, Netherlands

Abstract:

Most of the heritability of complex diseases and gene expression is driven by *trans* effects downstream of GWAS genes. However, decoding the *trans*-regulatory architecture is challenging due to the limited sample size and smaller causal effect. The integration of multiple molecular QTL (xQTL) studies can aggregate *trans* effect gene programs downstream of multiple *trans*-e-SNPs and gene perturbations across various tissues and cell types. Here, we propose a new approach, transCCA, that uses single- and multi-view canonical correlation analysis framework to identify gene programs shared with multiple *trans*-acting QTLs. We further propose a strategy

(cTrans) to identify an optimal combination of these *trans* effects for a complex disease of interest. We conducted extensive simulation studies to benchmark the performance of transCCA; single-context transCCA exhibited reduced false discoveries compared to the recently proposed ARCHIE method while maintaining similar accuracy. Multi-context transCCA further improved the performance over the single-context version in varying signal-to-noise ratio settings. We applied transCCA to three large-scale eQTL studies: MetaBrian, eQTLGen, and a new single-nucleus RNA sequencing dataset from 737 ROSMAP individuals across 13,690 *trans*-e-genes. Single-context transCCA gene programs from eQTLGen showed 3.3%-4.6% higher heritability enrichment for blood cell traits compared to ARCHIE and were found to be 1.3x enriched in E3ligase perturb-seq gene programs. Combining the single- and multi-context transCCA gene programs from 3 pseudo-bulk cell types (microglia, excitatory and inhibitory neurons) revealed up to 2.1x higher heritability signals for brain-related traits. Targeted transCCA applied to AD fine-mapped variants reveals gene programs that were 1.4x-2.2x enriched ($p<0.001$) in differentially expressed genes in microglia upon treatment with IFNβ, including genes *NUB1*, *PCMT1*, *HAVCR2*, et al. One of the transCCA gene programs, notably driven by *MAP2K6*, *RTN3*, *PSMC6* genes, demonstrated significant enrichment in the neurodegeneration pathway ($p<0.001$). Meanwhile, cTrans aggregation of *trans* effects for blood and brain-related complex diseases show high enrichment in phenotypically matched Mendelian disorder genes and drug targets; for example, cTrans aggregation of MetaBrain *trans* effects showed 4.1x enrichment in drug targets for brain disorders. Overall, we provide new computational approaches to learn disease-critical *trans*-regulatory gene programs from *trans*-xQTL data spanning multiple cell types and assays.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4062W: Identifying candidate risk and protective rare variant effects on Alzheimer's Disease using multi-omic variant prioritization

Authors:

E. Robb[1], T. Jensen[1], B. Ni[2], D. Nachun[1], G. Wang[3], S. Montgomery[1]; [1]Stanford Univ., Stanford, CA, [2]Johns Hopkins Univ., Baltimore, MD, [3]Columbia Univ., New York, NY

Abstract:

An estimated 30% of Alzheimer's disease (AD) heritability is driven by common variants, but rare variants can have high penetrance for certain AD-associated genes. While a small number of highly penetrant coding variants have been identified, little is known about non-coding rare variants that can impact AD risk. Detecting impactful rare non-coding variants remains a

significant challenge despite their abundance, largely due to limitations with statistical tools such as GWAS and QTL mapping in the rare variant setting. To enrich for impactful rare variants, prior work has shown that trait-associated rare variants can be prioritized using multi-omic outlier signals.
We hypothesized that rare variants associated with increased or decreased impact on Alzheimer's disease can be identified using multi-omic outlier signals in AD-associated genes.
We used the Religious Orders Study and Memory and Aging Project (ROSMAP) cohort, which has profiled subjects with whole genome sequencing, RNA-seq in multiple brain regions and proteomics in the DLPFC. Several AD phenotypes are also available for ROSMAP subjects including clinical diagnosis and pathological scores.
We prioritized rare variants in ROSMAP disease and control samples based on genomic annotations and multi-omic outliers with Watershed, including outliers for gene expression, splicing, and protein expression. We selected rare variants near known AD risk genes, and identified individuals with outlier AD phenotypes to prioritize protective or risk variants. We define outlier phenotype carriers as individuals whose clinical diagnosis conflicts with pathology or genetic risk factors.
Among functional rare variants prioritized by our model in ROSMAP, we found that individuals with outlier AD phenotypes were enriched for prioritized rare variants near AD-associated genes identified by NIAGADS. Our model identified 35,137 candidate functional rare variants near 6234 genes, and within the AD gene set, we prioritized 121 rare variants on 26 genes. We found a 1.4x increased risk of carrying prioritized rare variants for individuals with outlier pathology, and 3.6x increased risk within the AD gene set. We found a 2.2x increased risk of prioritized rare variants in the AD gene set for individuals with outlier genetic risk. For one APOE-e4 carrier without AD, we found 8 prioritized regulatory rare variants on APP associated with an underexpression outlier, suggesting a possible protective effect.
Our results suggest that multi-omic rare variant prioritization can help discover candidate rare variants with impact on AD risk, and that these functional rare variants are associated with phenotypic outliers in AD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4063W: Identifying genetic association with secondary phenotypes in a case-control study using tilted knockoffs

Authors:

Q. Zhao[1], S. Service[2], C. Bearden[3], C. Lopez-Jaramillo[4], N. Freimer[2], C. Sabatti[5]; [1]Univ. of Massachusetts, Amherst, Amherst, MA, [2]Univ. of California Los Angeles, Los Angeles, CA, [3]Univ.

of California, Los Angeles, Los Angeles, CA, [4]Univ. de Antioquia, Medellin, Colombia, [5]Stanford Univ, Stanford, CA

Abstract:

Leveraging data on secondary phenotypes can elucidate biological pathways underlying diseases; such traits may be more precisely and objectively assessed than disease diagnoses. Genetic association analyses of secondary phenotypes are often conducted in case-control samples, because this design predominates in studies aimed at elucidating the genetic contribution to diseases. This sampling strategy can lead to spurious association between genetic variants and a secondary phenotype when both affect the case-control status, resulting in a biased sample for the analysis of that phenotype. While current approaches can test association under such selective sampling, they do not apply when the number of variables is large.

We developed a tilted knockoff method that extends the model-X knockoff to selected samples. The model-X knockoff can be used with any model to test conditional association while controlling the false discovery rate when the observations are a random sample from a population of interest (Candès et al. 2018 J.R.Stat). Like the knockoff method, the tilted knockoff constructs synthetic knockoff variables to estimate the proportion of false discoveries. To account for distribution shifts due to selection, the tilted knockoff constructs knockoffs assuming the data are sampled from a tilted distribution. In simulated studies, the tilted knockoff method controls the proportion of false discoveries when the selection probability is known or estimated.

We applied the tilted knockoff to identify genetic variants associated to phenotypes associated with severe mental illnesses from a data set with 5119 cases and 1158 controls from the Paisa region in Colombia. These phenotypes describe neurocognitive functions, including attention and processing speed, measured by the Penn Computerized Neurocognitive Battery. Consistent with the simulation, the tilted knockoff produces fewer selections compared to the model-X knockoff. It both validates results of the model-X knockoff, while also generating additional results not found using that method. In summary, our findings from simulations show the higher validity of tilted knockoff selections when the samples are selected, and we suggest using the tilted knockoff approach to validate findings from the knockoff method and prioritize associations for future investigation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4064W: Identifying loci with pleiotropic effects on orofacial clefts and facial morphology using conditional false discovery rate analyses

Authors:

N. Herrick[1], S. Goovaerts[2,3], M. K. Lee[1], M. L. Marazita[1,4,5], P. Claes[2,3,6,7], S. M. Weinberg[1,4,8], J. R. Shaffer[1,4]; [1]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Univ. of Pittsburgh, Pittsburgh, PA, [2]Dept. of Human Genetics, KU Leuven, Leuven, Belgium, [3]Med. Imaging Res. Ctr., Univ. Hosp. Leuven, Leuven, Belgium, [4]Dept. of Human Genetics, Univ. of Pittsburgh, Pittsburgh, PA, [5]Clinical and Translational Sci., Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA, [6]Dept. of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium, [7]Murdoch Children's Res. Inst., Melbourne, Victoria, Australia, [8]Dept. of Anthropology, Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Human facial shape is influenced by genetic factors driving variation among normal-range morphologies and dysmorphologies. Recent GWASs have identified hundreds of loci associated with normal-range facial variation, furthering our understanding of the genetics of craniofacial morphology and development. Multiple lines of evidence suggest a (partly) shared genetic architecture of normal-range facial variation and non-syndromic orofacial clefts (OFCs), the most common facial birth defect, which affects about 1 in 700 live births. To explore their shared genetic underpinnings and boost discovery of OFC susceptibility loci, we leveraged the power of summary statistics from two independent GWASs in European cohorts of OFCs (n = 3,969) and normal-range facial morphology (n = 4,680) using conditional false discovery rate (FDR) analyses that detect pleiotropic loci. Next, we conducted replication tests in a second facial morphology European cohort (n = 3,566). Specifically, non-syndromic cleft lip with or without cleft palate (CL/P) and quantitative, multidimensional facial traits representing the shape of the full face were used for conditional FDR analyses. In total, we detected 21 significant SNPs at six independent genomic loci associated with OFCs at 1% FDR. Five of these loci have been previously reported in OFC studies. Two of the five loci, 2p21 and 17q22, were replicated; the sixth locus was found only in the replication study and has not been previously implicated in OFCs. Although we demonstrate the capability of this approach to identify shared genetic architecture of OFCs and normal-range facial morphology, future research would benefit from testing other OFC subtypes in more diverse population cohorts, which could reveal additional potential OFC risk loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4065W: Identifying Rare Variants Associated with hypermobile Ehlers-Danlos Syndrome Using a Case-Only Cohort and Biobank Controls: Overcoming Ancestry and Technical Differences in Separately Sequenced Samples

Authors:

A. Seth[1,2], W. He[1,2], R. Handsaker[2,3], V. Janoušek[4], M. Galindo[5], D. Subramanian[6], C. A. Francomano[7,8], W. Gandy[8], HEDGE Consortium, C. Laukaitis[9,10,11], J. Hirschhorn[1,2,3]; [1]Boston Children's Hosp., Boston, MA, [2]Broad Inst., Cambridge, MA, [3]Harvard Med. Sch., Boston, MA, [4]Biodviser Ltd, Oldham, Manchester, United Kingdom, [5]Univ. of Arizona, Tucson, AZ, [6]Peter MacCallum Cancer Ctr., Melbourne, Australia, [7]Indiana Univ. Sch. of Med., Indianapolis, IN, [8]The Ehlers-Danlos Society, London, UK and New York, NY, [9]Carle Illinois Coll. of Med., Urbana, IL, [10]Carle Hlth., Urbana, IL, [11]Inst. for Genomic Biology, Urbana, IL

Abstract:

Background: Hypermobile Ehlers-Danlos Syndrome (hEDS) is the most prevalent of a group of rare, inherited connective tissue disorders. The underlying genetic architecture is unknown, but rare loss-of-function and missense variants with large effect are hypothesized to contribute to disease risk. The hEDS Genetic Evaluation (HEDGE) study, a case-only cohort, aims to identify genes with variants that contribute to hEDS pathogenesis. We selected controls from a separately sequenced cohort (AllofUs, AoU), requiring us to resolve ancestry and technical differences between cases and controls. **Study Design & Methods:** We analyzed whole genome sequence (WGS) data from HEDGE (973 unrelated participants: met 2017 diagnostic criteria for hEDS, WGS data passed standard quality control (QC)). We selected 5 ancestry-matched AoU controls per case: the nearest neighbors via Mahalanobis distance in AoU's principal component space. Systematic technical differences between the cohorts were identified and mitigated with an additional QC pipeline, using synonymous variant burden tests for calibration. This pipeline will be used to perform burden tests of nonsynonymous variants. **Results:** We could match all HEDGE cases to AoU controls regardless of ancestry, resolving ancestry differences between the cohorts. Our initial calibration test with synonymous variants had residual inflation from technical differences: 42 likely false positive genes with Bonferroni corrected p-value (p_{adj}) < 0.05 and Genomic Inflation Factor (λ_{GC}) = 1.14. After additional QC steps (harmonizing transcript annotation sets, excluding sites with 50+ alternate alleles, excluding segment duplication regions), only 8 genes have $p_{adj} < 0.05$ ($\lambda_{GC} = 0.95$). Variants flagged by gnomAD for discrepant frequencies or allele-specific quality filters are likely driving the few remaining false positives. We are exploring additional filters to exclude these problematic variants before proceeding to burden tests of nonsynonymous variants. **Conclusion:** We developed a QC pipeline to calibrate

rare variant burden tests in an understudied rare disorder, using a case-only WGS data set and separately sequenced AoU participants as controls. After addressing the few remaining batch effects, we will further annotate variants (e.g. LOFTEE for loss-of-function variants) and implement nonsynonymous variant burden tests. The strength of this study is its novel ancestry-based matching and the systematic measures that mitigate technical differences between separately sequenced cohorts, generating a single, multi-ancestry, well-matched case-control cohort that includes all cases, regardless of ancestry.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4066W: IGEM: Novel software provides up-to-date biological knowledge for analyzing exposure-exposure interactions

Authors:

M. Hall[1], J. Zhou[2], M. D. Ritchie[3], N. Palmiero[1], K. Passero[4], A. Garao Rico[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Stanford Univ., Palo Alto, CA, [3]Univ. OF PENNSYLVANIA, Philadelphia, PA, [4]The Pennsylvania State Univ., State College, PA

Abstract:

Sophisticated approaches are needed to model complex exposure-exposure interactions (ExE) at the exposome level and meaningfully connect them to health outcomes. Filtering methods alleviate multiple testing burden issues associated with interaction analysis in big data. Main effect filtering sets a statistical threshold for subsequent interaction analysis, while knowledge-based filtering provides a way to incorporate publicly available database knowledge of biological relationships (e.g., pathways, ontology, and biological interactions), thereby reducing the interaction models to only pairs of features that are significantly related to the outcome (main effect) or biologically related to each other (knowledge-based). However, current knowledge-based methods have yet to include exposome knowledge, thereby limiting their applicability to ExE. To address this gap, we developed the Integrative Genome-Exposome Method (IGEM), a novel software with: 1) the Genome-Exposome knowledge database (GE.db), a unified repository of knowledge of the biological relationships between genes and exposures and between exposures across numerous open-source databases, 2) Genome-Exposome-filter (GE.filter), which generates a list of filtered and biological-relevant candidate pairs for interaction analysis using knowledge from GE.db, and 3) statistical interaction analysis with main effect and knowledge-based filtering. We applied IGEM to evaluate the role of ExE in body mass index (BMI) using data from the National Health and Nutrition Examination Survey (NHANES; N discovery cohort=9,063, and N replication cohort=9,874). We replicated 26 and 4

ExE models using main effect and knowledge-based filtering, respectively. The top interactions involved trans-β carotene and cis-β carotene (Bonferroni-adjusted $p = 2.41\times10^{-28}$), α-carotene and trans-β carotene (Bonferroni-adjusted $p = 1.38\times10^{-14}$), and α-carotene and cis-β carotene (Bonferroni-adjusted p 3.21×10^{-13}). Several lipids and fatty acids were found to interact with folate (e.g., MFA 16:1: Bonferroni-adjusted $p = 1.23\times10^{-12}$; SFA 16:0: Bonferroni-adjusted $p = 9.98\times10^{-12}$; total fat: Bonferroni-adjusted $p = 9.13\times10^{-10}$; and total monounsaturated fatty acids: Bonferroni-adjusted $p = 1.65\times10^{-9}$; MFA 18:1: Bonferroni-adjusted $p = 3.90\times10^{-9}$; and PFA 20:4: Bonferroni $p = 6.41\times10^{-6}$). Both filtering methods facilitate interaction analysis, and IGEM offers an efficient and user-friendly tool to evaluate complex interactions using exposome data for a variety of health outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4067W: Impact of sex, ancestry, and local ancestry on Apolipoprotein E (APOE) risk for Alzheimer's disease (AD): a pooled case-control analysis of Alzheimer's Disease Genetics Consortium (ADGC)

Authors:

O. Garcia Rodriguez[1], L. Wang[2], J. Young[1], Alzheimer's Disease Genetics Consortium (ADGC), M. Pericak-Vance[1], E. Martin[1], B. Kunkle[1]; [1]Univ. of Miami, Miami Beach, FL, [2]Univ. of Miami, Miami, FL

Abstract:

Background/Significance: Previous studies have shown that non-Hispanic white (NHW) females (vs. males) carrying the APOE ε4 and ε2 allele differ in risk of developing Alzheimer's disease (AD). Moreover, few studies found that the association between APOE ε4 and AD risk may be modified by age and its local ancestry in admixture populations s, there is still lack of evidence on how sex could interact with these factors. Thus, we aimed to evaluate the association of sex, ancestry, *APOE*, and local ancestry with AD in multiple cohorts. **Materials/Methods:** A pooled case-control study was conducted from 50 independent datasets (n= 44,556; age≥ 55 years) in the Alzheimer's Disease Genetics Consortium. Logistic regression, linear, and Cox regression models were used to study the associations between disease endpoint and independent variables including age, sex, race, and *APOE* genotype. Gee models were run for the local ancestry (European -EU-; African -AF-; Amerindian -AI-; and EA), APOE haplotype analyses. **Results:** Regardless of ancestry, APOE ε4 or ε2 carries did not differ in the risk of

developing AD by sex. When comparing subjects with APOE ε3/ε4 or ε4/ε4 (vs. ε3/ε3), East-Asian (EA) females (*OR*=5.64; 95% CI: 4.46-7.13) and males (*OR*=5.41: 4.04-7.24) had the strongest *ORs*; whilst the weakest risk was for South-Asian (SA) men (*OR*=1.14: 0.61-2.11) among all races. Age-stratified analyses showed that NHW females (*OR=5*.57: 4.94-6.27 vs. males: *OR*=4.66: 4.06-5.35) had a higher risk of developing AD between ages 66-75; whereas African American (AA), EA and SA exhibited this pattern at 55-64. In Hispanics, the relationship is reversed, with males aged 55-65 years having a stronger association with AD risk than their counterparts. In NHW (55-64 years) and NHB (66-75 years) individuals, there was interaction between age and APOE for AD risk by sex (p_{int}<0.05). NHW males and females; NHB males; and EA female carriers of ε3/ε2 and ε2/ε2 were protected from AD vs. their counterpart carriers of ε3/ε3. Although races did not differ by sex and age in terms of AD risk, those NHW women who carried APOE ε3/ε2 and ε2/ε2 alleles presented a slightly stronger protective risk of developing AD than males in the same ethnical group for all age categories. Local ancestry analyses showed that AD risk (ε4 vs. ε3) did not differ by sex and AD risk local ancestry across races; yet we observed that in HL with EU local ancestry for APOE both women and men had an increased risk of developing AD compared with their sex counterparts with local ancestry of the AF descendent. **Conclusions:** Our findings suggest differences by sex, ancestry, and local ancestry for APOE in the risk for developing AD in carriers of *APOE ε4 and ε2*.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4068W: Inferring the evolutionary architecture of complex traits through the WF process

Authors:

M. Carr, X. Rui, N. Mancuso; Univ. of Southern California, Los Angeles, CA

Abstract:

GWAS and eQTL studies have identified genetic variants which, in aggregate, differ in functional and evolutionary character. GWAS hits tend to be under selective constraint, whereas eQTL hits tend to experience less constraint. These differences suggest that it is not solely genetic architecture which explains differences in eQTL data, but an *evolutionary* architecture.
To date, there have been a handful of models which incorporate selection-based architecture in an inferential statistical setting. Among these, the core idea traces its roots to a paper of Eyre-Walker. Eyre-Walker proposed a model linking complex traits and fitness together, mediated through selection. However, Eyre-Walker's model is not easily adapted to an inferential setting. This observation forms the basis for more recent developments incorporating selection through

ad-hoc models using minor allele frequency (i.e. MAF) into statistical models with the goal of learning parameters such as heritability and identifying causal variants. More recently, Simons et al. have proposed a model based on an assumption of *stabilizing* selection. While this model links allelic effects to selection through a formal model, it requires some additional input in the form of a distribution of selection coefficients determined through simulations, which may be challenging to obtain in practice.
To address these limitations, we propose an efficient sparse linear model, which links allele frequencies with stabilizing selection under the Wright-Fisher (i.e. WF) process. To do this, we build upon the work of Spence et al. scaling the discrete-time WF process to efficiently compute allele frequencies, conditional on the strength of stabilizing selection. To perform inference, we developed a fully differentiable WF process, coupled with a sum-of-single effects (i.e. SuSiE) model. We implemented these concepts together using JAX, which enables efficient computation across CPUs, GPUs, or TPUs.
In summary, our work builds upon, and connects, two active areas of research in statistical and population genetics. By feeding in ideas from population genetics to a model firmly grounded in statistical genetics, we expect to gain more insights into the evolutionary architecture of complex diseases and eQTL variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4069W: Inferringcell-type-specific methylation quantitative trait lociwith a hierarchical Bayesian interaction model: a simulation-based benchmark study

Authors:

Y. Cheng[1], C. Biao[1], H. Li[1], X. Zhang[1], K. Xu[2], H. Zhao[3]; [1]Yale Univ., New Haven, CT, [2]Yale Sch. Med., New Haven, CT, [3]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract:

Methylation Quantitative Trait Loci (meQTLs) are chromosomal regions that harbor genetic variants affecting DNA methylation levels. The identification of meQTLs can be accomplished through quantifying the effects of genetic variants, e.g., single nucleotide polymorphisms (SNPs), on DNA methylation levels, and these inferred meQTLs can shed light on the interplay between the genome and methylome. However, most meQTL studies to date utilize bulk methylation datasets composed of different cell types that may have distinct methylation patterns in each cell type. Current technological challenges hinder the comprehensive collection of large-scale, cell-type-specific (CTS) methylation data, which limits our understanding of CTS

methylation regulation. To address this challenge, we propose a hierarchical Bayesian interaction model (HBI) to infer CTS meQTLs from bulk methylation data. Our HBI method integrates bulk methylations data from a large number of samples and CTS methylation data from a small number of samples to estimate CTS meQTLs. To compare the performance of HBI with other methods, we considered three main simulation scenarios: (1) there are genetic effects only in the most abundant cell type; (2) there are genetic effects only in the least abundant cell type; and (3) there are correlated genetic effects in all cell types. In each scenario, we further investigated whether the proportion of causal SNPs, minor allele frequency of causal SNPs, and noise/inaccuracy in the cell type proportions would affect the performance. Across these scenarios, we found that HBI generally improved the estimation of CTS meQTLs by achieving higher correlation, lower mean squared error (MSE) between estimated and true effect sizes, and higher power. There was substantial improvement in scenarios 2, when there were genetic effects in the least abundant cell type. For example, in this in scenario when the proportion of causal SNPs was fixed at 10%, using HBI to incorporate CTS data increased the power by 60%. In conclusion, across our simulation settings, HBI can incorporate strong and robust signals from MC-seq data to improve the estimation of CTS meQTLs. The applications of HBI to real data will enhance the understanding of the cellular specificity of complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4070W: Integrating clinical and genetic data into graph attention networks for disease risk prediction using Biobank data

Authors:

X. Huang, Z. Wang; Yale Univ., New Haven, CT

Abstract:

In the digital medicine era, electronic health records (EHRs) contain extensive patient data from diverse sources. Leveraging this readily available information is essential for personalized medicine and predictive healthcare. This study investigates the potential of integrating multimodal data into an attention-based graph network to enhance disease risk prediction and patient segmentation. We introduce a novel method, Bio-ALIGATEHR (ALIgning Graph Attention neTworks for EHR using Biobank), which combines information from various modalities, including clinical events and genetic data from Biobank. Bio-ALIGATEHR utilizes the kinship coefficient of patients from the UK Biobank to model genetic relatedness as edges in the attention network, which captures the genetic relationships among patients. This network is further augmented with an attention-based medical ontology representation to account for the

complex influence of disease dependencies. We demonstrate that explicitly modeling genetic relatedness significantly improves prediction accuracy for fibrotic diseases. Our results show average areas under the receiver operating characteristic curves (AUCs) of 0.82 for Ulcerative Colitis (UC), 0.76 for Systemic Sclerosis (SSc), and 0.65 for Crohn's Disease (CD), representing improvements of 9.2%, 11.3%, and 12.1%, respectively, compared to the baseline models GRAM and XGBoost that do not model genetic relatedness. Furthermore, the patient representations learned from our method effectively distinguish patient groups for these three fibrotic diseases. Overall, our findings highlight the importance of integrating diverse patient data sources through advanced computational models for a better understanding of the underlying mechanisms of diseases, leading to better-informed clinical decision-making and more effective patient care.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4071W: Integrating Perturb-seq and GWAS summary statistics identifies core genes for human complex disease ⭐

Authors:

S. Dorn[1], J. Miao[2], Z. Zhao[1], Q. Lu[1]; [1]Univ. of Wisconsin-Madison, Madison, WI, [2]Univ. of Wisconsin–Madison, Madison, WI

Abstract:

For many diseases, genome-wide association studies (GWAS) have found hundreds of statistically significant SNPs with small effect sizes, which are often hypothesized to have regulatory effects on the expression of nearby genes. However, it remains challenging to identify disease genes with large effects and biological mechanisms underlying infinitesimal associations across many genomic loci. We hypothesize that there are a small number of core genes that are key regulators of many important disease genes with GWAS hits. However, since these core genes have large effects on disease risk, their expression levels and regulatory genomic elements are highly constrained, making them challenging to uncover through GWAS. In Perturb-seq, highly constrained genes can be perturbed, allowing us to experimentally infer these genes' effects on expression levels. In this work, we present a statistical framework which identifies highly constrained core genes for complex disease through integrated analysis of Perturb-seq and GWAS summary statistics. Our approach measures the concordance of Perturb-seq effects, which quantify the direct impact of perturbations on downstream genes, and the transcriptome-wide gene-level associations for the disease in a matched cell type to identify crucial genes with large, synergistic regulatory effects on many downstream genes implicated in

disease risk. We applied our method to study coronary artery disease (CAD) using the largest publicly available CAD GWAS and a Perturb-seq dataset from human-derived endothelial cells (hECs) focusing on genes near known CAD GWAS loci. We identified three significant genes associated with CAD after multiple testing correction, including *NPR1* (p=7.24e-06), a gene encoding the receptor for natriuretic peptides ANP and BNP which are the targets of multiple therapeutic drugs for heart failure. *NPR1* has not previously been implicated in CAD through GWAS, but rare variants in its promoter cause heart failure in humans and *NPR1* knockout in mice and hECs lead to increased vascular endothelial cell adhesion. Our model can be applied in future Perturb-seq studies to associate experimental perturbation effects with complex human traits, providing novel insights into disease mechanisms and therapeutic targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4072W: Integration of Phenotype Similarity Embedding Model in Identifying Disease Genes Using Whole Exome Sequencing Data from Biobanks

Authors:

W. Zheng, Y. Xie, J. Gu, H. Zhao; Yale Univ., New Haven, CT

Abstract:

Following the success of Genome-Wide Association Studies (GWAS) that primarily focus on common variants, recent years have seen the discoveries of associations between rare variants and phenotypes through the analysis of Whole Exome Sequencing (WES) and/or Whole Genome Sequencing (WGS) data. The establishment of many large biobanks has facilitated much of this progress. However, many rare diseases are still underrepresented in large population cohorts, e.g., the UK Biobank and the All of Us Research Program, hindering the identifications of pathogenic genes associated with many rare diseases. Few methods take advantage of the extensive Electronic Health Records (EHR) from large biobanks to study rare disease genetics. We hypothesize that by aggregating all phenotype information for each individual, we can better define rare disease status based on the phenotype relationships among different diseases, and better infer pathogenicity of genes on diseases.We used static embedding methods to better embed the diagnosis data based on the phenotype relationships and identify phenotype differences between the potential pathogenic variant carrier group and the non-carrier group. We have developed statistical methods using those generative embeddings to better represent each individual's disease status and rank the potential risk genes for the disease of interest. We found that embedding representations reveal significantly

higher intra-group similarity scores for disease and rare loss-of-function (LoF) carrier groups compared to non-disease and non-carrier groups. Furthermore, our risk gene prioritization tool can identify risk genes for various Mendelian diseases, such as autosomal dominant polycystic kidney disease (ADPKD), in cases where large cohorts like UKBB have a very limited number of ADPKD patients, including both well-established ADPKD-associated genes (such as PKD1, PKD2, and IFT140), as well novel genes (GANAB, HNF1B, and DNAJB11) associated with this disease that traditional rare variant tests fail to detect. We also applied our approach to other diseases, such as neurofibromatosis, osteogenesis imperfecta, Marfan syndrome and so on. The identification of these additional risk genes highlights the power of our approach in detecting associations that may be overlooked by traditional methods with limited sample sizes. By leveraging the similarity between phenotype patterns of rare LoF variant carriers and disease patients in the embedding space, our proposed general framework can leverage the extensive EHR in biobanks to uncover novel risk genes, leading to better understanding of disease mechanisms and potential treatments.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4073W: Integrative genomic and transcriptomic investigation of human heart failure mechanisms - Trans-Omics for Precision Medicine in Congestive Heart Failure (TOPCHeF) ⭐

Authors:

C. Yang[1], M. Taylor[2], L. Mestroni[3], S. Chen[4], S. Graw[5], A. Karimpour-Fard[6], A. Manichaikul[7]; [1]Univ. of Virginia, Charlottesville, VA, [2]U of Colorado Anschutz Med. Campus, Aurora, CO, [3]Univ Colorado AMC, Aurora, CO, [4]Univ. of Colorado Cardiovascular Inst., Univ. of Colorado-Anschutz Med. Campus, Aurora, CO, [5]Colorado Genetics Lab., Dept. of Pathology, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [6]Dept. of BioMed. Informatics, Univ. of Colorado Sch. of Med., Aurora, CO, [7]Univ Virginia, Charlottesville, VA

Abstract:

Heart failure (HF) is a major public health problem in the United States, affecting >6.2M Americans, causing >80K deaths and >800K hospitalizations annually, and costing >$30B each year. The HERMES consortium carried out a GWAS of HF in 47,309 cases and 930,014 controls and identified the *PSRC1*, *ABO* and *MYOZ1* as HF putative target genes. However, the genetic mechanisms of HF have not been comprehensively explored. In the present study, we aim to elucidate the biological function of HF GWAS loci by integrating them more extensively with eQTL from heart tissue. To carry out this work, we build on summary statistics from the most

recent HERMES consortium GWAS of HF. In addition, we will leverage eQTL resources from GTEx left ventricular (LV) and Trans-Omics for Precision Medicine in Congestive Heart Failure (TOPCHeF). TOPCHeF study samples are derived from heart tissues from failing and non-failing (control) patients and TOPCHeF includes both transcriptome and whole genome sequencing data derived from >750 human LV tissue. In our initial analysis, we carried out genome-wide colocalization analysis to examine overlap of the HERMES HF GWAS signals with LV eQTL signals from GTEx v8. In addition to recapitulating results from the HERMES publication, we also identified colocalization of a genome-wide significant HF signal on chromosome 10 with a non-coding RNA RP11-464F9.21 near *MYOZ1*, suggesting this non-coding transcript may play a role in mechanism of the genetic signal and downstream regulation of *MYOZ1.* We further examined colocalization with suggestive signals of association for HF as a strategy to improve power to detect signals that do not reach a standard genome-wide significance threshold. We identified colocalization with expression of one or more genes for six regions demonstrating suggestive associations with HF ($P<5x10^{-6}$). Among these, a single region on chromosome 7 showed colocalization with expression of nine different genes in LV (*GTF2IP1*, *GTF2I*, *GTF2IRD2*, *PMS2P5*, *STAG3L1*, *PMS2P2*, *NSUN5P1*, *NCF1C*, *SPDYE5*). This group of genes overlaps a gene cluster encoding proteins containing helix-loop-helix motifs and have been implicated in Williams-Beuren syndrome which includes a substantial overlap with congenital heart disease. Our current results demonstrate the potential to elucidate overlap of GWAS signals with known and novel gene targets by leveraging the disease relevant eQTL that are forthcoming from the TOPCHeF study. In next steps, we will conduct eQTL mapping of LV RNA-seq from the TOPCHeF study in both diseased and control samples, and perform further integrative analysis to gain deeper insights into the putative targets underlying HF GWAS loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4074W: Investigating mode of inheritance in rare disease using aggregated rare variant association on phenotypes derived from phenome-wide deep learning

Authors:

D. Jordan, B. Petrazzini, R. Chen, J. Park, I. Forrest, H. Vy, A. Duffy, G. Rocheleau, R. Do; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Recent studies on the genetic architecture of human traits and diseases have suggested that traits thought to be recessive or dominant may actually have a range of intermediate

dominance. However, it has been difficult to test this hypothesis on rare variants or rare diseases, as non-additive genetic association studies require very large sample sizes to be powered. We have previously developed a deep learning phenotype imputation approach that allows us to conduct association studies on rare phenotypes (prevalence < 1 in 2,000) for aggregated ultra-rare coding variants (minor allele frequency < 10^{-5}). We report recessive and additive ultra-rare variant burden association tests for 155 rare phenotypes conducted using REGENIE in approximately 500,000 UK Biobank participants. Even with the lower power of recessive model association tests, we find many more significant disease-associated genes with the recessive model than the additive model (143 recessive vs. 39 additive), as well as a substantial bias of sub-significant effects towards recessivity. We find no clear examples of intermediate dominance, in contrast to results in common-variant and common-disease regimes. Overall, our findings suggest that the effects of rare variants on rare disease are largely recessive, with an additional significant role for additivity and minimal intermediate dominance. Testing these hypotheses has not previously been possible, providing a powerful example of how state-of-the-art deep learning methods applied to population-scale biobanks can reveal insights about the genetic architecture of human traits and diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4076W: Kinship estimation bias carries over to heritability estimation bias using variance components

Authors:

Z. Hou, A. Ochoa; Duke Univ., Durham, NC

Abstract:

Heritability is a fundamental parameter of diseases and other traits, quantifying the contribution of genetics to that trait as opposed to non-genetic environmental factors. Heritability is reflected in the extent to which relatives have similar phenotypes. The estimation of kinship matrices, also known as Genetic Relatedness Matrices or “GRMs”, is required for heritability estimation with many common approaches based on variant components. However, the most common “standard” kinship estimator employed by these approaches, including the popular GCTA package, can be severely biased in structured populations. In this study, we first characterize the theoretically predicted heritability estimation bias in the GCTA model due to kinship bias. Heritability estimation in this model requires unbiased estimates of the random effect coefficient, which we find is biased when the standard kinship estimator is used, and this bias depends only on the mean kinship value and the true heritability. We conduct simulation

studies to evaluate heritability estimation with various kinship matrices for scenarios such as admixture structure only and admixture plus family structure. The simulation results validate our theoretical prediction of the bias of the standard GCTA approach and show that our unbiased kinship matrices result in unbiased heritability estimates. A further complication is that upweighting rare variants in these kinship estimates, which is common practice, introduces an additional bias that has not been characterized in closed form. Then we apply various kinship matrices to several real datasets, such as the San Antonio Family Study and a Nephrotic Syndrome multiethnic cohort, to further characterize the source of biases and illustrate their extent in practice. We find that the heritability estimations using our unbiased kinship matrices lead to higher values, which are more consistent with previously published estimates in unstructured populations. Overall, we find that the most commonly used kinship estimator downwardly bias heritability estimation when there is population structure, and using an unbiased kinship estimator addresses this source of bias.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4077W: Leveraging Mendelian randomization for target discovery from machine-learning predicted phenotypes

Authors:

D. Amar, H. Somineni, C. O'Dushlaine; Insitro, SSF, CA

Abstract:

Predicting phenotypes is an appealing approach for increasing the sample size of a genetic study. However, the causal structure of the variables used for prediction is often unknown, which can lead to an inflation of false positive genome wide association studies (GWAS) hits due to the inclusion of downstream traits or confounded variables. Previous approaches for GWAS from predicted phenotypes focus on analysis of the predicted phenotypes in comparison to the true phenotypes. Mendelian Randomization (MR) offers an alternative approach, where genetic variants can be used as instrumental variables for identifying the causal connections between the outcome of interest and the features used as input for prediction. Recent progress in the MR field includes clustering-based methods like MR-Clust that offer a powerful way to address the heterogeneity of the causal estimates of single variants, while improving power by clustering independent variants into groups. Here, we study the utility of Mendelian randomization (MR) as an approach for delineating causal connections to help in target prioritization. We hypothesized that MR can help by both identifying mediators and consequences of the studied outcomes, and highlight genetic targets that are likely pleiotropic.

On simulated data we show that MR approaches can identify consequences of the disease and thus highlight both obvious false positive genetic targets, and suggest which phenotypes should be discarded from the prediction model. Of particular interest, clustering-based MR approaches can also highlight clusters of horizontally pleiotropic variants. While these pose a challenge for causal inference, they nevertheless are not considered false positives from a target discovery perspective. Finally, we apply the same methodology for the task of predicting PDFF liver fat from biomarkers using the UK-Biobank data. We show that the genetic variants of urate that are expected to be proper instrumental variables largely cluster around a null effect, suggesting that urate has null or negligible effect, and is therefore of minimal consequence as a causal factor of liver fat. However, we also show that there exists a large cluster of genetic variants that likely represent horizontal pleiotropy. We discuss similar results for ALT, but illustrate that unlike the urate example, the clusters that likely represent horizontal pleiotropy pathways constitute the majority of the variants, posing a challenge for MR methods that rely on the majority or plurality assumptions. In summary we show that MR pipelines can both improve power and reduce false positives when a set of features are considered for phenotype prediction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4078W: Limited overlap between genetic effects on disease susceptibility and disease survival: evidences from seven biobanks and nine common diseases

Authors:

Z. Yang[1], F-d. Pajuste[2], K. Zguro[3], Y. Cheng[4], D. Kurant[5], A. Eoli[6], J. Wanner[6], B. Jermy[7], J. Rämö[8], FinnGen, Estonian Biobank research team, S. Kanoni[9], D. A. van Heel[9], C. Hayward[10], Genes & Health Research Team, R. Marioni[11], D. McCartney[4], A. Renieri[3], S. Furini[3], R. Mägi[2], Genomics England Research Consortium, INTERVENE consortium, A. Gusev[12], P. Drineas[13], P. Paschou[14], H. Heyne[6], S. Ripatti[15], N. Mars[1], A. Ganna[16]; [1]Univ. of Helsinki, Helsinki, Finland, [2]Univ. of Tartu, Tartu, Estonia, [3]Univ. of Siena, Siena, Italy, [4]Univ. of Edinburgh, Edinburgh, United Kingdom, [5]Dana-Farber Cancer Inst., Boston, MA, [6]Hasso Plattner Inst., Potsdam, Germany, [7]BioMarin, Helsinki, Finland, [8]Broad Inst. of MIT and Harvard, Cambridge, MA, [9]Queen Mary Univ. of London, London, United Kingdom, [10]Univ Edinburgh, Edinburgh, United Kingdom, [11]edinburgh, stockport, United Kingdom, [12]Danaa Farber, Cambridge, MA, [13]Purdue Univ., West Lafayette, IN, [14]PURDUE Univ., West Lafayette, IN, [15]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [16]Inst. for Molecular Med. Finland, Helsinki, Finland

Abstract:

Understanding disease progression is of a high biological and clinical interest. Unlike disease susceptibility whose genetic basis has been abundantly studied, less is known about the genetics of disease progression and its overlap with disease susceptibility. Considering nine common diseases (N cases ranging from 11,980 to 124,682) across seven biobanks, we systematically compared the genetic architecture of susceptibility and progression, defined as disease-specific mortality. We identified only one locus significantly associated with disease-specific mortality and show that, at a similar sample size, more genome-wide significant loci can be identified in a GWAS of disease susceptibility. Variants that were significantly affecting disease susceptibility were weakly or not associated with disease-specific mortality. Moreover, susceptibility polygenic scores (PGSs) were weak predictor of disease-specific mortality while a PGS for general lifespan was significantly associated with disease-specific mortality for five out of ninediseases. We explored alternative definitions of disease progression and show that genetic signals for macrovascular complication in type 2 diabetes overlap with similar phenotypes in the general population, but effects are attenuated. Overall, our findings point to limited similarity in genetic effects between disease susceptibility and disease-specific mortality and suggest that either larger sample sizes, different measures of progression or integration of related phenotypes from the general population are needed to identify the genetic underpinning of disease progression.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4079W: Local ancestry adjusted *cis*-eQTL mapping of TB-T2D comorbidity elucidates the involvement of African ancestry in TB susceptibility.

Authors:

Y. Swart[1,2], C. Uren[1,2], C. Eckold[3,4], J. Cliff[5,6], S. Malherbe[1,2], K. Ronacher[7,8], V. Kumar[9,10], H. Dockrell[11,12], R. van Crevel[13,14], L. Kleynhans[7,15], M. Möller[1,2]; [1]Stellenbosch Univ., Cape Town, South Africa, [2]DSI-NRF Ctr. of Excellence for BioMed. Tuberculosis Res., South African Med. Res. Council Ctr. for Tuberculosis Res., Div. of Molecular Biology and Human Genetics, Faculty of Med. and Hlth.Sci., Cape Town, South Africa, [3]Imperial Coll., London, United Kingdom, [4]Faculty of Infectious and Tropical Diseases and Tuberculosis Ctr., London Sch. of Hygiene & Tropical Med., London, United Kingdom, [5]Brunel Univ., Uxbridge, United Kingdom, [6]Dept. of Life Sci., Brunel Univ. London, London, United Kingdom, [7]Univ. of Queensland, Brisbane, Australia, [8]Translational Res. Inst., Mater Res. Inst., Brisbane, Australia, [9]Radboud Univ. Med. Ctr., Groningen, Netherlands, [10]Dept. of Internal Med. and Radboud Ctr. for Infectious Diseases, Nijmegen, Netherlands, [11]London Sch. of Hygiene & Tropical Med., London, United

Kingdom, [12]Faculty of Infectious and Tropical Diseases and Tuberculosis Ctr., London, United Kingdom, [13]Univ. of Oxford, Oxford, United Kingdom, [14]Ctr. for Tropical Med. and Global Hlth., Nuffield Dept. of Med., Oxford, United Kingdom, [15]Translational Res. Inst., Mater Res. Inst. – The Univ. of Queensland, Brisbane, Australia

Abstract:

The validation of genome-wide association signals for tuberculosis (TB) susceptibility as well as the development of type 2 diabetes (T2D) across diverse populations are ongoing challenges. Identifying ancestry-specific expression quantitative trait loci (eQTLs) can aid in distinguishing the most probable disease-causing variants for population-specific therapeutic interventions. Therefore, *cis*-eQTL mapping was conducted in healthy controls, T2D, TB and TB-T2D patients. Both genotyping (Infinium H3A array with ~2.3 M markers) and RNA sequencing data of 96 complex multi-way admixed South Africans were used. Unique gene-variant pairs were associated with TB-T2D on chromosome 7p22 whilst adjusting for Bantu-speaking African ancestry (*PRKAR1B*:rs4464850; P=7.68e-07) and Khoe-San ancestry (*PRKAR1B:*rs117842122; P=3.66e-07). In addition, *IFITM3* (a biomarker for the development of TB) was associated with three eSNPs (rs11025530, rs3808990, and rs10896664) on chromosome 11p15 while adjusting for Khoe-San ancestry. Our results also indicated that the upregulation of the *NLRP6* inflammasome is strongly associated with people with TB-T2D while adjusting for Khoe-San ancestry. Two African-specific eGenes (*IFITM3* and *PRKAR1B*) would therefore have been missed if local ancestry adjustment was not conducted. The incorporation of local ancestry in *cis*-eQTL mapping led to the discovery of ancestry-specific eQTLs in a five-way admixed southern African TB-T2D cohort.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4080W: Mendelian Randomization Analysis of Latent Exposures Co-regulating Multiple Correlated Traits

Authors:

Y. Yu, A. Lakkis, J. Jin; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Mendelian Randomization (MR) is an increasingly popular tool for conducting causal inference using observational data from genome-wide association studies. In some settings, we want to identify causal signal among a set of correlated traits, such as phenotypes associated with different subtypes of a disease; in other words, the underlying exposure of interest (chronic

inflammation or mechanisms of different subtypes of a disease) may not be directly observable, but measurements can be available on multiple traits that are coregulated by the exposure. We propose methods for conducting MR analysis on latent exposures, which test the significance for, and the direction of, the effect of latent exposures by leveraging information from multiple related traits. Simulation studies show that the proposed methods have well-controlled type I error rates and enhanced power compared to single-trait MR tests under various types of pleiotropy. We will also illustrate the performance of the method for jointly analyzing the causal effects of different subtypes of psychiatric disorders on heart diseases and digestive disorders.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4081W: Mendelian Randomization does not implicate MS risk factors as having a role in NMOSD susceptibility.

Authors:

E. Misicka[1], S. Liu[1], D. Crawford[1], F. Briggs[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Miami, Miami, FL

Abstract:

Background. Neuromyelitis Optica Spectrum Disorder (NMOSD) is a rare autoimmune demyelinating disease of the central nervous system, affecting ~22,000 Americans. Due to the challenges of studying a rare disease, very few NMOSD risk factors have been identified. However, it is possible that NMOSD shares risk factors with similar diseases like multiple sclerosis (MS). Past modest-in-size epidemiologic studies have suggested a limited role for MS risk factors in NMOSD; however, such studies may be vulnerable to confounding, reverse causality, and measurement error. Robust causal analyses are needed to confirm these findings. Genetic instrumental variable (GIV) analyses that utilize summary statistics from genome-wide association studies (GWAS) can examine relationships between factors that may otherwise be confounded in epidemiologic studies or not possible to explore for a rare disease.
Objectives. To use Mendelian randomization (MR) to investigate the potential causal relationships between various MS risk factors and NMOSD.
Methods. Two-sample (2S) MR was conducted to investigate the relationships between MS risk factors and NMOSD. MS risk factors for 2SMR included serum Vitamin D (VitD), body-mass index (BMI), age at menarche (AAM), and telomere length (TL). Genetic instruments were identified from GWAS top hits for the respective exposures. Summary statistics for NMOSD IgG seropositivity were extracted from a GWAS of 230 IgG+ NMOSD cases and 718 healthy controls (HCs) from the Accelerated Cure Project, adjusting for age, sex, and population substructure.

Inverse-variance weighted (IVW) meta-analysis was conducted for each exposure-NMOSD relationship, and tested for horizontal pleiotropy using the Egger intercept test. Single-SNP and Leave-one-out sensitivity analyses were also conducted.
Results. Genetic instruments for VitD, BMI, AAM, and TL were constructed from 141, 656, 377, and 197 GIVs, respectively. After extraction from the NMOSD summary statistics and aligned for effect alleles, the final instruments were composed of 112, 650, 340, and 131 GIVs. Under IVW analysis, only BMI had a trending association with NMOSD IgG seropositivity (p=0.07). No horizontal pleiotropy was detected via Egger intercept test for any analysis (p>0.05 for all exposures). Sensitivity analyses did not indicate effects driven by individual SNPs in any exposure-outcome associations (p>0.05).
Conclusions. Genetically-driven VitD, BMI, AAM, and TL that confer MS risk do noi influence NMOSD susceptibility. Future work will seek to refine genetic instruments for the exposures under study as well as expand the search for NMOSD risk factors.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4082W: Meta Analysis of Rare Disease Cohorts Discovers Novel Disease-Causing Genes and Processes Governing Disease Prevalence

Authors:

M. Moldovan, V. Seplyarskiy, S. Kobren, Undiagnosed Diseases Network Tool Building Coalition Working Group, S. Sunyaev; Harvard Med. Sch., Boston, MA

Abstract:

Trio sequencing and analyses of de novo variation in patients affected by rare conditions has not only yielded multiple new diagnostic disease gene discoveries, but has also resulted in the availability of a substantial amount of genomic data for the research community.We recently developed RaMeDiES-DN, a powerful and unbiased statistic for assessing cohort-level enrichment of de novo mutations. Our statistic is closed-form, which leads to a substantial decrease in runtime without loss of statistical power compared to the state-of-the-art DeNovoWEST algorithm. RaMeDiES-DN operates on cohort-level summary statistics, enabling the simultaneous analysis of multiple cohorts without sharing any individual variant-level data. Here, we aggregated 53 published datasets of de novo mutations in cohorts of patients with conditions ranging from neurodevelopmental disorders (NDDs) to heart failure. We developed a phenotype-agnostic approach to maximize statistical power for gene discoveries by merging cohorts in a specific order. With this procedure, the power of diagnostics and novel disease-causing gene discoveries is largely enhanced in the joint analyses relative to single phenotype-

cohort analyses and naïve aggregation of all datasets into one. So far, this has yielded 12 novel genes passing the formal Bonferroni-corrected significance threshold. For instance, joint analysis of heart Disease and schizophrenia cohorts yielded, apart from known genes causing respective conditions, a calcium ion release regulator ITPRID1 as a potential discovery.In large cohorts like ones available for NDD it is possible to estimate gene-level penetrance, that is, how frequently a de novo mutation in a gene will drive the phenotype. Intriguingly, 11 genes were found to have more de novo mutations than expected under the conservative assumption of full penetrance. This paradox could be explained by clonal selection in sperm, where certain variants in sperm progenitors are favored over ancestral states, leading to their propagation. Indeed, disruptions to four of these genes (PTPN1, MAP2K1, FOXG1 and PTEN) are known to cause both severe congenital conditions and clonal expansions in sperm. The other seven genes are chromatin regulators and protein kinases, four of which have been associated with other types of clonal expansions, such as cancers. Overall, de novo mutations in these genes comprise about 10% of all diagnostic de novo mutations in the NDD cohort.Our work shows that reanalysis and joint analysis of multiple cohorts of patients affected by rare diseases has a potential to uncover both novel genes and pathways contributing to individual and comorbid conditions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4083W: Metabolic and lipidomic profiles reveal crucial mediators and pathways influencing bone mineral density

Authors:

K. Su[1], L. Anqi[1,2], Z. Qi[3], L. Jiang[1], A. Liu[1], X. Zhang[1], J. Genzer[1], S. Sajjan[1], Z. Luo[1], Q. Tian[1], C. Qiu[1], H. Shen[1], H-W. Deng[1]; [1]Tulane Ctr. for BioMed. Informatics and Genomics, Deming Dept. of Med., Sch. of Med., Tulane Univ., New Orleans, LA, [2]Dept. of Biostatistics and Data Sci., Sch. of Publ. Hlth.and Tropical Med., Tulane Univ., New Orleans, LA, [3]Dept. of Preventive Med., Coll. of Med. – Memphis, The Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

Abstract:

Osteoporosis is a significant global health issue characterized by reduced bone mineral density (BMD), increased fracture risk, and diminished quality of life. Despite advancements in research, particularly regarding women, understanding in men—who face higher fracture and post-fracture mortality rates—remains limited. Our study examines the impact of metabolite mediators, which can be influenced by nutrient intake, on male bone health. We analyzed 989 males from the Louisiana Osteoporosis Study, including 391 African Americans and 598 White

individuals, aged 20 to 51 years (mean age 37). Data collection involved metabolomics and lipidomics profiling, alongside comprehensive nutrient assessments. BMD measurements were taken at the forearm, hip, spine, and femoral neck using dual-energy X-ray absorptiometry to explore how metabolites mediate nutrient effects on BMD.Our statistical methods included pairwise and multivariate mediation analysis and dimensional reduction, aiming to pinpoint mediators' roles. Pathway analysis helped elucidate the biological functions of these metabolites. We identified 88 metabolic and 15 lipidomic mediators that consistently affect BMD across different nutrient conditions. Notably, 10-undecenoate (11:1n1) and undecenoylcarnitine (C11:1) were found to have bidirectional effects on all BMD measurements, indicating their complex roles in bone metabolism. Phosphatidylethanolamine was highlighted as crucial in bone development. We observed minimal overlap among mediators across different BMD sites, suggesting unique regulatory mechanisms for trabecular and cortical bones. Additionally, we discovered novel pathway-level mediators, such as those in porphyrin metabolism, which negatively impacted femoral neck BMD under various nutrients. Complex bidirectional effects in pathways like bile acid synthesis and fatty acid biosynthesis further illustrate the intricate regulation of BMD. This study clarifies the relationships between metabolites, metabolic pathways, and BMD, offering new insights to enhance osteoporosis prevention and support personalized dietary strategies to optimize bone health.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4084W: Metabolite mediation of pathways from genotype to disease in the Canadian Longitudinal Study on Aging

Authors:

J. Min[1], O. Vishnyakova[2], A. Brooks-Wilson[3], L. Elliott[1]; [1]Simon Fraser Univ., Burnaby, BC, Canada, [2]Genome Sci. Ctr., BC Cancer, Vancouver, BC, Canada, [3]BritishColumbia Cancer Res Ctr, Vancouver, BC, Canada

Abstract:

Causal mechanisms linking genetic variants and health outcomes is a difficult follow-up analysis in genome-wide association study with no current best practice. Causal mediation analysis is a robust approach to identifying, estimating, and interpreting causal effects of exposures on outcomes. This method decompoeses the total effect of exposure into direct and indirect effects, uncovering potential pathways between the exposure, mediator, and outcome. We present a novel machine-learning-based causal mediation analysis framework to uncover mechanisms linking genetic variants, the metabolome, and age-related diseases using data from

the Canadian Longitudinal Study on Aging (CLSA), allowing for multiple mediating metabolites for each variant/disease pair. Our framework advances the understanding of complex interactions between genetics, metabolomics, and diseases, and may have impact in targeted interventions and personalized medicine.Our framework integrates weighting-based causal mediation analysis with machine learning algorithms. This approach flexibly estimates causal effects while accommodating multiple mediators, avoiding biases from incorrect model specification inherent in regression-based methods. By incorporating ML-based variable selection techniques like LASSO, we identify sparse sets of mediating metabolites. Their joint causal effects and the uncertainties of the effects are estimated through bootstrapping,and are assessed with sensitivity analysis. The direction and magnitude of individual mediators' contributions to the joint indirect effects are determined via permutation tests. This comprehensive method can be expanded to account for interactions and non-linear associations.This study has identified 53 significant genotype-metabolome-disease mediation effects in CLSA, including 12 for osteoarthritis, 7 for osteoporosis, 7 for Parkinson's disease, and 27 for type-2 diabetes. The findings include pathways connecting known gene-metabolome and metabolome-disease associations. For instance, the SNPs in *SLC13A5*, *GLDC*, and *NAT8* may play causal roles in the risk of osteoarthritis through citrate, glycine, and methionine sulfone, respectively. These findings provide insights into underlying genotype-to-disease mechanisms and offer a foundation for future research.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4085W: MetaGLIMPSE: Meta Imputation of Low Coverage Sequencing Data for Modern and Ancient Genomes

Authors:

K. Kumar[1], S. Rubinacci[2], S. Zollner[3]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Brigham and Women s Hosp., Boston, MA, [3]Univ Michigan, Ann Arbor, MI

Abstract:

Imputation was initially developed for SNP array data. Now, imputation can enhance low coverage sequencing, which offers a cost-effective, ancestry-unbiased alternative, aiding rare variant identification in non-European populations. The most commonly used tool designed for imputing low coverage sequencing data is GLIMPSE (Rubinacci et al. 2021). As for all imputation methods, its accuracy varies by reference panel ancestry and size. A single, large reference panel is difficult to obtain for non-European ancestries as they are underrepresented in large reference panels and may be better represented in smaller ancestry specific panels. The optimal

method leverages genotypes from multiple reference panels, creating a 'mega' panel. However, creating a mega panel is often not possible because the individual data is protected for privacy reasons.We design a novel meta-imputation method, MetaGLIMPSE, that combines estimates from multiple reference panels using weights tailored for each individual and genomic region. We use a diploid Li and Stephens-inspired Hidden Markov Model to account for the difficulty of pre-phasing. Unlike existing meta-imputation methods, we account for genotype uncertainty by integrating over genotypes in the emission probability, resulting in meta-imputed genotype dosages.We evaluate our method by comparing it to the gold standard of imputing with a 'mega' panel- a single panel combining both ancestry-specific panels. While using a mega panel is not possible in practice due to privacy concerns, we use 1000 Genomes data to make this possible. We meta-impute a range of populations: Europeans with European reference panels, African-Americans with African and European reference panels, and East Asians with African and European reference panels, which mimics samples that are not in any available reference panel. We find our meta-imputed samples perform better than the best single panel imputation at coverages 0.5x - 8x. In addition, we find that at very low coverages for East Asians imputed with African and European panels, the performance of the mega imputation is worse than single panel imputation, contradicting commonly held beliefs about imputation panel design.This observation could reshape imputation panel design principles, especially for non-European populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4086W: MetaSCAMPI: A scalable variance- and covariance-based meta-analysis framework for genome-wide interaction testing utilizing multiple traits

Authors:

S. Bian[1], X. Li[2], M. P. Epstein[3]; [1]Emory Univ., Atlanta, GA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]EMORY Univ., Atlanta, GA

Abstract:

Exploring non-additive effects in genetic studies is challenging due to the modest magnitude of interaction effect sizes, multiple-testing burden, complications in determining interaction exposures, and extensive computational costs. We previously proposed a powerful, scalable screening method, SCAMPI (Scalable Cauchy Aggregate test using Multiple Phenotypes to test Interactions), which leverages the variance-covariance structure of multiple phenotypes to improve performance in prioritizing SNPs with interaction effects. Simulations and real data

analysis using UK Biobank data demonstrated that SCAMPI preserved type-I error and improved power compared to traditional univariate variance-based interaction testing across various scenarios.
Meanwhile, with the availability of multiple large cohort and biobank genetic studies, meta-analysis of such studies provides a natural and cost-effective solution to augment sample size in interaction studies. Compared to the analysis of individual-level data, meta-analysis requires each participating study to share only summary statistics, which protects the data privacy of study participants and reduces challenges in sharing and integrating large subject-level data. To harness summary statistics for interaction testing, we propose MetaSCAMPI, which combines effect estimates from multiple participating studies for inference without the need to pool individual-level data.
We implement MetaSCAMPI in two ways: first, by aggregating matrix-level summary information of SCAMPI from multiple studies; second, by approximating the distribution of the pooled statistics of the 2-degree-of-freedom test in SCAMPI. We demonstrate via simulations that MetaSCAMPI can preserve type-I error and enhance power under various scenarios, including different sample size, effect size, distribution, and number of studies. We also apply MetaSCAMPI to UK Biobank via meta-analysis of multiple subsets of the data (total N=337,422) to investigate the interaction effects for 288,910 SNPs post QC on four phenotypic outcomes, HDL-C, LDL-C, triglycerides, and BMI. MetaSCAMPI delivered very similar results compared to the pooled individual-level analysis results using traditional SCAMPI. We implement MetaSCAMPI within an R package for ease of use.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4087W: *ML-MAGES*: a computationally efficient tool for multivariate genetic association analyses with effect size shrinkage via machine learning ⭐

Authors:

X. Liu, L. Crawford, S. Ramachandran; Brown Univ., Providence, RI

Abstract:

A fundamental goal of human genetics studies is to identify genetic variants associated with complex traits and diseases; this is often achieved through genome-wide association (GWA) studies. Methodological challenges exist when analyzing GWA results, including accounting for inflation in GWA effect estimates and investigating more than one trait simultaneously. We leverage machine learning approaches for these two challenges, developing a computationally

efficient method called *ML-MAGES*. It takes GWA summary statistics of multiple traits and linkage disequilibrium (LD) between variants as inputs, and summarizes the associations of variants as well as genes to the traits, estimating strength and identifying shared associations. We first shrink the inflation in GWA effect sizes caused by LD via neural networks, and then cluster variant associations among multiple traits via variational inference. In approximating simulated true effects, our neural network shrinkage (RMSE: 0.047-0.071) outperforms regularized regression and fine-mapping, two approaches used for addressing inflated effects but in focal regions of different sizes (RMSE: 0.118 and 0.219). Our infinite mixture clustering offers a flexible, data-driven way to distinguish different types of associations across traits—trait-specific, shared across traits, or spurious. Importantly, our shrinkage method drastically reduces runtime and memory, processing 15,000 variants in 0.8 seconds with linear memory usage (regularized regression: 15.3s; fine-mapping: 31.8s; both with exponential memory). We demonstrate the application of *ML-MAGES* on association analyses of several quantitative traits in the UK Biobank. Genes identified from single-trait enrichment tests overlap with those significantly enriched in relevant gene sets from existing knowledge bases. Besides trait-specific associations, our method identifies shared associations across traits, suggesting putative shared genetic architecture. For example, when analyzing the GWA effects of high-density and low-density lipoprotein cholesterol (HDL and LDL), *APOC4* shows a shared association with both, while *LPL* (HDL) and *PCSK9* (LDL) show evidence of being trait-specific.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4088W: Model introspection techniques enable scalable variant effect prediction from genomic sequence models and improve mechanistically-driven functional fine-mapping

Authors:

C. Cole, T. T. Lau, R. E. Denroche, K. Callegari, B. Frey, T. Masud; Deep Genomics, Toronto, ON, Canada

Abstract:

Genomic sequence models have demonstrated great promise for assessing the impact of naturally occurring variation on molecular phenotypes. However, leveraging their predictions to identify causal variants underlying genome-wide association study (GWAS) loci has been challenging due to the scale of inference required and a lack of a systematic framework to assess their impact on fine-mapping. In this study, we extend the BaselineLF-v2.2 annotation set using gradient-based attribution methods and a recently described foundation model, BigRNA.

We score variants within each protein coding gene and a surrounding 50 kilobase flanking window for impacts on several molecular phenotypes including aberrant splicing, binding affinity for a variety of miRNAs, and altered RNA-seq coverage profiles. Our dataset of predictions covers over 1.7 billion positions of the human genome. Using the UK Biobank we investigate patterns of enriched heritability across 20 representative traits and find evidence for significantly enriched heritability and $\tau_{;}^{*}$ for our new annotations meta-analyzed across traits (Bonferonni corrected P-value < 0.05). We extend stratified LD score regression (S-LDSC) to create a novel simulation framework to assess the empirical impact of adding a new annotation with a given $\tau_{;c}$. Using a before-and-after comparison, we introduce annotations with a varying amount of per-variant heritability within 176 real GWAS loci and find that a 1% increase in conditionally independent heritability increases the posterior inclusion probability (PIP) of simulated causal variants by 0.4 in approximately a quarter of cases. We find that our annotation is calibrated to achieve differences in PIPs for causal variants and conduct an investigation using the newly introduced replication failure rate (RFR) paradigm by downsampling the same 20 traits to 100k individuals and comparing the proportion of variants that replicate in the full set with the BaselineLF-v2.2 model and our AI-augmented set. We find decreases in the RFR across traits, indicating improved calibration after the inclusion of AI predictions. We apply our new framework to the UK Biobank and identify case studies where our improved fine-mapping resolves complex loci and, in several cases, alters the suspected causal gene. Post-hoc predictions suggest new mechanisms and recapitulate known effects on disease-altering molecular phenotypes. Together, these results suggest that genomic sequence models represent a valuable addition to functionally-informed fine-mapping techniques and can provide valuable insights into disease-altering processes with post-hoc interpretability.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4089W: MRScan enables heritable confounder selection from public GWAS databases for bias reduction in multivariable Mendelian randomization

Authors:

R. Shi, J. Morrison; Univ. of Michigan, Ann Arbor, MI

Abstract:

Mendelian randomization (MR) is a method to estimate causal effects using genetic variants as instruments. A key challenge in MR studies is heritable confounding, which occurs when instruments affect traits that confound the exposure and outcome, leading to bias in causal

estimates. As more traits have been shown to have heritable components, we are increasingly aware of the commonality of heritable confounding. Multivariable Mendelian randomization (MVMR) can adjust for heritable confounding when GWAS summary data of known confounders are available. MVMR studies typically select a small set of confounders based on domain knowledge. However, this strategy may omit key confounders, inducing residual bias, false positives and decreased power. We propose a novel phenome-wide heritable confounder selection algorithm, MRScan, using public GWAS data in the MRC-IEU OpenGWAS database. MRScan enables users to identify previously unknown heritable confounders which can be adjusted in MVMR, enhancing accuracy in causal estimates. MRScan begins with a phenome-wide scan for traits that share associated variants with the main exposure. After quality control, we perform bi-directional MR and include traits as candidate confounders if they may be upstream of either the exposure or outcome and are not causally downstream of either one. We then clump traits by genetic correlation to remove redundancy, and filter out traits with insufficient instrument strength. We recommend conducting MVMR analysis including the full final set of traits using weak-instrument robust methods. MRScan also provides alternative options including LASSO, marginal, and stepwise selection. MRScan, implemented as an R package and Snakemake pipeline, offers multiple customizable settings. Users can adjust parameters, manually modify the confounder list, choose among multiple MVMR methods, and generate reports and diagnostic plots. We demonstrate the utility of our tool in the study of the causal effects of C-reactive protein (CRP) levels on multiple diseases. Prior MR studies of CRP either perform univariable MR (UVMR) or adjust only for BMI. In contrast, MRScan detects confounders between CRP-level and each disease. We find that confounder adjustment substantially impacts conclusions for some outcomes. For example, UVMR suggests significant causal effects of CRP-level on the risk of coronary artery disease, stroke, knee osteoarthritis, and diabetes. However, these effects are nullified after adjusting for confounders. This highlights the need to account for heritable confounders and MRScan's ability to identify them.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4090W: Multi-Ancestry Analysis Identifies Susceptibility Variants and Improves Polygenic Risk Scores for Breast Cancer Subtypes

Authors:

H. Zhang[1], X. Wang[2], T. Ahearn[3], K. Michailidou[4], R. Milne[5], J. Simard[6], P. Pharoah[7], S-H. Teo[8], M. Schmidt[9], D. Easton[10], P. Kraft[11], W. K. Ho[12], M. Garcia-Closas[3], Breast Cancer Association Consortium; [1]Natl. Cancer Inst., Cambridge, MA, [2]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst.,, Rockville, MD, [3]Natl. Cancer Inst, Rockville, MD, [4]The Cyprus Inst. of

Neurology and Genetics, Nicosia, Cyprus, [5]Cancer Council Victoria, Melbourne, Australia, [6]CHU de Québec - Université Laval Res. Ctr., Quebec, QC, Canada, [7]Univ Cambridge, Cambridge, United Kingdom, [8]Breast Cancer Res. Programme, Cancer Res. Malaysia, Subang Jaya, Malaysia, [9]17Div. of Molecular Pathology, The Netherlands Cancer Inst., Amsterdam, Netherlands, [10]Univ. of Cambridge, Cambridge, United Kingdom, [11]Natl. Cancer Inst., Rockville, MD, [12]21Sch. of Mathematical Sci., Faculty of Sci. and Engineering, Univ. of Nottingham Malaysia, Jalan Broga, Malaysia

Abstract:

Background: Breast cancer is a heterogeneous disease with diverse molecular subtypes, each exhibiting distinct clinical features and outcomes. Most studies focus on European populations, potentially missing key insights from non-European groups. Current PRS models primarily address overall risk and estrogen receptor (ER) status, without considering subtypes defined by progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and tumor grade. We propose a multi-ancestry analysis to identify novel genetic variants with heterogeneous associations across subtypes and use advanced PRS methods to develop comprehensive risk prediction models considering subtypes defined by ER, PR, HER2, and tumor grade.

Methods: This multi-ancestry genome-wide association study (GWAS) includes 119,962 invasive breast cancer cases and 106,377 controls from the Breast Cancer Association Consortium (BCAC), comprising 81% European (EUR), 16% East Asian (EAS), 2% African (AFR), and 1% Hispanic (HIS) populations. We applied two-stage polytomous logistic regression within each ancestry to identify variants associated with breast cancer subtypes, followed by fixed-effect meta-analyses. PRS models were developed using seven methods, including clumping and threshold (CT), LDpred2, CT-SLEB, PRS-CSx, and PROSPER. A super learning model combined these PRSs for improved risk prediction.

Results: Our analysis identified 11 novel susceptibility loci ($P < 5 \times 10^{-8}$), with two specific to EAS (rs138209001, $P_{EAS} = 8.72 \times 10^{-10}$; rs7615125, $P_{EAS} = 3.04 \times 10^{-10}$). The remaining loci were identified through multi-ancestry meta-analyses. For overall breast cancer risk, the best PRS model achieved an AUC of 0.664 for EUR, 0.621 for EAS, 0.572 for AFR, and 0.621 for HIS. The best PRS in EUR generated an OR per SD of 1.84 (95% CI: 1.79-1.89). Compared to the CT approach, the multi-ancestry PRS approach (CT-SLEB) improved AUC by 7.2% on average across ancestries. PRS for the subtype, defined as ER or PR+, HER2-, and tumor grade 1 or 2, showed higher AUCs than for triple-negative tumors in EUR (0.68 vs. 0.64) and EAS (0.63 vs. 0.59), but not in AFR (0.57 vs. 0.62) or HIS (0.61 vs. 0.69). Notably, for triple-negative subtypes, subtype-specific PRSs increased AUC by 9.6% on average across all ancestries compared to using overall PRS.

Conclusion: Our study demonstrates the effectiveness of multi-ancestry GWAS and PRS

methods in identifying novel genetic variants and improving risk prediction for breast cancer subtypes. Including diverse populations enhances our understanding of genetic contributions, paving the way for more personalized risk assessment and prevention strategies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4091W: Multi-ancestry gene expression prediction identifies novel genes regulating complex traits.

Authors:

S. Golzari[1], K. Akamatsu[2], T. Amariuta[3]; [1]Shu Chien-Gene Lay Dept. of Bioengineering, UC San Diego, San Diego, CA, [2]Sch. of Biological Sci., UC San Diego, San Diego, CA, [3]Dept. of Med., Div. of BioMed. Informatics, Halıcıoğlu Data Sci. Inst., UC San Diego, San Diego, CA

Abstract:

Understanding how genetic variation modifies gene expression can implicate mechanisms of disease-associated GWAS variants. However, eQTL and GWAS data from non-European populations is severely limited, restricting our understanding of disease to European genetic architectures. While polygenic risk scores are improved by multi-ancestry data integration, this paradigm is underexplored in the genetic prediction of gene expression, a critical step in associating gene regulation with disease. We developed MAGEPRO (Multi-Ancestry Gene Expression Prediction Regularized Optimization) to build powerful gene expression prediction models using eQTL summary statistics from diverse ancestries. MAGEPRO estimates posterior eQTL effect sizes in a target population with limited genotype/gene expression data as a function of a target-specific noisy prior and fine-mapped eQTL summary statistics from European, African American, and Latino populations. We applied MAGEPRO to monocyte gene expression from 242 Latinos and LCL gene expression from 346 African Americans and 364 Europeans across thousands of *cis*-heritable genes. MAGEPRO improved the accuracy of Hispanic models by 8.3% relative to the conventional noisy prior model (p=3x10^-102), by 6.3% relative to fine-mapping eQTL in the target population (p=3x10^-45), and by 0.6% relative to a modified version of the widely used PRS-CSx method, that now models data from the target population (p=0.01). Notably, MAGEPRO increased the accuracy of 556 genes by >50% relative to the conventional model and generated a predictive gene model (accuracy>0, p<0.05) for 45.9% of genes where conventional models had failed to do so. While MAGEPRO and PRS-CSx performed comparably well in predicting LCL gene expression, MAGEPRO outperforms PRS-CSx in smaller target cohorts (e.g., N=100, 4.6% improvement, p=8x10^-11). In out-of-sample validation, MAGEPRO modestly outperformed PRS-CSx (2% improvement, p=9x10^-6),

demonstrating the increased generalizability of our models beyond the target cohort. We used MAGEPRO LCL models to perform a transcriptome-wide association study (TWAS) across 15 blood cell traits in African (N=13,298) and European (N=511,471) individuals (Chen et al. 2020 Cell), identifying 111 and 802 trait-associated genes (p<0.05/6,894 genes), respectively, not detected by standard TWAS. Gene ontology analysis revealed these genes participate in immune cell differentiation and activation processes. In this work, we demonstrate how ancestrally diverse data can enhance eQTL effect size estimation, thereby nominating new risk genes in the population-specific regulation of complex traits and disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4092W: Multi-modal Polygenic Risk Scores and Functional Annotations Improve Glaucoma Prediction Across Ancestries

Authors:

X. Gao, M. Chiariglione; The Ohio State Univ., Columbus, OH

Abstract:

Glaucoma is the leading cause of irreversible blindness, affecting 70 to 90 million individuals worldwide, with primary open-angle glaucoma (POAG) constituting approximately 90% of cases in North America. Currently, there is no cure for glaucoma, making early detection vital. Although polygenic risk scores (PRSs) have been used to predict various diseases, their performance has often been limited by inadequate genomic coverage and insufficient functional integration. This study demonstrates that aggregating genetic factors related to POAG, intraocular pressure, and vertical cup-to-disc ratio, along with their associated functional annotations, markedly enhances the predictive accuracy of PRSs for POAG. We utilized genome-wide association study (GWAS) summary statistics for glaucoma and its endophenotypes derived from measured and imaging phenotypes (e.g., Gao *et al.* 2018; Gharahkhani *et al.*, 2021; Han *et al.*, 2021), predominantly from individuals of European descent. Our multi-modal PRSs, incorporating over seven million genetic markers and more than 70 functional annotations, were evaluated using the UK Biobank dataset (n = 324,713, independent of GWAS samples) and the Mexican American Glaucoma Genetic Study (MAGGS) dataset (n = 4,549). For European participants in the UK Biobank, our PRSs achieved an area under the receiver operating characteristic curve (AUC) of 0.814 (95% CI: 0.806-0.823), representing a 9.3% ($P < 2 \times 10^{-16}$) improvement over the baseline model that included only age and sex. In Latino participants of the MAGGS, the PRSs reached an AUC of 0.792 (95% CI: 0.763-0.819), improving prediction accuracy by 5.2% ($P = 2.13 \times 10^{-6}$) over the baseline. These findings underscore the

significant impact of extensive genomic coverage and functional annotations on enhancing the precision of glaucoma predictions, supporting advancements in personalized medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4093W: Multipopulation Genome Wide Association Study of Age at Menarche reveals potentially novel variants linked to TGF-β and ER stress

Authors:

H. Rao[1], K. L. Young[2], K. E. North[2], S. Buyske[3], S-S. Park[4], I. Cheng[5], C. Kooperberg[6], J. Haessler[6], L. Wilkens[4], L. Le Marchand[4], T. Matise[3], L. Fernandez-Rhodes[1]; [1]Pennsylvania State Univ., University Park, PA, [2]Univ North Carolina, Chapel Hill, NC, [3]Rutgers Univ, Piscataway, NJ, [4]Univ. of Hawai'i Cancer Ctr., Honolulu, HI, [5]Cancer Prevention Inst California, Fremont, CA, [6]Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Abstract:

Age at menarche (AAM) is an important milestone for female reproductive health and early onset has been associated with breast cancer, endometrial cancer, pre-eclampsia, low birthweight, childhood obesity and smoking. While the population mean AAM is 13 years, there is large variation across populations with some having significantly earlier onset. Multipopulation studies of AAM can assist in identifying new loci for AAM and generating new hypotheses that link reproductive timing to disease outcomes.
We performed a discovery GWAS in 56,092 (22,801 African American (AA), 15,309 Hispanic/Latina, 5498 Asian, 2282 Native Hawaiian, and 888 American Indian/Alaskan Native) females as part of the Population Architecture using Genomics and Epidemiology (PAGE) study. We used a generalized estimating equation model to predict AAM, adjusting for age, center, and genetic ancestry principal components. Subsequently we meta-analyzed findings from PAGE with an additional 19,646 diverse females from ReproGen/UK Biobank.
Average AAM ranged from 11-13 years across our studies. Our analyses identified 22 genome-wide significant loci ($p<5x10^{-8}$); with 3 potentially novel signals. One variant, rs34155040, is 22 kb upstream of *SKOR2*, a coding gene with a role in the negative regulation of the transforming growth factor-β (TGF-β) pathway; previously linked to ovarian reserve and anti-Müllerian hormone levels. This marker is also 95 kb upstream of *IER3IP1*, an endoplasmic reticulum (ER) stress moderating gene that has been previously associated with diabetes. We also identified a signal (rs79656726, which is monomorphic in European and Asian populations) near *RPS24* which encodes ribosomal proteins and has been linked to congenital anemia.

Race/ancestry specific analyses in AA revealed one potentially novel genome-wide significant locus on chr4. The intronic variant in the locus, rs4699646, is in the *TSPAN5* gene which plays a role in epithelial mesenchymal transition (EMT) and TGF-β pathway. Together, our results suggest the role of AAM in TGF-β, ER stress, and EMT pathways, which have linked functions with other reproductive hormones as well as cell growth and adhesion.
Our study reveals potential links between pubertal timing and TGF-β, ER stress, and EMT pathways in a diverse populations. Efforts to study variants associated with AAM can be instrumental in advancing public health measures and better understand reproductive timing, follicle reserve, cancers, and pregnancy outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4094W: Multi-Trait GWAS Identifies Shared and Novel SNPs in Fibroproliferative Diseases Among Non-Hispanic Black (NHB) and White (NHW) Populations

Authors:

G. Hampton[1], J. Hellwege[2], C. Greene[3], J. Jaworski[4], M. J. Puckelwartz[5], Y. Luo[6], D. Velez Edwards[7], T. Edwards[8]; [1]Vanderbilt Univ., Madison, TN, [2]Vanderbilt Univ. Med. Ctr.,Div. of Genetic Med., Dept. of Med., Vanderbilt Genetics Inst.(VGI), Vanderbilt Epidemiology Ctr., Nashville, TN, [3]Vanderbilt Univ., Nashville, TN, [4]VUMC, Nashville, TN, [5]Northwestern Univ., Assistant Professor, Pharmacology, Chicago, IL, [6]Northwestern Univ.,Director, Inst. for Artificial Intelligence in Med. - Ctr. for Collaborative AI in Hlth.care Associate Professor, Preventive Med. (Hlth.and BioMed. Informatics), Pediatrics Associate Professor, McCormick Sch., Evanston, IL, [7]Vanderbilt Univ., Professor of Obstetrics and Gynecology, Director Div. of Quantitative Sci., Director Women's Hlth.Res., Dept. of Obstetrics and Gynecology Dept. of BioMed. Informatics, Nashville, TN, [8]VUMC,Dept. of Med., Div. of Epidemiology Human, Genetics Graduate Program, Epidemiology Graduate Program, Inst. for Med. and Publ. Hlth., Nashville, TN

Abstract:

Fibroproliferative diseases (FPDs), such as hypertension and asthma, share a common etiology and are more prevalent among individuals of African ancestry (AA), suggesting shared genetic risk factors. Racial disparities reveal a higher prevalence among AA individuals, possibly due to a Th2-favored genome from generational exposure to helminth parasites in Africa. To investigate this hypothesis, we conducted a meta-analysis using BioVU, UK Biobank, and eMERGE datasets across seven FPDs. NHB and NHW meta-analyzed summary statistics were used in a Multi-trait GWAS analysis (MTAG) to identify significant SNPs across populations. MTAG analysis identified

significant SNPs for asthma, hypertension, fibroids, and other FPDs. For asthma, top SNPs included rs200363056 *(ASAH2C*, $p=6.27x10^{-23}$) for NHB and rs12290395 (*GSDMB,* $p=3.61x10^{-10}$) in NHW. *GSDMB* (rs12290395, $p=3.61x10^{-10}$) was also identified in the hypertension NHB population along with SNPs near *FBXO42* (rs12091661, $p = 2.79x10^{-13}$) in NHB and *PRDM8* (rs11099097, *MTAG_p=*$8.13x10^{-33}$) in NHW. In fibroids, the top significant SNPs were rs10713830 (*GPHN,* $p=3.79x10^{-18}$) in NHB and rs78378222 (*TP53,* $p=1.14x10^{-25}$) in NHW. GPHN also showed a strong association with keloids (rs10713830, *MTAG_p =* $9.23x10^{-18}$) and fibroids (rs10713830, p=$3.79x10^{-18}$) NHBs, indicating a shared genetic risk factor. For glaucoma, significant SNPs included rs569061728 (*ACOT12,* $p= 1.65x10^{12}$) in NHB and rs565631101 (*RP11-457K10.1,* $p= 2.31x10^{-45}$) in NHW. Our findings shed light on genetic loci contributing to FPD susceptibility and unveil several novel SNPs across different phenotypes. These discoveries hold immense potential for the future of fibroproliferative disease research. Notably, many of these significant genes, such as *ASAH2C, TP53, PRDM8*, and *GSDMB*, are key players in immune system regulation. For instance, *ASAH2C* contributes to an immunosuppressive environment by down-regulating T-cell responses. *PRDM8* is highly expressed in regulatory and memory T cells, impacting T-cell function. *GSDMB* contributes to fibrosis by promoting pyroptosis, leading to inflammation and subsequent fibrotic processes. *FBXO42* is a known regulator of the *NOTCH* family of proteins, particularly *NOTCH4*, a gene highly associated with glaucoma, indicating another shared genetic locus between fibroproliferative diseases. These insights support the hypothesis that a robust Th2 immune response, historically favored in African ancestry, contributes to FPD susceptibility. This study provides direction for understanding FPDs in NHB populations and establishes a foundation for further functional studies and potential therapeutic targets

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4095W: Multi-trait polygenic risk score and all-cause and cardiovascular disease mortality -A cohort study of 14,086 Japanese individuals-

Authors:

R. Fujii[1,2], N. Mako[2], N. Masahiro[2], S. Shuntaro[3], T. Yoshiki[1], S. Koji[1], W. Kenji[2], M. Keitaro[4,2]; [1]Fujita Hlth.Univ., Toyoake, Japan, [2]Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [3]Nagasaki Univ. Hosp., Nagasaki, Japan, [4]Aichi Cancer Ctr. Res. Inst., Nagoya, Japan

Abstract:

Backgrounds: Although the utility of composite trait-specific polygenic risk score (multi-trait PRS) has been examined among European ancestries, few studies investigated among East Asians. We examined the associations of six trait-specific PRSs and multi-trait PRS for cardiometabolic factors with all-cause and cardiovascular disease (CVD) mortality.**Methods:** The Japan Multi-Institutional Collaborative Cohort (J-MICC) Study recruited 14,086 Japanese adults (Mean age: 55±9; 55.8% women) between 2005 and 2014. We calculated six PRS for cardiometabolic traits (systolic blood pressure, diastolic blood pressure, body mass index, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and estimated glomerular filtration rate) by using the PRSice-2 method. Based on six trait-specific PRSs, we developed multi-trait PRS and considered as the primary exposure. All-cause and CVD mortality (ICD 10: I00-I99) were acquired under the permission from the Ministry of Health, Labor, and Welfare.**Results:** During a median 12.1-year follow-up period, a total of 472 all-cause and 79 CVD mortality cases were documented. For trait-specific PRSs, BMI PRS was associated with all-cause mortality, while SBP and DBP PRS was associated with CVD mortality. Compared with 0-90% of multi-trait PRS, adjusted HRs (95% CI) among the top 10% of multi-trait PRS were 1.05 (0.79, 1.52) for all-cause mortality and 2.10 (1.18, 3.57) for CVD mortality. Participants with three or more trait-specific PRS above 90% had a HR (95% CI) for CVD mortality of 4.88 (1.66, 14.39) compared to that of all trait-specific PRS below 90%. Estimated years of life lost by CVD mortality in the top 10% multi-trait PRS was 7.4 (1.7, 12.7) years compared with the rest.**Conclusions:** Our cohort study found that multi-trait PRS for cardiovascular-related factors may contribute to identify the high-risk populations for CVD mortality among Japanese adults. Further studies are expected to examine the associations of CVD subtypes with multi-trait PRS and to add non-genetic determinants into genetic predisposition.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4096W: Multivariate SEM-PGS Model: Using Polygenic Scores to Investigate Cross-Trait Indirect Genetic Effects

Authors:

X. Lyu[1], J. Balbona[2], T. Chen[3], Y. Kim[1], M. Keller[4]; [1]Univ. of Colorado at Boulder, Boulder, CO, [2]Washington Univ. in St. Louis, St. Louis, MO, [3]Penn State Univ., University Park, PA, [4]Univ Colorado, Boulder, Boulder, CO

Abstract:

Recent research has uncovered indirect genetic effects including genetic nurture and assortative mating across a wide range of human behaviors, primarily observed and analyzed within the

same trait between parents and offspring. Importantly, indirect genetic effects also occur across different traits as a consequence of cross-trait vertical transmission. In this study, we proposed the multivariate SEM-PGS model, an extension of the previously proposed SEM-PGS model to further account for familial similarities. Using polygenic scores (PGS) in a structural equation model of nuclear family members, the multivariate SEM-PGS model provides a novel tool for gaining more in-depth insights into the complexity of gene-environment interplay in human development. We formulated and validated mathematical expectations for the estimated parameters in matrix form, including estimates for direct genetic effects, as well as same-trait and cross-trait genetic nurture, vertical transmission, and assortative mating effects. Nine conditions of forward-time Monte-Carlo simulations were performed to evaluate the precision of parameter estimation and bias under various sample sizes, different levels of genetic variance explained by PGSs (r^2_{PGS}), assortative mating effects (μ) and vertical transmission effects (f). By fitting the model with simulated data in the OpenMx software, we demonstrated that the cross-trait estimates were unbiased when model assumptions were met. The estimates showed a reasonable standard error (SE) with more than 36,000 nuclear families under conditions comparable to empirical analysis. Moreover, the SEs decreased with well-established PGSs, and larger effect sizes of both within-trait and cross-trait f and μ. Additionally, we showed that same-trait direct and indirect genetic effects are prone to overestimation when cross-trait effects exist but not incorporated in the model. We also discussed promising model modifications to capture a broader spectrum of familial gene-environment interplay using PGSs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4097W: Network biology algorithms reveal biological pathways underlying cigarette smoking behaviors

Authors:

G. Keele[1], K. Sullivan[2], A. Townsend[2], J. Miller[2], D. Kainer[2], M. Lane[2], M. Cashman[2], M. Garvin[2], P. Kruse[2], B. Quach[3], C. Willis[4], K. Xu[5], B. Aouizerat[6], E. Johnson[7], D. Hancock[3], D. Jacobson[2]; [1]RTI Intl., Research Triangle Park, NC, [2]Oak Ridge Natl. Lab., Oak Ridge, TN, [3]RTI Intl., Durham, NC, [4]RTI Intl., Chapel Hill, NC, [5]Yale Sch. Med., New Haven, CT, [6]Univ California San Francisco, San Francisco, CA, [7]RTI Intl., Duluth, MN

Abstract:

Genome-wide association studies (GWAS) from the GWAS & Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) have identified hundreds of genome-wide significant loci contributing to the heritability of cigarette smoking behaviors. However, for most variants, their target genes,

pathways, and neurobiological relevance are poorly understood. To better understand the biological pathways underlying smoking behaviors, we applied two network biology algorithms (Gene set Refinement through Interacting Networks [GRIN] and Multiplex Embedding of Networks for Team-based Omics Research [MENTOR]) to GSCAN GWAS summary statistics from European ancestry subjects for Smoking Cessation (SmkCess; 388,313 individuals) and Cigarettes Per Day (CigsPerDay; 326,497 individuals). We assigned genes from genome-wide significant ($p < 5e^{-8}$) single nucleotide polymorphisms (SNPs) using multiple methods, including nearest gene, MAGMA, and H-MAGMA with Hi-C data from multiple brain regions. We first built a multiplex network of 10 gene-gene layers from distinct types of biological experimental evidence (including two human brain prefrontal cortex-specific layers) and then used GRIN to filter out likely false positive genes based on the premise that correctly assigned genes would be highly interconnected within the networks. We assigned 526 unique genes targeted by SNPs associated with CigsPerDay, of which 235 genes were retained by GRIN. We assigned 310 unique genes targeted by SmkCess-associated SNPs, of which 149 genes were retained by GRIN. We then used MENTOR with our 10-layer multiplex network to identify functional groupings of smoking-associated genes, followed by gene set enrichment of biological pathways. Of the GRIN-retained genes 46 were common to both phenotypes. MENTOR-identified clades of highly interconnected genes suggesting several mechanisms shared across phenotypes, including the expected nicotinic receptor function as well as neuronal development and differentiation functions. Drug database analyses highlight a number candidate compounds that target biology outside the nicotinic receptor genes, suggesting the potential for novel smoking cessation treatments. These systems biology approaches linked related but unique GWAS-based genetic architectures to shared biological mechanisms, demonstrating how network integration of biological data with GWAS can reveal plausible mechanisms underlying complex phenotypes and behaviors.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4098W: New-Onset Refractory Status Epilepticus (NORSE) in Individuals with POLG Gene Variants: A Systematic Review

Authors:

M. Kellogg[1], A. Oliger[1], I. Bushlin[2], M. White[2], L. Behrndt[3]; [1]Portland VA HCS & Oregon Hlth.& Sci. Univ. (OHSU), Portland, OR, [2]Oregon Hlth.& Sci. Univ. (OHSU), Portland, OR, [3]Oregon Hlth.& Sci. Univ., Portland, OR

Abstract:

Rationale:
NORSE is defined as de novo refractory status epilepticus (RSE) in someone without a prior history of seizures and without an acutely identifiable cause; FIRES is a subset of NORSE preceded by a febrile illness. Heterogenous autoimmune, post-infectious, and/or genetic factors have been implicated in cryptogenic NORSE pathogenesis, including POLG gene variants. The POLG gene plays a role in the replication and repair of mitochondrial DNA, and mutations in POLG are associated with intractable epilepsy and RSE. This systematic review sought to determine the spectrum of 1) age of onset of POLG-related NORSE/FIRES; 2) specific POLG mutations associated with NORSE/FIRES; and 3) suspected triggers and prognosis of POLG-related NORSE/FIRES, particularly in adult-onset cases.
Methods:
We performed a systematic review of peer-reviewed studies that described individual cases of children or adults with RSE, NORSE or FIRES found to have identifiable POLG gene variants. Exclusion criteria included absence of individual case data, age of onset of seizures or RSE was unclear, or a specific POLG gene variant was not specified.
Results:
From a total of 166 screened articles published through April 2024, 61 met eligibility criteria for data extraction. Out of 97 cases of RSE, 65 (67%) met criteria for NORSE, of which 19 (29%) met criteria for FIRES and 13 cases (20%) were adult-onset (age>=18). Median age of onset for NORSE was 7yrs (range 7 mos-31yrs). The most common variants detected included: pAla467Thr (31 cases, 48%), pTrp748Ser (14 cases, 22%), and pGlu848Glu (13 cases, 20%). 34 cases (52%) were compound heterozygous, and 18 (28%) were homozygous for POLG variants. Familial inheritance was confirmed with genetic testing in 32 (49%) cases; only 1 POLG mutation was explicitly described as de novo and 2 were described as VUS. Mortality was seen in 54% of cases. Mean timing of mortality after onset was 9 months (SD 0.78).
Conclusion:
This systematic review demonstrates that POLG mutations are implicated in late-onset NORSE. About half the cases of NORSE occurred in individuals aged 10 years and older, and 20% occurred in adults (aged 18+). However, while genetic testing is routine in the pediatric evaluation of NORSE, it is not yet standard of care for adult-onset NORSE. This study's findings suggest the need for genetic testing for NORSE at all ages. More research is needed to determine the specific triggers for POLG-related NORSE and optimal treatment.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4099W: No evidence of synthetic association across multiple complex traits

Authors:

P. Romero[1], M. C. Keller[2]; [1]Univ. of Colorado Boulder, Boulder, CO, [2]Univ Colorado, Boulder, Boulder, CO

Abstract:

Genome-wide association studies (GWAS) were developed to investigate the association between single nucleotide polymorphisms (SNPs) and a trait of interest. For decades, GWAS have uncovered disease-associated loci and helped us understand the genetic architecture of complex traits. However, Goldstein warned of the possibility that rarer variation not tagged by GWAS might explain common variant GWAS findings. Goldstein described this phenomenon as 'synthetic association'. Many disagreed, arguing that synthetic association is an insufficient explanation for almost all GWAS results. Here, we present the first empirical examination of synthetic association across different complex traits, including height, body mass index, diabetes, and cigarettes per day. We did not find evidence of synthetic association across any of our traits; however, we did find substantial evidence of fine-scale population stratification. In short, we echo Wray et al.'s claim that synthetic association is not a concern in GWAS and highlight the need to develop methods that properly control for population stratification reflective of recent demographic history.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4100W: Novel genome-exposome integration tool identifies gene-environment interactions for autism spectrum disorder and attention deficit hyperactivity disorder

Authors:

N. Palmiero[1], A. Rico[1], J. Zhou[2], M. D. Ritchie[3], H. Volk[4], C. Newschaffer[5], M. A. Hall[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Stanford Univ., Palo Alto, CA, [3]Univ. OF PENNSYLVANIA, Philadelphia, PA, [4]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [5]Pennsylvania State Univ., University Park, PA

Abstract:

As genome- and exposome-wide data have become available for increasing numbers of samples, gene-environment interaction (GxE) tools need to limit computational and multiple testing burden and avoid testing all combinations of exposures and single nucleotide polymorphisms (SNPs). While effective, the major drawback to EWAS- and GWAS-filtered GxE (i.e., main effect filtering) are the chosen variables are limited to those that have a demonstrated main effect on the phenotype. Knowledge-based filtering has been used as a creative alternative to main effect filtering for gene-gene interactions; however, these approaches have focused on relationships between genes. Therefore, there is need for an integrated software to perform knowledge-driven filtering and analysis for GxE. We developed the Integrative Genome-Exposome Method (IGEM) to address this need. Knowledge of the biological relationships (e.g., pathway, biological interactions, associations) between environmental exposures (including toxins, exogenous small molecules, nutrients, drugs, pollutants) and genes are housed across a number of open-source database resources including Human Metabolome Database (HMDB), Comparative Toxicogenomics Database (CTD), DigChem, BioSNAP, KEGG, and PharmGKB. IGEM integrates the knowledge across these databases to create putative, biologically related gene-environment models for interaction analysis, thereby reducing the computational and multiple testing burden associated with GxE. Besides filtering, IGEM facilitates exposome data quality control, statistical interaction analysis, and visualization. We used IGEM to evaluate the role of GxE in autism spectrum disorder (ASD; n=10,599) and attention deficit hyperactivity disorder (ADHD; n=17,425) in the Environmental Influences on Child Health Outcomes (ECHO) cohort. IGEM identified 2 and 9 GxE models significantly associated with ASD and ADHD, respectively (Bonferroni-adjusted $p<0.05$). For ASD the top result was for an interaction between a SNP near gene *pleckstrin homology domain-containing family G member 2 (PLEKHG2)* and child exposure to any secondhand e-cigarettes (Bonferroni-adjusted likelihood ratio test p-value [LRT-p]: 0.0185). The top result for ADHD involved a SNP in the gene *low-density lipoprotein receptor-related protein 12 (LRP12)* and exposure of child dietary potassium mg/day intake (Bonferroni-adjusted LRT-p: 9.69x10e-4). Each child exposed to the environmental factor was above the age of 1. These results demonstrate the importance of evaluating GxE for neurodevelopmental phenotypes using knowledge-driven filtering and analysis tools like IGEM.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4101W: Novel pipeline to genotype trisomic loci for the study of congenital heart defects in Down syndrome

Authors:

E. Feldman[1], D. J. Cutler[1], T. C. Rosser[1], S. B. Wechsler[1], L. Sanclemente[2], A. Rachubinski[3], M. Galbraith[3], K. R. Rabin[2], M. Wagner[4], B. D. Gelb[5], J. M. Espinosa[3], P. J. Lupo[2], S. L. Sherman[1], E. J. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Baylor Coll. of Med., Houston, TX, [3]Crnic Inst. at Univ. of Colorado Anschutz, Aurora, CO, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Congenital heart defects (CHDs) are the most common structural birth defect and are present in 40-50% of children with Down syndrome (DS). While increased gene dosage on chromosome 21 contributes to the risk for abnormal heart development, about 50% of infants with DS have structurally normal hearts (NH). Thus, additional modifying genetic or environmental factors may exist. To characterize the genetic architecture of various DS-associated CHDs, we sequenced genomes of a multiethnic group of children with DS and a CHD (n=886) or DS and a NH (n=572) as part of the Gabriella Miller Kids First Pediatric Research Program and the INCLUDE (INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndromE) Project. While our previous studies suggest that there is a role for common variants and rare truncating variants which interact closely with genes found on chromosome 21, they do not completely explain CHD susceptibility in DS. However, variants on chromosome 21 remain uninvestigated in this dataset due to their triploid nature. In DS, there are four possible genotypes at chromosome 21 loci: AAA, AAa, Aaa, and aaa. Since each chromosome may be sequenced at a different rate, we expect that true heterozygous calls could correspond to low allele balances and be inaccurately called as homozygotes based on the standard pipelines. Therefore, for trisomic loci, re-genotyping with respect to allele balance must be considered. To call genotypes on chromosome 21 in DS, we applied Bayes' Theorem with a uniform prior, which was then used to calculate the posterior probability for each of the four possible genotypes based on the number of sequencing reads representing the reference and alternate alleles. Then we selected the genotype with the greatest calculated probability. Finally, in order to prepare an analytic dataset, we only kept genotypes with a probability greater than 90% and removed variants and samples with high missingness. Analyses are underway to compare our pipeline to other existing methods for calling triploid genotypes, such as GATK haplotype caller with triploid mode, and to study common and rare variants on chromosome 21 associated with CHD risk in children with DS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4103W: Outlier rare-variants implicate CTCF-bound promotor like sequences upstream of core genes in children with high risk for obesity

Authors:

C. LeMaster[1], C. Schwendinger-Schreck[1], B. Ge[2], W. Cheung[3], J. Johnston[3], C. Smail[3]; [1]Children's Mercy Kansas City, Kansas City, MO, [2]McGill & GQ Innovation Ctr., Montreal, QC, Canada, [3]Children's Mercy Kansas City, Kansas City, MO

Abstract:

Rare variants associated with gene expression outliers occur in less than 1% of the population but often exhibit higher pathogenetic or functional effects than common variants. Recent studies have shown that these rare outlier variants (ROVs) can be implicated in conditions such as obesity, even in individuals with a low polygenic risk score (PRS). We investigated how these ROVs impact cis-regulatory elements upstream of genes linked to the PRS for BMI in children from the Genomic Answers for Kids (GA4K) study. Obese individuals (BMI > 30, n=57) had a mean PRS Z-score of 0.56, while non-obese individuals (BMI < 30, n=482) had a mean Z-score of -0.01 ($p < 0.0001$). In obese individuals, 44% had a z-score < 0.5, and in non-obese individuals, 29% had a z-score > 0.5. These outliers had fewer upstream ROVs on average (obese mean=27, non-obese mean=27) compared to inliers (obese mean=31, non-obese mean=29). Evaluation of ROVs for overlap with candidate cis-regulatory elements (cCREs) revealed that all participants had at least one overlap with a cCRE. Overlaps with CTCF-bound promoter-like signatures (cctcf-PLS) were found in 63% of the cohort, CTCF-bound distal enhancer-like signatures (ctcf-dELS) in 70%, CTCF-bound proximal ELS (ctcf-pELS) in 74%, pELS in 56%, and dELS in 73%. Other cCREs were overlapped by less than 10% of the cohort. We also assessed the weight of the gene downstream of the ROV-cCRE, the Combined Annotation Dependent Depletion (CADD) score, and the gene expression (GE) z-score. dELS overlaps were associated with higher weight genes (mean weight=7.65x10^-5), ctcf-dELS had the highest GE outcomes (mean GE z-score=0.57), ctcf-PLS had the lowest GE outcomes (mean GE z-score=0.002), and ctcf-PLS had the highest CADD score (CADD=10.12). In 16 obese outliers with ctcf-PLS overlap, the average CADD was 8.7 and GE was 0.07. In non-obese outliers, 19 overlaps were repeated with an average CADD of 10.5 and GE of -0.16. Our data suggests that in non-obese individuals with a high BMI PRS, their BMI could be explained by ROVs upstream of core genes at regulatory regions like CTCF-bound PLS, driving reduced expression.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4104W: PanTrio: Bespoke pangenome reference enables highly accurate genotyping of proband-parent trio sequencing samples

Authors:

S. Demirkaya-Budak[1], H. Feng[2], G. Budak[1], **D. Freed**[2], D. Turgut[1], A. Jain[1]; [1]Velsera, Boston, MA, [2]Sentieon Inc., San Jose, CA

Abstract:

Accurate genotyping of proband-parent trio sequencing samples is essential for identifying common genetic causes of early onset congenital disorders: *de novo* mutations (DNMs) and loss of heterozygosity (LoH). However, the standard methods that individually genotype each sequencing sample and then compare the alleles at corresponding loci have low precision and recall as they are confounded by sequencing errors and uneven coverage in many sections of the genome, especially centromeres and allosomes. We present a new paradigm for accurately analysing family trio samples, by first collating individual variant calls to construct a bespoke family pangenome, remapping the sequencing samples against this pangenome reference, using Velsera-GRAF aligner, to boost the recall of variants in low coverage regions, whilst preserving high precision by using a state-of-the-art variant caller, Sentieon-DNAscope, incorporating a machine learning based filter, and finally annotating DNMs and LoH by comparing offspring haplotypes with each parent respectively. This workflow resulted in significant increase in genotyping accuracy when tested on the two families included in the Genome in a Bottle benchmark set, comparing favourably against DeepTrio and GATK. The computational cost of our workflow was comparable to individual genotyping with GATK, and 10% lower than family genotyping with DeepTrio. We also assessed the DNM identification accuracy by analysing 10 family trios where a total of 766 DNMs have been previous identified and validated via manual inspection and Sanger sequencing. Our method identified 711 true DNMs which is significantly more than DeepTrio (654) and GATK (628), as well as a dedicated DNM discovery workflow (665) that implements a strategy of identifying consensus among multiple variant-calling workflows, incurring twofold increase in computational cost. Our method called a total of 881 Mendelian inconsistent genotypes in functionally relevant genomic regions across all samples, demonstrating significantly better precision than DeepTrio (2063 inconsistent calls) and GATK (1347 inconsistent calls), and approaching the state-of-the-art precision afforded by the consensus DNM workflow (671 inconsistent calls). We have implemented this method as an automated bioinformatics workflow suitable for large-scale analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4105W: Pathway-level network-informed transcriptome-wide association study leveraging prior gene regulatory network information

Authors:

D. Lee, J. Jin; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Transcriptome-wide association studies (TWAS) test for associations between genetically regulated portion of gene expression (GReX) and traits of interest. The standard TWAS pipeline consists of 2 steps: first, GReX model is trained on a reference dataset with both genotype and gene expression levels available, and then is used to impute gene expression in another dataset, where associations between imputed GReX and traits are tested. Typically, a GReX model is trained using *cis*-single nucleotide polymorphisms (SNPs) for each single gene, which restricts the maximum variability that can be explained by the GReX model to the *cis*-SNP heritability. To overcome this limitation, we propose a network-informed TWAS (NTWAS), which leverages external gene regulatory network (GRN) information to jointly model all genes in a gene pathway. Information of GRN, which could be represented by a graph describing the regulatory interactions among genes and gene products, can be extracted from public gene pathway databases. For each pathway, we use a structural equation model to construct a GReX model for each gene, where the predictors consist of the associated genes in the same pathway and the *cis*-SNPs of the target gene. The elastic net is applied only to the *cis*-SNPs to induce sparsity in the identified *cis*-eQTLs. We evaluated the performance of the NTWAS GReX model using the Religious Orders Study and Rush Memory and Aging Project (ROSMAP) data, in which 586 individuals had both genotype and anterior cingulate cortex tissue gene expression data. We trained NTWAS and FUSION GReX models on 466 individuals and then calculated R-squared (R^2) values between measured and imputed gene expression data in the remaining 120 individuals. The GRN information was extracted from the Kyoto Encyclopedia of Genes and Genomes. Among 334 genes with nominally significant ($p<0.05$) NTWAS and FUSION model R^2, 244 genes had greater median R^2 in the NTWAS model. At Bonferroni significance threshold, 54 out of 99 genes had greater median R^2. Through the incorporation of GRN, NTWAS identified *trans*-eQTLs for each gene that are *cis*-eQTLs of other genes in the same pathway that contribute to the expression of the gene of interest and can explain greater variability of GReX than *cis*-eQTL-based TWAS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4106W: PhenoLink: Using a fine-tuned large language model to extract HPO terms from natural language texts

Authors:

J. Cai[1], Z. Chen[2], G. Li[1], J. Yang[1], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Nanfang Hosp., Southern Med. Univ., Guangzhou, China

Abstract:

The Human Phenotype Ontology (HPO) project provides standardized terms and codes for phenotypic abnormalities, increasingly utilized in clinical software tools, disease databases, and phenome research. In genetic disease diagnosis, many phenotype-driven differential diagnosis, genomic diagnosis, and translational research software tools require the input of HPO terms. The hierarchical structure of HPO terms enables the computational analysis of phenotypes, establishing HPO terms as the phenotypic standard for phenome-wide association studies (PheWAS) and disease-phenotype databases. However, since phenotypic information is often recorded in natural language, converting it to HPO terms demands additional time and familiarity with HPO terminology. Current phenotype extraction tools are rule-based and struggle to handle complex contexts. Large language models, with their semantic understanding capabilities, excel in managing complex contexts, making them a promising approach for improving phenotype extraction tools. Here, we introduce PhenoLink, a phenotype extractor based on a fine-tuned large language model capable of extracting phenotypes from natural language texts and converting them into HPO terms. PhenoLink extracts phenotypes from input texts using a fine-tuned large language model, subsequently outputting HPO terms most semantically similar to each phenotype. To achieve this, we constructed a reference dataset (n=102,461) comprising HPO names, definitions, and synonyms. To further enhance the performance of PhenoLink, we utilized ChatGPT 4.0 to paraphrase the names and definitions of 13,000 HPO terms and incorporated the paraphrased content into the HPO reference dataset. Finally, we evaluated the performance of PhenoLink using 403 independent phenotypes manually annotated by clinicians proficient in HPO terminology.PhenoLink achieves an extraction accuracy of 82.05% for 403 independent phenotypes, surpassing the HPO official Application Programming Interface (43.42%), ClinPhen (67.24%), and ChatGPT 4.0-turbo (43.18%). Using the ability to rank the correct HPO term within the top 10 as a performance metric, PhenoLink improves accuracy by 14.87%-34.72% compared to these tools. In summary, PhenoLink can accurately and swiftly extract phenotypes from natural language texts and match them to HPO terms, which can be used for genetic disease diagnosis, phenotype research, and the construction of disease-phenotype databases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4107W: Polygenic risk and progression prediction for longitudinal and familial data in large biobanks

Authors:

A. Abdel-Azim[1,2], T. Chen[3], H. Li[3], H. Zhang[4], X. Lin[3,1]; [1]Dept. of Statistics, Harvard Univ., Cambridge, MA, [2]Dept. of Molecular and Cellular Biology, Harvard Univ., Cambridge, MA, [3]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD

Abstract:

Background: Polygenic risk prediction facilitates precision medicine, enabling earlier diagnoses, refined patient stratification, and genetically informed treatment plans. As large biobanks emerge globally, such as the UK biobank (UKB), various risk prediction frameworks have been developed, aggregating an individual's genetic insights. However, the application of these frameworks in large biobanks and their integration into clinical practice is hindered by several challenges. State-of-the-art risk prediction frameworks assume individual records are independent and do not directly account for genetic similarity between relatives or longitudinal measurements, resulting in substantial loss of information and accuracy.
Methods: We introduce and validate a statistically rigorous risk prediction framework to solve this open problem, leveraging mixed-effects models to adeptly handle correlated data. We evaluated the performance of this framework on real and simulated data in two contexts: related subjects and longitudinal observations. In related cohorts, additive genetic effects (SNP-based) are disentangled from genetic effects due to relatedness, thereby mitigating their confounding influences. With longitudinal observations, we disentangle a more accurate genetic risk value and an individual-level progression assessment. To our knowledge, there are no risk prediction frameworks that have been introduced for longitudinal observations.
Results: In several realistic simulations, with different sample sizes, genetic architectures, and nested cohort structures, our methods which account for relatedness and longitudinal data outperform standard polygenic risk score (PRS) approaches and conventional mixed-effects models. For longitudinal data, in simulation, we more accurately estimate a genetic risk value and forecast polygenic longitudinal profiles, for example showing 15% improvement over conventional mixed models in forecasting accuracy for individuals with four longitudinal records. We applied our models to blood biomarker traits in the UKB, achieving both more accurate individual-level risk assessments and progression insights. Our methods are implemented in Julia and C and leverage efficient iterative strategies, leading to several-fold reductions in memory and time.

Conclusions: We offer a valuable framework for calculating improved PRSs in large biobanks by effectively handling correlated data structures due to family relatedness and longitudinal repeated measures. These methods lead to more precise genetic risk predictions for diseases and progression as well as clinical translation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4108W: Polygenic risk for poor lung function and its association with COVID-19 non-recovery outcomes

Authors:

A. Goemans[1], O. C. Leavy[1,2], B. Guillen-Guio[1,2], E. Pairo-Castineira[3], K. Rawlik[3], K. Baillie[3], A. SIngapuri[2], R. A. Evans[2], C. E. Brightling[2], L. V. Wain[1,2], T. C. D. Lucas[1], PHOSP-COVID Collaborative Group; [1]Dept. of Population Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom, [2]The Inst. for Lung Hlth., NIHR Leicester BioMed. Res. Ctr.-Respiratory, Univ. of Leicester, Leicester, United Kingdom, [3]Roslin Inst., Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract:

Background: Individuals with COVID-19 can experience protracted recovery with a wide range of symptoms affecting multiple organ systems. The biological mechanisms remain unclear and predicting those at highest risk of ongoing symptoms and impact on quality of life after COVID-19 remains challenging. We hypothesised that genetically predicted specific organ impairment might be associated with specific long term recovery outcomes irrespective of pre-existing comorbidities.
Aims: As SARS-CoV-2 infection initiates as a respiratory infection, we hypothesised that the relationship between polygenic risk for poor lung function would be more strongly correlated with persistence of respiratory-related symptoms than with overall non-recovery. We used polygenic risk scores (PRS) to determine if genetic risk for poor lung function was associated with non-recovery outcomes post COVID-19 in hospitalised patients.
Methods: Using the Post-HOSPitalisation COVID-19 study, we calculated genome-wide PRS for lung function using publicly available GWAS data for obstructive (FEV1/FVC), restrictive (FVC) and fibrotic (IPF) traits. PRS were optimised for patient reported recovery outcomes - overall recovery and respiratory-related symptoms - at 5 and 12 months. We tested for associations between these PRS and recovery outcomes, adjusting for pre-existing respiratory comorbidities.
Results: A PRS of FVC was not associated with non-recovery at 12 months, but association with breathlessness was nominally significant (OR= 0.82; 95% CI=0.71 - 0.96; P=0.0135; N variants=472910). At 5 months, a PRS for pulmonary fibrosis was associated with non-recovery

(OR= 1.15; 95% CI=1.01 - 1.31; P=0.0359; N variants=5653), and increased self-reported cough (OR= 1.16; 95% CI=1.02 - 1.33; P=0.0235; N variants=2464), but reduced breathlessness (OR= 0.82; 95% CI=0.68 - 0.97; P=0.0266; N variants=63086).
Conclusions: Our results suggest that there is some genetic overlap between lung traits and respiratory-related symptoms in patients hospitalised for COVID-19 irrespective of baseline comorbidities. These associations should be prioritised for further enrichment to identify pathways important in non-recovery in the subgroup of patients with high genetic risk for poor lung function.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4109W: Pooled calling on rare noncoding variants in samples with low pass whole genome sequencing

Authors:

R. Zhang[1,2], T. Ge[1], H. Huang[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst., Cambridge, MA

Abstract:

Genome-wide association studies (GWASs) have identified thousands of loci associated with diseases or traits. However, previous studies mainly focus on coding variants or common variants, leaving non-coding rare variants largely unexplored. One major hurdle has been the prohibitive cost of capturing rare variants in non-coding regions, especially for large cohorts at the population-level. A recently developed sequencing technology, Blended Genome Exome (BGE), produces high-coverage whole exome (30-40X) and low-coverage whole genome (2-3X) data in a single CRAM file. This cost-efficient and ascertainment-unbiased technology makes it feasible to analyze non-coding rare variants. However, there is a pressing need for an accurate and robust variant calling method for low-coverage rare variants.
In this study, we propose a pooled calling method to estimate the allele frequency of low-coverage rare variants, with a goal to enable efficient allele frequency comparison between cases and controls. Specifically, for each variant, we estimate the posterior probability of each candidate genotype for each individual and organize these probabilities into a matrix encompassing all individuals. Using this matrix, we calculate the probability of each allele frequency for the population, assigning the allele frequency with highest probability to the variant. We compared our method with the state-of-the-art variant calling method, GATK HaplotypeCaller. Utilizing 85 individuals from Asian Bipolar Genetics Network (A-BIG-NET) study with high-coverage whole-genome sequencing (30X) as the ground truth, we used GATK

DownsampleSam to downsample the original data to 20X, 10X, 8X, 6X, 4X, 2X, 1X, and 0.5X respectively. We then assessed the concordance of the allele frequency of rare variants between high-coverage and low-coverage data for each method. We are currently in the process of further developing and evaluating this method.
Non-coding rare variants constitute the majority of the genome. This method can provide accurate estimation of the allele frequencies of low-coverage rare variants within non-coding regions, with the potential to offer new insights into the genetic basis of complex disorders and diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4110W: Predicting Polygenic Score Accuracy with Neighbors in the Genetic Ancestry Continuum

Authors:

K. Feng, L. Kachuri, J. S. Witte; Stanford Univ., Stanford, CA

Abstract:

Background: Polygenic scores (PRS) have poor predictive accuracy when applied to individuals from ancestries that are underrepresented in the model's training set. Recent work has demonstrated that an individual's Euclidean distance to the center of the training set in principal component (PC) space is inversely correlated with model accuracy. However, modeling PRS accuracy across PC space is difficult in practice, as most PRS models are trained on meta-analysis of summary statistics.
Methods: We propose a non-parametric approach to estimating PRS accuracy in a test population that uses an external calibration population, bypassing the need for individual-level PRS training data. Our approach leverages the intuition that calibration samples with high genetic similarity to a test sample (i.e. close in PC space) experience similar model accuracy. We train genome-wide PRS-CSx models for multiple traits using GWAS data from the Million Veteran Program (n=322K), and evaluate them in the All Of Us (n=235K) and UK Biobank (n=429K) cohorts using a *k*-nearest neighbors kernel.
Results: We demonstrate that genetic neighbors can reliably predict PRS accuracy for a given test sample across multiple traits. For height, our model exhibits varying test accuracy across population clusters defined by genetic similarity to major continental ancestry groups. In both test datasets, accuracy was highest for European (incremental R^2=0.092-0.099), lower for East Asian (0.056-0.059) and South Asian (0.052-0.063), and lowest for African (0.050-0.051). Calibration neighborhood R^2 is strongly correlated with test sample accuracy. The strength of

this correlation depends on the size and density of the neighborhood. In All of Us, varying the *k*-nearest neighbors across 10 PCs gave R^2=0.85 (when *k*=100), R^2=0.98 (*k*=1000), and R^2=0.99 (*k*=10,000); in UK Biobank R^2=0.43 (*k*=100), R^2=0.74 (*k*=1000), and R^2=0.99 (*k*=10,000). **Conclusion:** Our method allows PRS accuracy to be assessed at the individual level using a continuous representation of genetic ancestry, which is especially useful for admixed individuals. In future work, we will demonstrate how this can be used to select the best model for a given individual.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4112W: Prioritization of mouse strains for modeling genetic variant effects in rare and common disease

Authors:

R. Ball, R. Duggan, M. Gerring, A. Berger, M. Berry, G. Mukherjee, **E. Chesler**; The Jackson Lab., Bar Harbor, ME

Abstract:

Prioritization of mouse models for rare and common disease research is critical to identify treatments and interventions that benefit human health, yet it is often a time-consuming process that requires expert knowledge of both the mechanism of disease and the phenotypic profile of mouse backgrounds. It is particularly challenging in rare disease research where the precise mechanism of disease is generally less understood, often leading to an extended delay in identifying an appropriate mouse model. In most cases, mouse models are created by modelling human genetic variants through genome editing, choosing from a small set that are convenient but may not be vulnerable to the manifestation of disease-related phenotypes. These strains may not be vulnerable even when the condition is thought to be caused by a single gene perturbation (i.e., rare, Mendelian disease variants) because the perturbation functions within a biological network. For some strains, the biological networks are robust and can compensate for the effect of deleterious alleles while other strains harbor molecular networks that are more vulnerable to deleterious variation, resulting in variable penetrance and heterogeneity in the impact of disease mutations on different strain backgrounds. We posit that strains that show more mild or more severe patterns of dysfunction are detectable through the manifestation of similarity to disease-associated phenotypic or transcriptional states. We developed a data-driven approach that prioritizes mouse strains by computing transcriptional and phenotypic similarity to the disease state. The tool prioritizes among 657 inbred mouse strains from a variety of collections including common inbreds, recombinant inbreds,

Collaborative Cross and wild-derived inbred mouse strains. Strains are prioritized based on their resemblance to genomic or phenomic representations of disease, annotated with disease vulnerability scores and associated confidence estimates along with breeding/engineering characteristics of interest to mouse researchers. A saturated and uniform representation of the genomic state of these strains across tissues and cell types is achieved with a machine learning (ML) model. We present the application of our software pipeline for prioritization of these strains, including validation of the approach through both prospective and retrospective analyses. Several examples are presented in rare and common disease, along with considerations for background strain selection from the prioritized results.
This work is supported by NIH U54 OD030187, NIH DA028420 and by The Jackson Laboratory, The Cube Initiative Program Fund.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4113W: pyGEMMA: A fast, scalable, and user-friendly Python/R implementation of linear mixed models for genome-wide association studies

Authors:

R. Langefeld, B. Gao, X. Zhou; Univ. of Michigan, Ann Arbor, MI

Abstract:

Linear mixed models (LMMs) are widely used association analysis tools for genome-wide association studies (GWAS) to account for population stratification and relatedness among individuals. While multiple efficient LMM software packages like GEMMA, Bolt-LMM, and REGENIE exist, they often lack an accessible user interface, require specific data formats, and provide minimal parallel scaling. Here, we introduce pyGEMMA, a Python/Cython-based implementation of GEMMA for Python and R that provides a user-friendly framework for conducting LMM-based analyses in genetic studies with parallel computing capabilities. With pyGEMMA, researchers can perform end-to-end experiments with LMMs using high-level functions and scripts. pyGEMMA provides functions for both Python and R (through reticulate) for model-fitting and visualization without the specific data format requirements of other standard software. It includes functions for standard GWAS visualization (e.g., Manhattan plot, Q-Q plot). We also include computational efficiency improvements by leveraging tensor-based storage in NumPy to take advantage of additional BLAS and LAPACK operations. The package includes functionality to run numerous SNPs in parallel in order to scale to the hardware available. Additionally, we implement multi-node, scalable relatedness matrix calculation and

eigendecomposition through the MPI communications protocol and SLATE (Software for Linear Algebra Targeting Exascale) library. We demonstrate the effectiveness of pyGEMMA through GWAS analyses on three datasets. We also demonstrate the scalability of pyGEMMA by running on over 300k individuals of European descent from the UK Biobank. Our results show that pyGEMMA produces identical results to GEMMA with runtimes faster than competing methods while being more convenient and user-friendly. Through its multi-node computing approach, pyGEMMA's implementation also scales to larger datasets than most competing methods and maintains a reasonable runtime. By integrating GEMMA into Python and R, researchers can leverage its extensive functionality for LMM analysis while enjoying the ease and flexibility of Python/R programming. pyGEMMA is freely available on GitHub.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4114W: Real-time dynamic polygenic prediction for streaming data

Authors:

J. Tubbs[1], Y. Chen[2], R. Duan[3], H. Huang[1], T. Ge[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst., Boston, MA, [3]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Polygenic risk scores (PRSs) hold promise in advancing precision medicine. However, existing PRS construction methods rely on static summary statistics derived from genome-wide association studies (GWASs), which are updated at lengthy intervals. As genetic data and health outcomes are continuously being generated at an ever-increasing pace, the current PRS training and deployment paradigm is suboptimal in maximizing the prediction accuracy of PRSs for incoming patients in healthcare settings.

We introduce real-time PRS-CS (rtPRS-CS), which uses a highly efficient stochastic gradient descent algorithm to update genome-wide SNP weights within seconds as each new sample is collected, enabling online, dynamic refinement and calibration of PRSs. Through extensive simulation studies, we demonstrate that rtPRS-CS is robust to various genetic architectures and training sample sizes.

Leveraging 21 quantitative traits measured in the Mass General Brigham Biobank (MGBB) and UK Biobank (UKBB), we show that rtPRS-CS can efficiently integrate massive streaming data to enhance PRS prediction over time. For instance, when applying rtPRS-CS to body mass index (BMI) with initial SNP weights trained on a GWAS of 27,582 MGBB participants, the variance explained in BMI increases from a baseline of 5% by an average of 0.3% for each additional

10,000 UKBB samples incorporated. In an independent holdout sample, across all traits examined, rtPRS-CS significantly outperforms the PRS constructed from baseline GWAS, with its predictive performance statistically indistinguishable from the theoretical upper limit.
We further apply rtPRS-CS to schizophrenia cohorts across 6 East Asian regions, totaling 25,294 cases and 31,304 controls, demonstrating the clinical utility of rtPRS-CS in dynamically predicting and stratifying disease risk. Across the 15 cohorts contributing to the target dataset with diverse genetic ancestries, rtPRS-CS produces well-calibrated PRSs and improves the median variance explained in schizophrenia liability from 1.3%, as observed when applying the PRS trained on the baseline GWAS, to 2.9%.
In summary, rtPRS-CS optimizes polygenic prediction for streaming data, with potentially widespread clinical applications in healthcare systems.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4115W: Reconstruct clonal relationship with mosaic variants at bulk and single-cell level

Authors:

X. Yang[1], C. Chung[2], K. Vong[2], J. Gleeson[2]; [1]Univ. of Utah, Salt Lake City, UT, [2]Univ. of California, San Diego, La Jolla, CA

Abstract:

Throughout life, cells in our bodies accumulate variants due to DNA replication, recombination, damage, and environmental exposure. While most of these variants are repaired by various mechanisms, a small portion remains unrepaired in the genome. It is passed to daughter cells originating from the same progenitor—this distribution of cells carrying different variants results in genomic mosaicism. We can trace the clonal relationships among cells by analyzing these mosaic variants and reconstructing potential lineage patterns.
Here, we present a computational methodology to reconstruct clonal relationships with naturally occurring mosaic variants. Our method is used for reconstructing early human developmental patterns. Our approach utilizes whole-genome deep sequencing and machine-learning-based variant calling pipelines on DNA extracted from bulk tissue to profile candidate mosaic SNV/INDELs with <0.05 false positive rate. Additionally, we obtain a validation dataset by integrating state-of-the-art transcriptome and genome analysis from single nuclei using ResolveOME. Mosaic variants from bulk samples were further quantified using massive parallel amplicon sequencing (MPAS) and single nuclei MPAS (snMPAS).
Lineages are deconvolved based on the unequal representation of fractions of mosaic variants

in different tissue samples and cell populations, reflecting the actual proportion of progenitors at the time the variants occurred. Our single-cell lineage reconstruction validates the hypotheses generated from bulk mosaic variant data.
Our strategy infers early human developmental milestones at both the organ and cell populations. It also provides new insights into the intricacies of human development. This top-down method ensures sufficient marker sharing between samples and reduces false positive detections compared to single-cell-based bottom-up methods. This is particularly significant in studying early human embryonic development, where interventions like CRISPR-based lineage tracing are not feasible. Additionally, our method aids in understanding finer details of human development that may remain elusive due to sample limitations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4116W: Reproductive factors and sex hormones levels on differentiated thyroid cancer risk A Mendelian Randomization study

Authors:

S. Park[1], P-E. Sugier[1], Y. Asgari[1], M. Karimi[1], EPIC consortium, EPITHYR consortium, H. Thomsen[2], A. Försti[3,4], R. Elisei[5], F. Gemignani[6], S. Landi[7], S. Rinaldi[8], A. Elbaz[1], C. Domenighetti[1], T. Truong[1]; [1]Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Villejuif, France, [2]MSB Med. Sch. Berlin, D-14197, Berlin, Germany, [3]Hopp Children's Cancer Ctr. (KiTZ), Heidelberg, Germany, [4]Div. of Pediatric Neurooncology, German Cancer Res. Ctr. (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany, [5]Dept. of Clinical and Experimental Med., Endocrine Unit, Pisa Univ. Hosp., Pisa, Italy, [6]Dept. of Biology, Genetic Unit, Univ. of Pisa, Via Derna 1, 56126, Pisa, Italy, [7]Dept. of Management, Università Ca' Foscari, Venezia, Italy, [8]Intl. Agency for Res. on Cancer, World Hlth.Organization, Lyon, France

Abstract:

Differentiated thyroid carcinoma (DTC) occurs three times more frequently in women than in men, yet the underlying biological mechanisms driving this discrepancy are not well understood. To investigate the causal role of sex hormones and reproductive factors in the DTC risk, we implemented a two-sample Mendelian Randomization (MR) analysis using genome-wide association studies (GWAS) summary statistics. GWAS on DTC were derived from a meta-analysis of six studies, including 7,705 cases and 963,612 controls of European ancestry, while GWAS summary statistics on sex hormones (estradiol, progesterone, sex hormone-binding globulin (SHBG), and testosterones), reproductive factors (age at first birth, age at menarche and menopause), and gynecological conditions (endometrial cancer, endometriosis, pelvic organ

prolapse, polycystic ovary syndrome and uterine fibroids) were retrieved from publicly available sources. We used the inverse-variance weighted (IVW) method to estimate odds ratio (OR), with additional multiple sensitivity analyses to validate the MR analysis, and also conducted multivariable MR (MVMR) to account for potential confounding by obesity-related traits and thyroid stimulating hormones (TSH). Our IVW analysis identified a positive association between SHBG and DTC (OR_{ivw}=1.13, P =0.046). After controlling for TSH and body mass index (BMI) in a MVMR analysis, the strength of this association remained similar but lost statistical significance. Bioavailable testosterone also showed a positive but marginally significant association with DTC after adjustment for BMI (OR_{ivw}=1.13, P =0.07). Putative causal association was observed with uterine fibroids in women less than 50 years-old (OR_{ivw}=1.52, P=0.017). Endometrial cancer was associated with DTC (OR_{ivw}=1.15, P=9.0×10^{-3}); however, a genetic correlation of r^2=13% suggested potential pleiotropy. No significant associations were observed for other investigated factors. In conclusion, our study does not provide strong evidence for a causal role of reproductive and hormonal factors in DTC risk, despite the observed sex disparity in incidence rates.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4117W: Robust tests for association between polygenic scores and ancestry

Authors:

J. Blanc, J. Berg; Univ. of Chicago, Chicago, IL

Abstract:

Polygenic scores have become an important tool in human genetics, enabling the prediction of individuals' phenotypes from their genotypes. Understanding how the pattern of differences in polygenic score predictions across individuals intersects with variation in ancestry can provide insights into the evolutionary forces acting on the trait in question, and is important for understanding health disparities. However, because most polygenic scores are computed using effect estimates from population samples, they are susceptible to confounding by both genetic and environmental effects that are correlated with ancestry. The extent to which this confounding drives patterns in the distribution of polygenic scores depends on patterns of population structure in both the original estimation panel and in the prediction/test panel. Specifically, the amount of bias in the association between a polygenic score and an axis of ancestry variation in the test panel depends on the extent to which that axis is represented in the GWAS panel. Standard approaches, such as including the top genetic principal components

as fixed effects when estimating effect sizes, are intended to solve this problem, but have been shown to fail in important cases. Here, we develop a simple method to estimate the proportion of genetic variance in a given GWAS panel that lies along the test axis. This method allows us to make quantitative statements about the strength of confounding that would need to be present in the GWAS panel to bias the resulting polygenic score association test. For cases where substantial potential for confounding exists, we develop a second method to determine whether the top genetic principal components of the GWAS panel are sufficient to eliminate the confounding. Finally, we apply our procedure to documented cases of polygenic selection acting on human complex traits; first testing their susceptibility to confounding, then determining to what extent the included principal components reduce or eliminate the susceptibility, and finally re-analyzing the results.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4118W: Scoring familial risks in 7million Finns using a pedigree-based BLUP method provides a unique angle to perceive and predict human phenotypes ⭐

Authors:

F. Wang[1,2], Z. Zheng[3,4,5], V. Pinto de Andrade Anapaz[1], Z. Yang[1], E. Pedersen[6,7], N. Mars[1,4], Y. Wang[4,3,5,8], E. Karaman[9], T. Kiiskinen[1,10], M. Ferro[1], K. Detrois[1], R. Rodosthenous[1], FinnGen, G. Genovese[3,5,11], M. Reeve[1], M. Kurki[4,3,5], A. Palotie[1,3,4,5], M. Perola[10], C-Y. Chen[12], M. Daly[1,4,13,5,8], T. Tuomi[1,14,15,16], P. Visscher[17,18], M. Pirinen[1,19,20], A. Ganna[1,5,4], A. Liu[3,1,4,5,8]; [1]Inst. for Molecular Med. Finland, Helsinki, Finland, [2]Ctr. for Population Hlth.Sci., Usher Inst. of Population Hlth.Sci. and Informatics, Univ. of Edinburgh, Edinburgh, Scotland, United Kingdom, [3]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [5]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA, [6]Natl. Ctr. for Register-Based Res., Aarhus Univ., Aarhus, Denmark, [7]Lundbeck Fndn. Initiative for Integrative Psychiatric Res., Aarhus, Denmark, [8]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [9]Ctr. for Quantitative Genetics and Genomic, Aarhus Univ., Aarhus, Denmark, [10]Finnish Inst Hlth & Welfare, Helsinki, Finland, [11]Dept. of Genetics, Harvard Med. Sch., Cambridge, MA, [12]Biogen, Lexington, MA, [13]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Boston, MA, [14]Abdominal Ctr., Endocrinology, Helsinki Univ. Hosp., Helsinki, Finland, [15]Folkhälsan Res. Ctr., Helsinki, Finland, [16]Lund Univ. Diabetes Ctr., Malmo, Sweden, [17]Univ Queensland, St Lucia, Australia, [18]Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [19]Dept. of Mathematics and

Statistics, Univ. of Helsinki, Helsinki, Finland, [20]Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland

Abstract:

Family members have shared genetic and environmental determinants and thus resemble in phenotypes; such resemblances can be quantified as heritability or used to address disease risk. Thus far, research in the field has concentrated on close relatives such as parents and siblings, albeit detailed familial information now becomes available in health registers of many nations. Here, to fully utilize information on any genealogy-traceable relatives, we propose a novel method - pedigree-based best linear unbiased prediction (BLUP), for which heritability is first estimated with Haseman-Elston regression, and then, for every individual linked in the pedigree, an aggregated familial score (FRS) is predicted with a conjugate gradient method based on estimated heritability and relative information. We applied it to two Finnish datasets - FinRegistry, which has a deep pedigree spanning up to 8 generations for the whole nation (N of individuals=7M, N of relative pairs=112M), and also FinnGen, a genotyped subset (N=520K, pedigree-based GRM 10 times denser than UKBB), both linked to electronic healthcare records available since the 70's. In FinRegistry, for most examined diseases, we observed improved prediction accuracy by integrating FRS to the baseline model of age and sex. For example, AUC increased from 0.74 to 0.77 for asthma, 0.67 to 0.70 for T2D, 0.53 to 0.58 for hypertension, and 0.65 to 0.69 for schizophrenia. The FRS also performed well in distinguishing high-risk individuals - for asthma, individuals among the top 1% of FRS confronted a 5-fold risk of asthma diagnosis than the remaining. In FinnGen, we compared FRS to polygenic risk score (PRS) and found that their joint contribution was stronger than that of either alone. This is because PRS captures the additive genetic value represented by common SNPs, while pedigree BLUP, by design, covers all factors shared among families, and thus also considers additional information such as monogenic determinants and common environments. We note that the method is computationally efficient, e.g. in FinRegistry, for a single trait, the computation took a few minutes and consumed less than 10GB of memory, demonstrating its feasibility of a phenome-wide scan. While PRS has gained recognition in medical genetics, its accessibility is restricted to genotyped individuals. We propose a novel score - FRS - that can predict disease risk at a nationwide scale to cover the large non-genotyped population, and also suggest a potential to further boost predictive power among genotyped individuals.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4119W: Sex-specific genetic resource for identifying genes influencing neurological traits and disorders through expression captured by neuroimaging

Authors:

T-C. Wang[1,2], X. Bledsoe[1], E. R. Gamazon[1], L. Dumitrescu[1,2], J. Below[1]; [1]Vanderbilt Genetics Inst. and Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Memory and Alzheimer's Ctr., Dept. of Neurology, Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Women are disproportionally affected by heritable, polygenic, neuro disorders, such as Alzheimer's disease and depression, yet, genetic effects underlying the observed sex difference are poorly characterized. In this study, we leverage two measures: genetically regulated gene expression (GReX) derived from the GTEx database and sex-stratified summary statistics of 3,935 neuroimaging-derived phenotypes (NIDPs) from UK Biobank MR imaging. We aim to identify genes whose GReX is linked to NIDPs separately in females and males to elucidate the sex-specific genetic influence on neurophysiological features captured by NIDPs.
We performed transcriptome-wide association (TWAS) analyses using SPrediXcan on the 3,935 sex-stratified summary statistics of NIDPs (22,138 subjects; 11,624 genetic females) and the GReX derived from 49 GTEx tissues. The GReX was generated by applying the Joint Tissue Model framework, which uses the common regulatory architecture of tissues to enhance gene expression prediction. For quality control, TWAS statistics with a low prediction ($R^2 < 0.1$), NIDPs with non-significant SNP-mediated heritability, and results with genes in inversion regions on chromosome (chr) 8 and chr 17, and the MHC region on chr 6, were removed. Next, the Benjamini-Hochberg false discovery rate (FDR) threshold of 0.05 was applied to correct for multiple testing.We identified 45,719 FDR-significant GReX-NIDP TWAS associations, 9,276 are shared, 18,499 are female- and 17,944 are male-specific. The associations shared between sexes have directionally concordant zscore estimates. The above numbers exclude 4,236 TWAS statistics from sex-specific tissues: ovary, vagina, uterus, testis, and prostate. The GReX-NIDP TWAS associations are enriched for cortical grey-white contrast features across shared and sex-specific associations. Most GReX in shared GReX-NIDP associations are enriched in brain tissues, while female- and male-specific associations are enriched in nerve-tibial, brain, thyroid, and other tissues. This resource can inform hypotheses about gene function. For instance, in males, lower expression of *SLC39A13* (zscore: -4.9) in the frontal cortex BA9 may impair zinc homeostasis in the right cingulate gyrus, affecting the cingulum bundle microstructural integrity, as reflected by dMRI data.
In conclusion, our study provides a rich resource linking GReX to neurophysiology in a sex-

specific context. This resource can be used to prioritize gene candidates for further validation and advance our understanding of how gene regulation affects brain physiology differently in women and men, and potentially to sex differences in disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4120W: Shared and unique 3D genomic features of substance use disorders across multiple cell types reveal a role for the pancreas in pathogenesis ⭐

Authors:

K. Trang[1], A. Chesi[2], S. Toikumo[2], J. Pippin[1], M. Pahl[1], M. S. Kayser[2], A. D. Wells[1,2], H. R. Kranzler[2], R. L. Kember[2], S. F. A. Grant[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Recent genome-wide association studies (GWAS) have revealed shared genetic components among alcohol, opioid, tobacco and cannabis use disorders. However, the extent of the underlying shared causal variants and effector genes, along with their cellular context, remain unclear. We leveraged our existing 3D genomic datasets comprising high-resolution promoter-focused Capture-C/Hi-C, ATAC-seq and RNA-seq across over 50 diverse human cell types to focus on genomic regions that coincide with GWAS loci. Using stratified LD regression, we determined the proportion of genome-wide SNP heritability attributable to the features assayed across our cell types by integrating recent GWAS summary statistics for the relevant traits: alcohol use disorder (AUD), tobacco use disorder (TUD), opioid use disorder (OUD) and cannabis use disorder (CanUD). Significant enrichments ($P<0.05$) were observed in 14 specific cell types, with heritability reaching 9.2-fold for iPSC-derived cortical neurons and neural progenitors, confirming that they are crucial cell types for further functional exploration. Additionally, several pancreatic cell types, notably pancreatic beta cells, showed enrichment for TUD, with heritability enrichments up to 4.8-fold, suggesting genomic overlap with metabolic processes. Further investigation revealed significant positive genetic correlations between T2D with both TUD and CanUD ($FDR<0.05$) and a significant negative correlation with AUD. Interestingly, after partitioning the heritability for each cell type's cis-regulatory elements, the correlation between T2D and TUD for pancreatic beta cells increased ($r=0.2$) compared to the genetic correlation value. Our study provides new genomic insights into substance use disorders and implicates cell types where functional follow-up studies could reveal causal variant-gene mechanisms underpinning these disorders.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4121W: Single-subject omics approaches unravel pathway-specific heterogeneity in Alzheimer's disease

Authors:

A. Podder[1], Y. J. Liew[1], R. S. Pandey[1], G. W. Carter[2,1], A. Uyar[1]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]The Jackson Lab., Bar Harbor, ME

Abstract:

Alzheimer's disease (AD), a common neurodegenerative condition linked with aging, manifests as progressive cognitive decline. Despite extensive research over several decades, the observed inter-individual variability in the clinical presentation of AD has impeded the development of targeted therapies. Hence, there is a pressing need for novel and holistic methodologies aimed at deciphering the precise molecular mechanisms underlying heterogenous disease phenotypes.

In this study, we propose for an individual-centric methodology to elucidate the unique molecular signatures associated with AD. Our endeavor aims to unveil critical molecular pathways essential for both the onset and progression of the disease, while assessing the relevance of heterogeneity in these pathways among individuals. We utilize single-sample omics approaches to unveil critical transcriptomic and proteomic signatures in post-mortem brain tissues retrieved from the Religious Orders Study and Memory and Aging Project (ROSMAP), two ongoing longitudinal studies of AD in the United States providing multimodal datasets. Employing single-sample Differential Gene Expression Analysis (ssDGEA), we identified genes exhibiting significant perturbations within individual AD samples. Following this, single-sample Gene Set Enrichment Analysis (ssGSEA) was conducted to detect dysregulated pathways associated with AD across multiple individuals.

Our preliminary results indicate significant sex-specific differences in AD pathophysiology. Specifically, immune response and metabolic process pathways were notably enriched in male AD cases, whereas synaptic receptor signaling pathways were predominant in female AD cases. These results were consistent at both the transcriptomics and proteomics levels. Additionally, based on the candidate pathway activity profile, we further categorized AD cases into biologically coherent subgroups using a graph-based clustering method. We detected molecularly distinct subtypes in AD implying that these subsets of individuals may exhibit a differential response to pathway-targeted treatments among AD cases. We are currently applying these measures to additional AD datasets to validate our findings and enhance our understanding of individual-specific pathway heterogeneity.

The proposed computational framework is a step towards the implementation of precision medicine principles in AD. The findings from this study will enhance our understanding of disease heterogeneity at the individual and subtype levels and will contribute in efforts to accelerate the discovery of targeted therapeutic interventions in the field.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4122W: Sleep Deprivation and Stress are additively associated with shortened methylation based telomere length

Authors:

J. Jung[1], D. Rosoff[2], J. Wagner[1], T. Perlstein[1], J. Reitz[1], F. Lohoff[1]; [1]NIH / NIAAA, Bethesda, MD, [2]Univ. of Oxford, Oxford, United Kingdom

Abstract:

Stress is associated with premature aging and impacts sleep quality and can contribute to insomnia. Conversely, sleep deprivation (SD) by itself can influence cortisol release through the activation of the hypothalamic-pituitary-adrenal (HPA) axis and active the stress system. Both stress and chronic SD have also been associated with age-related diseases; however, the underlying interrelationships and mechanisms contributing to aging have not been well studied. To better understand the roles of stress and sleep on biological aging, DNA methylation derived telomere length (DNAmTL), methylation score of stress (MSS), and Pittsburgh sleep questions index (PSQI) and a reduced sleep question through computerized patient record system were assessed in a sample of 615 participants (372 individuals with alcohol use disorder and 243 healthy controls). Methylation-based biological signature of severe SD was developed using a machine learning approach and studied with various alcohol-related clinical phenotypes.Reduced sleep was associated with average drinking per day and high binge drinking level ($p=0.001$), increased MSS ($p=0.0001$), and shortened DNAmTL ($p=0.015$). Furthermore, the increased total score of the PSQI showed a significant association with DNAmTL shortening ($p=0.04$) and elevated MSS ($p=0.0003$). Methylation-derived prediction of severe SD was related to DNAmTL compared to individuals without severe SD ($p=0.0002$). MS stress and SD showed additive decreases in DNAmTL. MSS was significantly associated with sleep duration, disturbance, day dysfunction, and overall quality ($p<0.05$). Sensitivity analysis of individuals without major depressive disorder still showed significant associations, but to a lesser extent.Our studies showed that SD is associated with stress-related clinical phenotypes as well as methylation-based cellular aging. We found that stress and sleep problems additively contribute to the acceleration of biological age, as evidenced by the shortening of telomere

length. Our findings add to the understanding of how stress and sleep interact to affect biological aging.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4123W: Somatic CNVs in oncogenic and vascular malformation genes enriched in Cerebral Cavernous Malformations

Authors:

A. Ressler[1], D. MARCHUK[2]; [1]Duke Univ., Durham, NC, [2]Duke Univ. Sch. of Med., Chapel Hill, NC

Abstract:

Cerebral Cavernous Malformations (CCMs) are vascular lesions which can cause brain hemorrhage and stroke, with no curative therapeutic options other than invasive surgical resection. Inherited forms of CCM, with multifocal vascular brain malformations, are driven by germline loss of function mutations in one of three CCM genes - KRIT1, CCM2 or PDCD10. Sporadic, solitary CCMs are driven by biallelic somatic loss of function variants in one of the three CCM genes or an activating variant in MAP3K3, alongside an oncogenic mutation, with somatic PIK3CA gain-of-function variants commonly identified. The past two decades of CCM research have led to the identification of at least one pathogenic variant in most recently resected lesions. However, most genetic approaches focus on identifying small deletions and single nucleotide events in known disease-causing genes leading to missed structural and copy number events. In our recently published work, we developed a statistical framework to identify pathogenic somatic loss-of-heterozygosity as a mechanism of lesion genesis in resected CCMs. Following the identification of pathogenic chromosomal events, we began using high density SNP-arrays to re-assess previously sequenced resected lesions with a spectrum of genotypic and phenotypic characteristics, hypothesizing copy number variants (CNVs) could identify previously missed events in known CCM-causing genes, in addition to identifying additional mechanisms that either lead to lesion genesis in the absence of CCM,MAP3K3 or PIK3CA variants or in a modulatory role. Critically, while the genetics of CCMs are relatively well understood, there remain a subset of lesions without a genetic diagnosis and there is paucity of data on how genotypic variation contributes to aggressivity of lesions. By quantifying CNVs enriched in CCM lesions compared to alternative vascular malformations and non-lesional brain, we identified multiple oncogenes and vascular malformation genes with deleterious somatic CNVs. We propose these CNVs set the foundation for the next generation of CCM genetic analyses focused on lesional genotypic and phenotypic heterogeneity and implicate alternative therapeutic

pathways that target gene networks independent of the canonical pathogenic CCM variants in one of three CCM genes, MAP3K3 or PIK3CA.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4124W: Sparse Bayesian variable selection using global-local shrinkage priors for the analysis of cancer datasets

Authors:

Z. Ma[1], M. Wang[2]; [1]Univ. of Texas Rio Grande Valley, Edinburg, TX, [2]Univ. of Texas at San Antonio, San Antonio, TX

Abstract:

In this research, we propose a sparse Bayesian procedure for the problems of variable selection and classification in high-dimensional logistic regression models based on the global-local (GL) shrinkage prior framework. Particularly we first consider two types of GL shrinkage priors, the horseshoe (HS) prior and the normal-gamma (NG) prior, for the regression coefficients and then specify a correlated prior for the binary vector to distinguish models with the same size. By using mixture representations of the logistic distribution and the considered GL shrinkage priors, we construct a Bayesian hierarchical modeling, which allows researchers to develop an effective Markov chain Monte Carlo (MCMC)-based computation algorithm to generate posterior samples for making the posterior inference. We carry out simulations to compare the finite sample performances of the proposed Bayesian method with existing Bayesian methods in terms of the accuracy of variable selection and prediction. Finally, a real-data application with cancer dataset is provided for illustrative purposes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4125W: Statistical and population genetic approaches to enhance the discovery of trait-associated loci in understudied populations with small sample size ⭐

Authors:

B. Dinh[1], X. Wang[1], H. Tian[1], J. Tang[1], X. Sheng[1], J. C. Carlson[2], D. E. Weeks[2], L. R. Wilkens[3], L. Le Marchand[3], C. A. Haiman[1], C. W. K. Chiang[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Univ. of Hawaii Cancer Ctr., Honolulu, HI

Abstract:

Genome-wide association studies (GWAS) have revealed the variation and biological mechanisms that contribute to human traits and diseases. Despite recent focus on non-European cohorts, studies for smaller and susceptible populations remain challenging due to their unique population structure, patterns of admixture, and small cohort size. We adopt several approaches aimed to improve the robustness and power of detecting population-enriched associations in a small population sample, focusing on the Native Hawaiians as a model. First, we combine genotype imputation results for 5,382 Native Hawaiians from a representative reference panel, the Multiethnic Cohort Study, with that from the TOPMed reference panel through meta-imputation, which we have shown in other work to improve the accuracy and coverage of imputation including at Polynesian-enriched variants. Second, we use the genetic relationship matrix (GRM) computed from inferred genome-wide genealogies in a linear mixed model (LMM) to improve the adjustment for fine-scale population structure. Compared to standard GWAS (an LMM with the standard GRM and TOPMed imputed genotypes), this approach results in a 1.6-2.3% increase in effective sample size across variants shown to be associated (P < 5e-6 - 5e-8) with 13 adiposity traits, and maintains robustness of the analysis (mean λ_{GC} = 1.0123). Third, we include the polygenic scores (PGS) from studies of European cohorts as covariates which has been shown to improve statistical power in GWAS. For BMI, we find that including PGS results in a 5.6-24.8% increase in effective sample size among associated variants (P < 5e-6 - 5e-8) compared to excluding the PGS. Lastly, we imputed missing phenotypes through a method combining the GRM and phenotypic correlation. The efficacy of this approach depends on the proportion of missing phenotypes, but the effective sample size increases up to 3.5% in traits such as hip circumference (where >25% of samples are missing). Incorporating all of these approaches into a GWAS and focusing on BMI as an illustrative example, we find the ensemble approach identified 202 SNPs at 47 loci at a P-value cutoff of 5e-6. 22 of 47 loci are potentially novel and 16 are not found using a standard GWAS; these loci will be prime targets for replication or further meta-analysis with other cohorts. Among 115 variants with P < 5e-6 across the 22 loci, the mean increase in effective sample size is 28.6% using our ensemble approach compared to the standard GWAS. Taken together, our results suggest this framework can be adopted to improve the power for discovering population-enriched novel loci associated with complex traits in cahoots with limited sample size.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4126W: Systematically quantifying gain- and loss-of-function biochemical consequences of missense variants using UK Biobank exomes ⭐

Authors:

A. Holleman, R. Hoffing, L. Krohn, P. LoGerfo, M. Plekan, S. Akle Serrano, A. Deaton, L. Ward; Alnylam Pharmaceuticals, Cambridge, MA

Abstract:

RNAi therapeutics are a new class of medicines that act through the cell's endogenous gene silencing mechanism. Human genetics has proven to be a useful way to discover novel targets for RNAi therapeutics. To determine whether new genetic discoveries are translatable to an RNAi therapy, we need to ascertain if the variants that cause disease are gain- or loss-of-function. There is no reliable bioinformatic prediction of missense direction of effect from sequence alone. However, rare protein-truncating variants (PTVs) provide a directional link between genes and phenotypes, and these can be used to infer the directionality of missense variants. For example, rare PTVs in *PCSK9* associate with lower LDL; because of this, we can infer that *PCSK9* R46L, which lowers LDL, is loss-of-function (LOF) and *PCSK9* R496W, which raises LDL, is gain-of-function (GOF). To test this inference method on a larger scale, we curated a gold standard of mechanistically understood PTV associations in the UK Biobank data spanning 173 genes (for example, *PCSK9* and lower LDL) for systematic downstream analysis. We then used Regenie and the UK Biobank to test 21,435 rare (MAF < 1%) missense variants in these genes for association with the known PTV phenotypes, 647 of which were significantly associated at $p < 2.3e-6$. Of the 647 missense variants, 610 were inferred to be LOF because they had the same direction as PTVs, and 37 were inferred to be GOF, including the known *PCSK9* R496W variant. Our quantified effect predictions for these 647 variants correlate well with established algorithms (for CADD, REVEL, and AlphaMissense, $r^2 = 0.24, 0.34, 0.43$, respectively). Compared to clinical annotations in ClinVar, half (305/610) of the predicted LOF missense variants are not annotated at all. Of the 305 annotated variants, only 88 are labeled as pathogenic or likely pathogenic; 176 are labeled as VUS (variants of unknown significance) or have conflicting interpretations, and the remaining are annotated as benign. This method provides a way of using biobank exome sequences to discover disease-causing GOF variants or disease-protective LOF variants in genes that could be targets for RNAi therapeutics. It could also be used to discover and re-classify missense LOF variants in known disease genes, and provide a more comprehensive training dataset than ClinVar for the new generation of AI methods that attempt to predict variant effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4127W: The comparison of data-driven and knowledge-driven approaches to clustering a Mendelian Randomization instrument, using multiple methods

Authors:

C. Fawcett, C. P. Nelson, V. Codd; Univ. of Leicester, Leicester, United Kingdom

Abstract:

Mendelian randomization (MR) can be used to investigate the causal relationship between an exposure and an outcome, using an instrument consisting of genetic variants associated with the exposure. Clustering methods for MR have been developed to allow variants in the instrument to be grouped into clusters, which may identify pathway specific exposure-outcome relationships. In addition to methods which can generate clusters based on genetic association, it is also possible to manually cluster genetic variants based on established scientific knowledge. A causal relationship between telomere length (TL) and abdominal aortic aneurysm (AAA) has previously been established via MR, although its underlying mechanism is not understood. Five manually-curated clusters for the TL instrument have been developed based on known telomere biology. We found that SNPs with known biological relevance to TL and the cell cycle were significantly associated with AAA, whilst the addition of other TL SNPs into the instrument may result in increased noise. Interestingly, however, a cluster containing SNPs with no known biological relevance to TL was also found to be statistically significant. Hierarchical-based data-driven clustering methods on the complete TL instrument identify only a single cluster, alongside a null cluster, with similar SNPs being identified across multiple hierarchical methods. Additionally, SNPs belonging to manually-curated clusters with both known biological relevance and unknown biological relevance are similarly prioritised during hierarchical data-driven clustering. Early work comparing hierarchical-based data-driven clustering and manually-curated clustering has been previously presented at the international 6th Mendelian Randomization Conference. Here we report on new advances to this earlier work, including additional data-driven clustering methods that provide further insight onto the underlying biological pathways linking TL and AAA, beyond those previously reported.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4128W: The Covariate Adjusted PERmutation (CAPER) tool for genetic association analysis

Authors:

R. Bohlender[1], Y. Yu[2], J-S. Chen[1], J. Below[3], C. Huff[4]; [1]U.T. MD Anderson Cancer Ctr, Houston, TX, [2]UT MD Anderson Cancer Ctr., Houston, TX, [3]Vanderbilt Univ Med Ctr., Nashville, TN, [4]U.T. MD Anderson Cancer Ctr, Pearland, TX

Abstract:

Covariate-adjusted permutation offers a robust and generalizable solution for assessing the statistical significance of rare variant association tests. In biobanks, where extremely skewed case/control sample sizes are common, it is important to have methods that are not sensitive to this imbalance. Due to computational complexity, covariate-adjusted permutation has failed to gain traction despite its usefulness. Thus, analysis of biobank scale sequencing datasets has generally been limited to statistical analysis methods with analytical approximations. Many statistics are biased in the face of significant differences in case/control sample size. Even using the saddlepoint approximation, inflation is still present when case/control sample sizes are highly imbalanced, which is a common occurrence in biobanks. Here, we introduce The Covariate Adjusted PERmutation (CAPER) Tool for Genetic Association Analysis, which is designed to conduct fast and memory efficient multithreaded association analysis of large sample sets, with support for highly scalable covariate adjusted permutation. CAPER currently supports 11 statistical tests, including CMC, SKAT, SKAT-O, WSS, VT, and VAAST. The core innovation in CAPER is an adaptive permutation engine that implements covariate-adjusted permutation of large datasets. In addition, CAPER supports approximate binning of individuals with similar odds of being a case, reducing the number of categories to permute. This binning greatly reduces the computational burden of permuting large datasets, while allowing users to optimize the tradeoff between computation and statistical accuracy. To limit memory usage, CAPER leverages sparse matrices and shared memory where possible. CAPER is implemented in C++ using efficient linear algebra libraries, which further enhances memory efficiency and computational scalability. Total run time for the analysis of a set of roughly 20,000 genes in 500,000 cases and 500,000 controls, running on 80 threads was approximately 121 hours, with maximum memory usage at roughly 31 GB using VAAST. SKAT with covariate-adjusted permutation on the same configuration took 1264 hours and used 141 GB of memory. As our results demonstrate, CAPER implements efficient covariate adjusted permutation at scale, providing the means to conduct robust rare variant association studies on very large data sets with highly asymmetric case/control sample sizes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4129W: The cumulative effect of protein-truncating variants on human life expectancy and overall disease burden.

Authors:

H. Vy[1], D. Balick[2], D. Jordan[1], R. Do[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Harvard Med. Sch., Boston, MA

Abstract:

Genetic variation is a critical determinant of an individual's overall health and life expectancy. Despite significant advancements in understanding the functional effects of genetic variants in specific genes, the combined impact of deleterious genetic variants on overall health remains poorly understood. Our earlier research demonstrated that the aggregated effect of protein-coding variants exhibits pleiotropic effects on various medical phenotypes. However, these analyses were limited by a lack of rare variants not captured in the imputed data, thus excluding many rare variants that potentially have large detrimental effects. In this study, we focus our analyses on protein truncating variants (PTVs). Previous studies have indicated that a higher number of PTVs in an individual's genome is associated with reduced lifespan and health span. We extended their methodology by introducing a weight parameter to account for variations in the effect sizes of PTVs. Specifically, we adapted the formula proposed by Gardner et al. (2022) to estimate the aggregated effects of PTVs on an individual's genome: $\Delta\omega_i=1-\exp(-\sum_g s_{het[i,g]})$, where $s_{het[i,g]}$ represents the pre-computed selection coefficients for gene *g* with PTVs in individual *i*. As outcomes, we estimated individuals' disability-adjusted life years (DALY) from electronic health records (EHR) as a proxy for overall disease burden, and additionally test human lifespan. We assessed the association of Δω with lifespan and DALY in the exome sequences of 366,731 individuals from UKBB and 29,681 individuals from Bio*Me*, and in the whole genome sequences of 211,240 individuals from *All of Us*. The association analyses were performed separately by genetic ancestry, followed by a meta-analysis. We observed that Δω is strongly associated with lifespan ($p = 6.51\times10^{-15}$) and DALY (2.86×10^{-15}). Among the two components of DALY, Δω exhibited a stronger association with years lived with disability (YLD) ($p = 2.96\times10^{-115}$) than with years of life lost (YLL) ($p = 7.24\times10^{-11}$). Limiting the analysis to the actionable genes listed by the American College of Medical Genetics and Genomics (ACMG v3.2), $\Delta\omega_{ACMG}$ appeared to drive the association with YLL ($p=3.08\times10^{-6}$), whereas Δω from the remaining genes drove the association with YLD ($p=3.38\times10^{-101}$). Our findings advance the understanding of the genetic basis of lifespan and overall disease burden. Importantly, we demonstrate that Δω can serve as a genetic risk factor for both mortality and overall disease burden across multiethnic populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4130W: The impact of Indigenous American ancestry on the risk of acute lymphoblastic leukemia and the implications for future study designs

Authors:

J. Langie[1], L. Morimoto[2], X. Ma[3], C. Metayer[2], J. L. Wiemels[1], A. J. de Smith[1], C. W. K. Chiang[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of California Berkeley, Berkeley, CA, [3]Yale Univ., New Haven, CT

Abstract:

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer, with Hispanic/Latino (H/L) children having up to 1.4 times the rate of ALL compared to their non-Hispanic White counterparts. This disparity has not been fully explained by environmental factors, suggesting the role of understudied genetic variants, particularly ones found in Indigenous American (IA) ancestry inherited by Latinos. In this study, we characterized the impact of IA ancestry on the frequency and effect size of known ALL risk alleles. We found that elevated risk of ALL in Latinos may be conferred by increased frequency of risk alleles due to IA ancestry. For instance, in an analysis of Latinos from the California Cancer Linkage Project (CCRLP; 1,930 cases, 2,103 controls) and the California Childhood Leukemia Study (CCLS; N = 605 cases, 515 controls), we found that, across independent known ALL loci (N=21), more loci than expected by chance showed a significant association between IA ancestry and increasing risk ($p = 0.015$). Compared to randomly-sampled, frequency-matched alleles from the genome, the risk alleles at all known ALL loci also showed a significant association with IA ancestry (46.9 % for known risk alleles vs. 43.5 % for matched alleles; $p= 3.8 \times 10^{-7}$). The IA haplotype had 1.39 times the odds (95% CI: 1.35 - 1.43; $p= 5.9 \times 10^{-46}$) of harboring the risk allele compared to the non-IA haplotype. Conversely, in a combined Latino sample of 2,535 cases and 9,035 controls, we found no evidence of GxAncestry interaction on ALL risk across all known loci (min P = 0.0026, prior to adjusting for multiple testing), suggesting that ALL risk alleles do not have increased effects on the IA ancestry background. This lack of synergism between local ancestry and genotype in ALL has methodological implications for gene discovery in admixed populations. To explore this, we implemented Tractor, a method that models genotype effects by ancestry, and thus is best powered in the presence of effect size heterogeneity by ancestry. Across known loci, we found that Tractor had 36.1% decreased power compared to standard GWAS. Taken together, our results suggest that the disproportionate burden of ALL in Latino populations may arise from

the enrichment of risk alleles within the IA ancestry component, possibly driven by immunity-related selective pressures. Finally, future genetic studies of ALL in Latino patients may benefit from methods that incorporate contributions from ancestry to increase the statistical power in identifying associations, rather than leverage effect size heterogeneity across ancestries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4131W: The Norwegian Childhood Diabetes Registry: clustering of pediatric diabetes cases

Authors:

K. Kuznetsova[1,2], M. A. Juárez Garzón[1,2], D. S. P. Tallapragada[1], M. Udler[3,4,5], A. Manning[4,3,6], M. Vaudel[1,2,7]; [1]Mohn Ctr. for Diabetes Precision Med., Dept. of Clinical Sci., Univ. of Bergen, Bergen, Norway, [2]Computational Biology Unit, Dept. of Informatics, Univ. of Bergen, Bergen, Norway, [3]Diabetes Unit, Massachusetts Gen. Hosp., Boston, MA, [4]Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [5]Metabolism Program, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Metabolism Program, Broad Inst. of MIT and Harvard, Cambridge, MA, MA, [7]Dept. of Genetics and Bioinformatics, Norwegian Inst. of Publ. Hlth., Oslo, Norway

Abstract:

The Norwegian Childhood Diabetes Registry (NCDR) records all incident cases of childhood diabetes nationwide, providing a unique dataset to explore this condition. Our study analyzes 2,890 children aged 1-18, including 2,560 autoantibody-positive (Ab+) and 330 autoantibody-negative (Ab-) cases. We conduct patient clustering based on clinical profiles and outcomes. This research aims to enhance understanding of diabetes heterogeneity and improve patient management strategies. The young age of the patients makes the clustering challenging as most of the previous studies have been done on cohorts of adult individuals.
Participants were manually clustered based on autoantibody status and sex, and then, k-means clustering was used to further separate these subgroups based on age at diagnosis, glucose, HbA1c, C-peptide, and BMI. C-peptide and BMI were adjusted using age and sex-appropriate SD scores before clustering. Finally, we computed polygenic risk scores (PRSs) for T1D and T2D and diabetes-related traits (β-cell function, insulin secretion, lipodystrophy, etc.) and compared phenotypic variables and PRSs between the different clusters.
Clustering of Ab+ cases resulted in two groups. After adjusting the variables for age and sex, older age at diagnosis was associated with lower BMI and C-peptide levels. Ab- cases clustered in three groups. Cluster 1 presented the lowest levels for glucose and HbA1c and the highest age at diagnosis, C-peptide, and BMI, aligning with the classical presentation of youth-onset

type 2 diabetes. Cluster 2 showed a similar age at diagnosis as Cluster 1 but an opposite phenotype, with the highest glucose and HbA1c levels and the lowest C-peptide and BMI. Cluster 3 presented the youngest age at diagnosis and values between Cluster 1 and 2 for the rest of the variables.
Comparison of PRSs between clusters of Ab+ patients showed significant differences in BMI PRS, with younger cases presenting higher BMI PRS ($p<0.01$) as well as non-significant trends for higher T1D-nonHLA PRS ($p=0.06$), however, no other significant differences were observed. Regarding Ab- cases, Cluster 1 showed significantly higher T2D PRS ($p<0.05$) and lower T1D PRS ($p<0.001$), T1D-nonHLA PRS ($p<0.001$), T1D-HLA PRS ($p<0.001$) and Liver Lipids PRS ($p<0.01$). Comparison between Cluster 2 and 3 showed trends but no significant differences were observed.
The Norwegian Childhood Diabetes Registry cohort provides data for pediatric diabetes investigations. In conclusion, our study shows distinct clustering in childhood diabetes, highlighting the need for personalized treatment approaches based on specific clinical and genetic profiles to optimize outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4132W: The PhenX Toolkit: New Protocols for Sickle Cell Disease Pregnancy, Medical Cannabis, Polysubstance Use and Recovery, and Updated Protocols for Tobacco Regulatory Research

Authors:

M. Sullivan[1], A. K. Malinowski[2], D. Sharma[3], A. Buu[4], J. Kelly[5], M. Morean[6], C. A. Stanton[7], J. Abedini[1], L. Cox[1], T. Hendershot[1], W. Huggins[1], C. Ives[1], M. C. Krzyzanowski[1], K. Washington[1], D. Williams[1], L. Wood[1], C. Hill[1], G. Swank[1], B. Cook[1], P. Ceger[1], E. Fallon[1], J. Ortiz[1], L. Gehtland[1], C. M. Hamilton[1]; [1]RTI Intl., Research Triangle Park, NC, [2]Univ. of Toronto and Mount Sinai Hosp., Toronto, ON, Canada, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]Massachusetts Gen. Hosp., and Harvard Med. Sch., Boston, MA, [6]Yale Sch. of Med., New Haven, CT, [7]Westat, Rockville, MD

Abstract:

The PhenX (consensus measures for **Phen**otypes and e**X**posures) Toolkit is a freely available web-based catalog of measurement protocols recommended for research with human participants. Use of PhenX protocols facilitates data sharing and cross-study analyses and can increase the impact of individual studies. PhenX protocols are recommended by Working Groups (WG) of domain experts using a consensus process that includes community input. The

Toolkit is an established resource that includes protocols across 30 broad research domains (e.g., Obesity, Rare Genetic Conditions) and six focused collections (e.g., Blood Sciences Research, Social Determinants of Health). Users may access the protocols, supporting information, and tools to help incorporate PhenX protocols into their studies. The Toolkit has over 4,900 registered users and has been recommended in more than 660 National Institutes of Health (NIH) Funding Opportunity Announcements. The Toolkit is continually updated and expanded and new content and features are highlighted in this presentation. Funded by the National Heart, Lung, and Blood Institute of the NIH, the PhenX Sickle Cell Disease (SCD) Pregnancy WG, an international panel led by Kinga Malinowski M.D. and Deva Sharma M.D., recommended the addition of nine new protocols and supplemental information to address the intersection of obstetrics, hematology, and SCD. Funded by the National Institute on Drug Abuse, the Medical Cannabis, Polysubstance Use, and Recovery WG, led by Dr. Anne Buu and Dr. John Kelly added new protocols to cover topics including motives, drug interactions, and recovery capital. Funded by the NIH Office of the Director, Office of Disease Prevention, a Content Expert Panel led by Dr. Meghan Morean and Dr. Cassandra Stanton updated or replaced 42 protocols in the PhenX Tobacco Regulatory Research Collection to support standardized data collection for the Tobacco Centers of Regulatory Science investigators. Examples include youth and adult e-cigarette flavor preferences, peer and family influences on smoking, tobacco brand and variety, and protocols from the Population Assessment of Tobacco and Health Study. In addition, new Toolkit features provide citations for protocols in My Toolkit and the ability to filter PhenX protocols used in major studies (e.g, All of Us) as the Who's Using PhenX resource continues to support collaborations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4133W: The role of sleep in the human brain and body: insights from multi-organ imaging genetics

Authors:

Z. Fan[1], Y. Yang[1], Y. Guo[2], Y. Li[2], J. Shu[3], X. Yang[2], B. Li[4], J. Lin[5], Q. Wang[6], C. Gibson[7], P. Paschou[8], T. Li[9], P. Gehrman[10], H. Zhu[9], B. Zhao[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Purdue Univ., West Lafayette, IN, [3]Purdue Univ., Plano, TX, [4]Univ. of California, Los Angeles, Los Angeles, CA, [5]Yale Univ., New Haven, CT, [6]Peking Univ. Sixth Hosp., Beijing, China, [7]Univ. of California, San Francisco, San Francisco, CA, [8]PURDUE Univ., West Lafayette, IN, [9]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [10]Univeristy of Pennsylvania, Philadelphia, PA

Abstract:

Sleep is crucial for overall physical and mental health, concerning organs such as the brain, heart, eye, liver, kidney, and lung. Nonetheless, a thorough understanding of how sleep relates to these organs, as well as their genetic bases, remains elusive. Here we conducted a systematic mapping of sleep-organ connections, focusing on 623 multi-organ imaging biomarkers and 10 sleep traits. Both phenotypic and genetic analyses uncovered robust associations between sleep and the structure and function of multiple organs, such as brain functions measured by functional magnetic resonance imaging (fMRI) and body composition traits in abdominal MRI. Sleep and imaging traits had shared genetic influences in 51 genomic loci, 23 of which had colocalized causal genetic effects. Considering the extensive genetic overlaps between sleep and multi-organ imaging biomarkers, we further described the patterns of genetic similarities between sleep and 50 diseases relevant to these organs. Notably, among all diseases examined, psychiatric disorders consistently exhibited the strongest genetic correlations and causal links with sleep. Mediation analysis further revealed that many of the identified sleep-imaging connections were mediated by intra or inter-organ diseases. Overall, our study demonstrates the broad implications of sleep on brain and body health, influenced in part by shared genetic factors.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4134W: Trajectory analysis of heart failure-related phenotypes from biobank data uncovers novel genetic associations ⭐

Authors:

T. Pottinger[1], L. Pesce[2], A. Gacita[3], L. Montefiori[4], N. Hodge[2], S. Kearns[2], T. Monroe[2], I. Salamone[3], J. Pacheco[2], L. Rasmussen-Torvik[2], M. Smith[2], R. Chisholm[3], M. Nobrega[5], E. McNally[3], M. Puckelwartz[2]; [1]Columbia Univ., New York, NY, [2]Northwestern Univ., Chicago, IL, [3]Northwestern Univ, Chicago, IL, [4]Univ. of Chicago, Chicagao, IL, [5]Univ. of Chicago, Chicago, IL

Abstract:

Background: Approximately 6 million adults in the US have heart failure (HF). HF progression is variable due in part to differences in sex, age, and genetic ancestry. Previous population-based genetic studies have largely focused on single timepoint data related to HF, a disease known to change over time. As an alternative to cross-sectional analyses, we used trajectory probabilities of longitudinal data as a continuous trait to query biologically relevant associations with genome-wide genetic variation in a medical biobank. **Methods**: We evaluated whole genome sequencing of 895 unrelated biobank participants, including 494 with at least 1 electrocardiogram and 324 who had more than 1 echocardiogram (~3 observations per person).

A censored normal distribution multivariable mixture model was used to cluster phenotype measures for genome-wide association (GWA) analyses. Variants that were found to be significantly associated with these longitudinal phenotypes were evaluated in the UK Biobank to test for replication with the first observed electrocardiogram or echocardiogram measurement. **Results**: GWA analysis on the trajectory probability of the corrected QT measurement (QTc) taken from electrocardiograms identified significant associations with variants in regulatory regions proximal to the *WLS* gene, which encodes the WNT ligand secretion mediator, Wntless. *WLS* was previously associated with QT length using approximately 16,000 participants, providing support for the utility of this method to uncover significant genetic associations in small datasets. GWA analysis on the trajectory probability of left ventricular diameter as taken from echocardiograms identified novel significant associations with variants in regulatory regions near *MYO10*, which encodes Myosin-10. These variants were found to be significantly associated with high QTc and derived left ventricular diameter in the UK Biobank replication data. **Conclusions**: We found that trajectory probabilities improved the ability to discover significant and relevant genetic associations to increase yield from small, well-phenotyped cohorts with longitudinal data from a medical biobank.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4135W: Transcriptome-wide association study using summary-level brain eQTL reference data to identify risk genes associated with adolescent neurocognitive function

Authors:

S. Abrishamcar, Q. Dai, J. Yang, A. Huels, M. Epstein; Emory Univ., Atlanta, GA

Abstract:

Introduction: Neurocognitive disorders in adolescents are increasingly common and can be highly heritable, but the underlying genetic mechanisms influencing them remain largely unknown. Transcriptome-wide association studies (TWASs) are important GWAS post-analyses that can be used to identify risk genes that regulate gene expression underlying complex traits. Here we conduct a TWAS of neurocognitive development in adolescence in hopes of gaining a better understanding of the complex genetic architecture underlying neurocognitive disorders. **Methods:** We used the recently published OTTERS tool to train multiple gene expression imputation models (P-value thresholding with linkage disequilibrium clumping [P+T], LASSO [lassosum], and Bayesian multivariable regression with continuous shrinkage [PRS-CS]) using brain eQTL summary statistics from the cortex (N=2683) region obtained from MetaBrain.

For each fitted imputation model per gene, we performed TWAS on independent GWAS summary data of 9 neurocognitive phenotypes measured by the Penn Computerized Neurocognitive Battery from the Philadelphia Neurodevelopmental Cohort (N=3116). Subsequently, for each gene, we combined the TWAS p-values across all imputation models using an aggregated Cauchy association test to yield a final omnibus TWAS p-value. We conducted functional follow-up by referencing the GWAS catalog and Open Targets Genetics. Further validation analyses using the FOCUS tool, which fine-maps TWAS summary statistics at genomic risk regions, are ongoing. **Results:** We identified 2 significant risk genes (FDR<0.05) of the working memory phenotype (*PITPNM2*; *PKNOX2*) and 3 suggestive risk genes (FDR<0.08) of the working memory (*KMT5A*; FDR=0.08), social processing (*CREB3L4*; FDR=0.06), and verbal reasoning (*LAMTOR5-AS1*; FDR=0.05) phenotypes, respectively. The significant risk genes (*PITPNM2*; *PKNOX2*) were enriched in important biological pathways involving neuronal signaling and synaptic function. Some of these genes have been identified in previous GWAS in association with neurocognitive outcomes including schizophrenia (*PITPNM2*; *PKNOX2*) and general cognitive and academic performance (*KMT5A*; *PITPNM2*; *LAMTOR5-AS1*) and neurodevelopmental outcomes including cortical thickness (*CREB3L4*; *PITPNM2*) and white matter connectivity (*PITPNM2*). Additionally, all genes were associated with brain morphology including cortical brain volumes/thickness in locus-to-gene studies. **Conclusion:** These findings provide insight into the biological mechanisms and complex genetic architecture underlying neurocognitive development in adolescents.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4136W: Transfer-learning of PRS with pseudo-validation

Authors:

B. Cho; Seoul Natl. Univ., Goyang-si, Korea, Republic of

Abstract:

Polygenic Risk Scores (PRS) are essential tools for predicting the risk of various phenotypes in individuals. However, their effectiveness is often limited by the scarcity of genetic data for certain populations or phenotypes. To address this issue, Transfer Learning for Polygenic Risk Scores (TL-PRS) has been introduced, leveraging PRS from large datasets to transfer to smaller target datasets. Despite its potential, the current implementation of TL-PRS has limitations: it requires individual-level validation data to select hyperparameters, which are often inaccessible due to privacy concerns. To overcome this, we extended TL-PRS by using a pseudo-validation approach. The new implementation, Pseudo-validated Transfer Learning for Polygenic Risk

Scores (PTL-PRS), incorporates a method to split summary statistics and perform pseudo-validation with these split statistics.
We applied PTL-PRS in two scenarios: using European-ancestry sample PRS to transfer to South Asian ancestry, and using PRS trained on type 2 diabetes (T2D) to transfer to T2D complications. Our analysis shows that transfer learning can improve prediction accuracy and that pseudo-validation provides performance comparable to using individual-level validation data. Specifically, the PTL-PRS model for eight traits, using two PRS methods—Lassosum and PRScs—demonstrated an average improvement of 7.06% in R^2 compared to standard PRS models. Additionally, the prediction accuracy of T2D complications increased by 6.25% in AUC with PTL-PRS compared to the baseline. We also developed software to optimize PTL-PRS implementation by incorporating early stopping and directly reading binary genotype data into matrices, reducing training times by 53.85% (from 26 hours to 12 hours for a million SNPs). Overall, PTL-PRS improves prediction accuracy while enhancing software usability.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4137W: Two decades of admixture mapping studies on human health: ancestry/sociodemographic diversity and insights

Authors:

E. Herrera Luis, C. Duncan, A. Sartori, J. Sharma, G. L. Wojcik; Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Abstract:

Introduction: Recently admixed populations remain underrepresented in genomic research, limiting our understanding of genetic factors affecting health within these groups. However, admixture mapping (AM) leverages local ancestry in admixed populations for powerful locus detection. To date, the impact of AM on genomic discoveries remains unassessed, in contrast with genome-wide association studies (GWAS) or polygenic scores, which benefit from the availability of databases of published studies and their accompanying metadata. **Aim:** To conduct a scientometric review of AM studies on human health and to create a curated database of AM studies. **Methods:** A systematic literature search was conducted of admixture mapping studies on human health published up to December 31, 2023 using PubMed and Clarivate Web of Science, using a predefined protocol for data extraction and synthesis. Regional representation of admixed population diversity was captured and assessed by building upon the GWAS Catalog ancestral framework. Traits and diseases were mapped to Experimental Factor Ontology (EFO) terms. **Results:** A total of 121 AM studies were identified, comprising 497

outcomes and 1,179 statistically significant AM signals. Cancer, cardiovascular, and diabetes/kidney/digestive traits were the most frequently studied outcomes. Out of the 1,083 variant-trait associations identified through fine-mapping of AM peaks, 9.68% were associated with GWAS traits, with AM variants linked to lipids/lipoproteins, metabolites, or protein levels showing more evidence of pleiotropic effects as per the GWAS Catalog. Across traits, African American/Afro-Caribbean and Hispanic/Latin American participants were disproportionately represented (97.7%) compared to other admixed populations. Moreover, most publications included participants from upper-middle-to-high income countries, primarily from the United States (89.26%). AM publications were predominantly authored by scientists affiliated with institutions in high income countries, reflecting the top funders and the participant composition in AM studies. **Conclusions:** Despite the focus of AM on ancestrally diverse groups, both study populations and researchers from diverse backgrounds remain broadly underrepresented. As recently admixed groups remain underrepresented with smaller sample sizes in genomics research, AM can offer a powerful method to leverage limited data. There is also an urgent need to promote multi-layered diversity in genomic research and the genomics workforce to promote community-engaged research and ensure equitable access to health benefits of genomics advances.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4139W: Two-sample Mendelian randomization with common and annotation-weighted rare variants

Authors:

Y. Zhang[1], X. Li[2], N. Liu[3], Q. Yan[4]; [1]Indiana Univ., Bloomington, Bloomington, IN, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Indiana Univ., Bloomington, IN, [4]Columbia Univ. Irving Med. Ctr., New York, NY

Abstract:

Mendelian Randomization (MR) has become an important technique for establishing causal relationships between risk factors and health outcomes, using genetic variants as instrumental variables to overcome confounding and reverse causation in observational studies. Traditional MR analyses have predominantly focused on common genetic variants, which only partially represent the genetic architecture of complex traits and diseases. Rare variants, which can have larger effect sizes and provide unique biological insights, have been largely overlooked due to statistical challenges and interpretive complexities. This study presents a novel two-sample MR framework named MR-CARV (two sample Mendelian Randomization considering both Common

and Annotation-weighted Rare Variants) that incorporates both common and rare genetic variants. This method aims to leverage the comprehensive genetic data made available by high-throughput sequencing technologies and large-scale consortia. The effects of rare variants are assessed using the STAAR pipeline, where rare variants are grouped into functional categories, including gene-coding, gene-noncoding, and non-gene regions. Variant annotations and biological impact are used as weights in test statistics construction. Our simulation studies demonstrate that MR-CARV has a robust type I error and higher power compared to traditional methods that rely solely on common variants. The application of this enhanced MR methodology to blood lipid levels and adverse pregnancy outcomes shows that HDL-C has a significant effect on pre-eclampsia (PE). It is important to note that the current analysis only includes rare variants in the gene-coding regions. Specifically, the estimate for the effect of HDL-C on PE using only common variants is -0.0329 (SE = 0.0128, P = 0.010). When rare variants are included, the estimate is -0.0323 (SE = 0.0124, P = 0.009). This suggests that incorporating rare variants can potentially improve the performance and precision of MR analyses. We are working on including rare variants from non-coding and non-gene regions, which could further enhance the robustness of MR analyses. Our findings emphasize the utility of considering the full spectrum of genetic variation to improve causal inference in genetic epidemiology. This comprehensive approach enhances the power and precision of causal inference, underscoring the importance of including rare variants in MR analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4140W: Unlocking the Power of Biobanks using GPUs - Accelerating GWAS with SAIGE on High Performance Computing and Clouds

Authors:

A. Rodriguez[1], Y. Kim[1], T. Nandi[2], K. Keat[3], R. Kumar[3], R. Bhukar[4], M. Conery[3], M. Liu[5], J. Hessington[6], D. Schmidt[7], E. Begoli[7], G. Tourassi[8], S. Muralidhar[9], P. Natarajan[10], B. Voight[3], K. Cho[11], M. Gaziano[11], S. Damrauer[12], K. Liao[13], W. Zhou[14], J. Huffman[15], A. Verma[3], R. Madduri[1]; [1]Argonne Natl. Lab., Lemont, IL, [2]Argonne Natl. Lab., Westmont, IL, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Broad Inst., Cambridge, MA, [5]7Dept. of Biostatistics, Columbia Univ.'s Mailman Sch. of Publ. Hlth., New York, NY, [6]Information Systems, Universtiy of Pennsylvania, Philadelphia, PA, [7]Oak Ridge Natl. Lab., Oak Ridge, TN, [8]Computing and Computational Sci. Directorate, Oak Ridge Natl. Lab., Oak Ridge, TN, [9]VHA Office of Res. and Dev., Washington, DC, [10]Massachusetts Gen. Hosp., Boston, MA, [11]VA Boston Hlth.care System, Boston, MA, [12]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, [13]Massachusetts

Veterans Epidemiology Res. and Information Ctr. (MAVERIC), VA Boston Hlth.care System, Boston, MA, [14]Massachusetts Gen. Hosp., Chestnut Hill, MA, [15]Palo Alto Veterans Inst. for Res., Boston, MA

Abstract:

The rapid expansion of biobanks is fueling breakthroughs in genetics research. Their size and growth allow researchers to analyze massive datasets and contributing important discoveries about genes and health. Analyzing large genetic datasets involves complex calculations, often comparing different segments of the data. Traditional CPU-based tools struggle with these tasks due to memory limitations. GPUs, with their high memory bandwidth and parallelism, offer a faster and more efficient alternative, particularly for large-scale matrix operations in population genomics analysis. High-performance computing resources can further accelerate these efforts. As an example, we implemented GPU-based distributed computational optimizations to the SAIGE (Scalable and Accurate Implementation of Generalized Mixed Model) algorithm. These optimizations enabled a large-scale genome-wide association study (GWAS) across 2,068 phenotypes derived from electronic health records of 635,969 diverse participants from the Veterans Affairs (VA) Million Veteran Program (MVP). Our strategies enabled scaling up the analysis to over 6,000 nodes on the Department of Energy (DOE) Oak Ridge Leadership Computing Facility (OLCF) Summit High-Performance Computer (HPC), resulting in a 20-fold acceleration compared to the baseline model. We are developing additional optimizations to allow users to run multiple phenotypes in parallel for both steps one and two. We provide a Docker container with our optimizations that were successfully used on multiple cloud infrastructures on UK Biobank and All of Us datasets where we showed significant time and cost benefits over the baseline SAIGE model. As GPU computing and advanced AI techniques become seamlessly woven into the biomedical research pipeline, researchers will unlock a new level of understanding. Hidden connections between genes and phenotypes will come to light, paving the way for a deeper comprehension of disease and the path towards personalized medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4141W: Unsupervised phenotyping and genetic discovery in 62,495 cardiac MRIs using autoencoders ⭐

Authors:

S. Ometto[1], S. Chatterjee[1], A. M. Vergani[1], E. Giacopuzzi[1], A. Landini[1], S. Sharapov[1], A. Visconti[2], E. Bianchi[1], F. Cisternino[1], E. Soda[1], F. Ieva[1], E. Di Angelantonio[1], N. Pirastu[1], C. A. Glastonbury[1]; [1]Fondazione Human Technopole, Milan, Italy, [2]Università di Torino, Turin, Italy

Abstract:

Temporal cardiac MRI is an information rich modality, but manually deriving multiple phenotypes of organ function via annotation is a rate limiting step. Autoencoders learn high-level (latent) semantic concepts from images in an unsupervised manner. These latent representations capture complex phenotypes that can then be used to drive novel genetic discovery. We developed a 3D diffusion autoencoder (3D DiffAE) to extract latent phenotypes from four chamber view long-axis cardiac MRIs across 62,495 subjects in UK Biobank (UKB). Latent phenotype-clinical associations were assessed using linear regression. Testing of common and rare (MAF < 1%) SNVs associated with latent phenotypes was performed using REGENIE. Conditional analysis was conducted with COJO and colocalisation using multi-trait coloc. Replication was performed in a held-out subset of UKB. Using 3D DiffAE, we inferred latent phenotypes that are reproducible, heritable (4-18%), and significantly associated with atrial fibrillation (P=8.5e-29), myocardial infarction (P=3.7e-12), and other cardiometabolic outcomes in concurrently measured EHR data. The discovery GWAS (n=47,827) identified 89 genome-wide significant associations in 182 endophenotypes, with 83 showing concordance of effect in replication. Significant variants were mapped to 44 independent loci, of which 7 have not been previously described. By integrating GWAS summary statistics from 56 cardiometabolic traits, we see that 13 loci colocalise (PP4>0.9) with at least one derived cardiac phenotype and 6 with cardiac traits and diseases. Two latent phenotypes shared the same causal SNV (rs422068) with descending aorta area, left and right ventricle stroke volume, pulse pressure and atrial fibrillation. Colocalisation with GTEX eQTLs linked the novel variant rs9388487 to CENPW expression in heart atrial appendage (NES=-0.4, P=1.7e-9). PRS computed using LDPred2 in the UKB non-discovery cohort (n=314,774) statistically improved AUC (DeLong's test) over the baseline (Age, Sex, BMI, Smoking status) for multiple diseases, including myocardial infarction and CAD. Rare variant association using SKATO-ACAT resulted in 17 significant genes. Mice knockouts of 3 of them (ATP5L, PHC3, BTBD8) showed an enlarged heart, and/or abnormal heart morphology. In summary, we performed the largest systematic characterisation of cardiac endophenotypes to date, demonstrating that unsupervised phenotyping of cardiac MRIs using 3D diffusion models capture structural and functional features of the heart. This approach provides a tractable means for novel genetic discovery and improvements in generalised disease risk prediction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4142W: Unveiling the Influence of Reference Sample Size on Ancestry Estimation of Population Structure.

Authors:

S. Tifour[1], A. Price[2], E. Austin[3], K. Marker[4], A. Hendricks[2]; [1]Univ. of Colorado Anschutz, Aurora, CO, [2]Univ. of Colorado Denver, Denver, CO, [3]Univ. of Colorado Denver, Aurora, CO, [4]Univ. of Colorado, Anschutz, Denver, CO

Abstract:

Summary genetic data is valuable for both research and clinical purposes, like prioritizing putative causal variants and polygenic scoring. Yet, this data can mask population structure limiting its utility, especially for admixed or understudied ancestral populations. Methods, such as Summix, accurately and precisely estimate and adjust for ancestry substructure proportions given sufficiently large reference sample sizes. However, it is unknown how well these methods perform when the reference sample size is small, which is common for underrepresented populations. Using a comprehensive range of simulation scenarios, we demonstrate that a reference sample size of ≥150 is needed for accurate (<2% bias) ancestry proportion estimates. Notably, we observed that bias increases inversely proportional to the reference sample size. We aim to leverage this finding to develop a bias correction for small sample sizes improving the utility of Summix. This helps to enable the robust use of a broader set of ancestral populations as reference groups in genetic studies, even in cases where obtaining a large sample size may not be feasible.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4143W: Using an external reference panel and single-variant summary statistics for rare-variant aggregation tests for association

Authors:

B. Ryan[1], R. Welch[1], M. Boehnke[1], C. Fuchsberger[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Eurac Res., Bolzano, Italy

Abstract:

Genome-wide association studies (GWAS) have identified hundreds of thousands of associations between common genetic variants and a range of human diseases and traits. These studies are often of low power to identify associations with rare genetic variants. To increase power, aggregation tests which pool the genetic signal across multiple variants in a gene or other

region of the genome to test the cumulative effect of these variants on a disease or trait are often used. There are many publicly available GWAS summary statistics, and aggregation tests can be performed using only single-variant score statistics, their covariances, and the allele frequencies for each variant. Estimates of single-variant test statistic covariances are required, and are easily obtained with individual-level genetic data. Due to privacy concerns, individual-level genetic data often cannot be shared. Sharing covariance matrices may also be infeasible due to data size. In this study, we apply a previously proposed method of estimating single-variant test statistic covariance from an external reference panel to perform aggregation tests. We propose a two-stage approach by first filtering genes using a null covariance where covariance between all variants is set to zero to perform aggregation tests, and in stage two using an external reference panel to repeat aggregation tests only on those genes passing a p-value threshold. Using the null covariance generally overestimates association significance of genes, allowing us to exclude non-significant genes in the first stage. We investigate the properties of our method across ten traits from the UK Biobank, three variant annotation masks, and three different aggregation tests. When p-values on the log scale produced using individual-level data and our two-stage approach across all ten traits are compared, squared Pearson correlation coefficients are between 0.984 and 0.996 across all scenarios using a reference panel of 1,000 individuals. Using an African American reference panel (n=1,000) from the InPSYght study, we show that this approach is robust to misspecification of the reference panel ancestry. By simulating a meta-analysis of 20 studies from the UK Biobank, we also show that our methods are applicable when estimating covariance to use in meta-analysis aggregation testing. Using a preliminary version of our R package, performing aggregation testing on 1,869 genes using the two-stage approach was on average 20 times faster than estimating the covariance for each gene. This two-stage approach will be a useful tool for researchers wishing to perform rare-variant aggregation tests using publicly available summary statistics.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4144W: Using partial polygenic scores to detect gene-by-ancestry interactions in recently admixed populations ⭐

Authors:

A. Aw, I. Mathieson; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Polygenic scores (PGS) trained on single ancestry cohorts have poor prediction accuracy in recently admixed individuals, for reasons that are not fully understood. Recent work suggests that causal effects are highly correlated across ancestries, so remaining explanations include interaction between tagging variants and local ancestry (e.g. differences in LD patterns) or interactions between tagging variants and global ancestry (i.e. *trans*-interactions).
Here, we describe a novel method, based on polygenic scores computed only from genomic chunks of one ancestry (partial polygenic scores), that quantifies how the effects of tagging variants in admixed individuals depend on local or global ancestry. We define two models; a local model in which the effect of any variant depends only on local ancestry and a global model in which it depends only on global ancestry. We show that the relationship between ancestry and predictive power is the same in both models for standard PGS, but different for partial polygenic scores. In particular, whereas predictive power varies linearly as a function of ancestry under the local model, it varies cubically under the global model. This difference can be used to distinguish the two models. Finally, we show both analytically and through simulations that we can identify mixtures of the two models, parameterized by the proportion of genetic effects that follow the local model (λ_{Loc}).
Applying our method to PGS derived from European ancestry summary statistics from UK Biobank and trained on PMBB European-ancestry individuals (n = 29,410), we recover established drivers of poor portability of PGSs within PMBB African Americans (n = 9,324). These include the large allele frequency difference between Africans and Europeans at the 1q23 DARC locus that determines monocyte and neutrophil counts (monocyte: λ_{Loc} = 0.8; neutrophil: λ_{Loc} = 0.6). For other traits, including BMI (λ_{Loc} = 0.41) and standing height (λ_{Loc} = 0.52), we estimate that the contributions of local and global ancestry to differences in predictive power are roughly equal.
In summary, our approach shows that partial polygenic scores can distinguish the impact of local and global ancestry on PGS performance, and reveals an unexpectedly high contribution of gene-by-global ancestry interactions to differences in predictive power across ancestries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4145W: VACANT: Variant Annotation Clustering AssociatioN Test

Authors:

S-H. Cho, C. Huff; Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract:

Advancements in sequencing technologies have enabled the investigation of rare genetic
Advancements in sequencing technologies have enabled large-scale investigations of the contributions of rare genetic variation to complex genetic diseases. Due to the low frequency of these variants, large sample sizes and/or substantial effect sizes are often required. Numerous gene-based association tests have been developed to provide increased statistical power over single variant tests by aggregating information across multiple variants. While a few of these tests integrate functional variant annotation weights to enhance power, the broad majority do not consider a priori information regarding the predicted functional consequences of the variants evaluated. Additionally, many gene- based tests are not robust in the presence of substantial imbalances in case- control sample sizes, which are commonly encountered in biobank-scale data.
To address these limitations, we have developed the Variant Annotation Clustering AssociatioN Test (VACANT). VACANT is an analytical framework for gene-based testing that incorporates quantitative functional annotations to account for the deleteriousness of each variant. By optimally grouping variants based on annotation scores, VACANT increases statistical power compared to tests that do not consider predicted variant severity. Furthermore, by employing an appropriate regression model, VACANT is scalable for analyzing biobank-scale sequencing association studies of both continuous and dichotomous traits, regardless of case-control ratio imbalances.
Currently, we are comparing VACANT to existing methods, including VAAST, SKAT-O, and RVT2. Preliminary results indicate that when achieving 80% statistical power with a downsampled case number of 3,000 and control number of 40,000, our approach outperforms these methods by at least 20% for well-known breast cancer genes such as ATM, BRCA1, BRCA2, and PALB2, considering rare variant signals with a minor allele frequency (MAF) smaller than 0.5%. VACANT thus provides a robust approach for conducting rare variant association studies on both existing datasets and the very large datasets anticipated in the future.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4147W: Validating the effects of genetic variants on protein level:Link between whole genome sequencing and proteomic data in UK Biobank

Authors:

T. Zou, M. Parker, L. Barrera; Beam Therapeutics, Cambridge, MA

Abstract:

Background: Accurately predicting the biological impact of both genetic variants and off-target edits remains challenging. Although many variant effect prediction tools exist, their validation on functional data is limited. We hypothesized protein level could be a readout informing variant function and characterized whole genome sequencing (WGS) variants associated with protein abundance.
Methods: We aggregated predicted loss-of-function (pLOF) variants in 88 genes on chromosome X into a gene set and tested the associations with their corresponding protein level in ~50k UK Biobank participants. We focused on X chromosome genes because males that carry LOF variants are expected to have low to zero protein. Next, we tested the associations between independent WGS variants located ±250kb of those significant pLOF genes and protein level in males only (n=15,986), and annotated the functional consequence of tested variants using CADD.
Results: Among 88 tested genes on chromosome X, pLOF variants in 25 genes were significantly associated with reduced protein level (P<0.00056), which were used in the following variant-level tests. Of 27,357 independent variants tested, 216 (0.8%) were significant (P<1.8*10-6). Variants associated with reduced protein level were likely to be annotated as functional. For example, 81.8% of stop gained, 62.5% of frameshift, 28.6% of splicing, and 10.3% of non-synonymous variants were significantly associated with reduced protein level, but only 0.5% of intronic variants were significant, which was similar to intergenic variants. Consistent with expectation, on average rare variants had larger effect sizes than common variants, and stop gained variants had largest effect sizes, followed by frameshift, splice site and non-synonymous variants. There were no clear positional patterns among significant variants as they could occur anywhere throughout a gene. Overall, CADD score was good at predicting significant functional variants, with 48.5% having CADD score >20, but only 6.8% of significant intronic variants had CADD score >20. One limitation of this work is having limited power to detect associations and draw conclusions for rare variants (minor allele count <10).
Conclusions: Using paired WGS and proteomic data, we successfully validated the effects of genetic variants on protein level. Our findings suggest that plasma proteomics data can be a readout for variant function, intronic variants are unlikely to impact protein levels, and that more accurate prediction tools for intronic variants are needed. These results can help inform the relative risk from different types of off-target edits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4148W: VINTAGE: A unified framework integrating gene expression mapping studies with genome-wide association studies for detecting and deciphering gene-trait associations

Authors:

Z. Li, B. Gao, X. Zhou; Univ. of Michigan, Ann Arbor, MI

Abstract:

Integrative analysis of genome-wide association studies (GWAS) with gene expression mapping studies holds the promise to unravel the molecular mechanisms underlying disease etiology. Here, we present VINTAGE, a unified statistical framework for such integrative analysis to identify and decipher gene-trait associations. VINTAGE explicitly quantifies and tests the proportion of genetic effects on a trait potentially mediated through gene expression using a local genetic correlation test, and further leverages such information to guide the integration of gene expression mapping study towards gene association mapping in GWAS through a genetic variance test. The explicit quantification of local genetic correlation in VINTAGE allows its gene association test to unify two seemingly unrelated methods, the sequence kernel association test (SKAT) and the transcriptome-wide association study (TWAS), into the same analytic framework and include both as special cases, thus achieving robust performance across a range of scenarios. The combination of the genetic variance test and local genetic correlation test allows us to assess the gene-trait relationship from two distinct angles, providing a comprehensive characterization of these relationships. We illustrate the benefits of VINTAGE through extensive simulations and integrative analyses of a large-scale cis-eQTL study from eQTLGen with eighteen complex traits from UK Biobank. In real data, VINTAGE not only improves the power for detecting gene-trait associations by an average of 5% and 87% over SKAT and TWAS, respectively, but, more importantly, stands out as the only effective method for assessing potential mediation effects of gene expression in the identified associations through the local genetic correlation test. Specifically, VINTAGE reveals that while most genes do not exhibit detectable mediation effects, with the median proportion of potential mediation estimated at only 12% across genes, there are 61 genes that display significant local genetic correlation with potential mediation effects across the examined traits. Among all the identifications, *CAMK1D* likely influences the systolic blood pressure (SBP) through gene expression regulation, with the proportion of mediation effects estimated to be 65%. In contrast, SNPs in the genic and adjacent regulatory regions of *FURIN* likely influence SBP through mechanisms other than the expression changes in *FURIN*, as indicated by the estimated proportion of mediation of 0.43%. Overall, VINTAGE serves as an effective tool for evaluating

the mediation effects of gene expression and comprehensively characterizing gene-trait relationships.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Wednesday Session

Board 4149W: XPRS: A Tool for Interpretable and Explainable Polygenic Risk Score

Authors:

N. Kim, S. Lee; Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

Polygenic Risk Scores (PRS) have emerged as crucial methods for assessing genetic susceptibility to various diseases. Despite their potential, the clinical utility of PRS can be limited by the lack of tools that help to explain PRS. To address these issues, we introduce eXplainable PRS (XPRS), a software designed to enhance the interpretability of PRS by decomposing them into gene/region and SNP contribution scores. By utilizing Shapley Additive Explanations (SHAP), XPRS calculates the attributed score of each gene or region, providing detailed insights into which genes significantly contribute to an individual's PRS. This decomposition allows for a more granular explanation, enabling the examination of specific SNPs within genes to determine their individual impact on PRS.

XPRS is composed of a multilevel visualization approach. At the population level, Manhattan plots and tables highlight important genes of PRS based on the highest variance in gene contribution scores. At the individual level, XPRS visualizes attributed gene values to pinpoint risk genes that drive the PRS value of the given individual and employs LocusZoom plots to show which SNPs influence those genes.

The software is implemented with a user-friendly web interface, allowing easy data input. Core computations are optimized with C++, and R is used for visualization and data preprocessing, leveraging its robust statistical and graphical capabilities. This integrated approach ensures that XPRS is both accessible and efficient in handling complex computational tasks.

In conclusion, XPRS can bridge the gap between complex genetic data and actionable clinical insights. Its ability to provide detailed, interpretable genetic risk assessments can pave the way for more precise and personalized healthcare interventions.

Thursday, November 7, 2024, 2:30 pm - 4:30 pm

Session Title: Cancer Poster Thursday Session

Board 8001T: A comparison of the diagnostic yield of comprehensive multigene panel testing with RNA analysis in hereditary cancer to smaller guidelines-based panels

Authors:

L. Fosler[1], B. Johnson[2], **S. Hamilton**[3], E. Chen[3], W. Korn[4]; [1]Invitae, Granger, IN, [2]Invitae, Fort Lauderdale, FL, [3]Invitae, San Francisco, CA, [4]Invitae, San Anselmo, CA

Abstract:

Introduction: Previous research has shown that the diagnostic yield of multigene panel testing (MGPT) for hereditary cancer is higher when compared to smaller guidelines-based panel testing. Research has also shown that RNA analysis in hereditary cancer can increase diagnostic yield by improving variant classification and identifying individuals with a deep intronic variant outside the standard DNA reportable range. We sought to analyze and compare the diagnostic yields of guidelines-based panel testing to MGPT with RNA analysis in individuals pursuing hereditary cancer testing. **Methods:** Individuals who underwent hereditary cancer MGPT for 84 genes and RNA analysis for 63 genes at Invitae were identified. Individuals with 2 clinician-reported test indications were included in the analysis: 1) personal and/or family history (FHx) of breast cancer (BC) and 2) personal and/or FHx of colorectal cancer (CRC) and/or polyposis. Positive rates of pathogenic/likely pathogenic (P/LP) variants were calculated for genes included on a 13-gene BC or 20-gene CRC guidelines-based panel for the respective indication and then compared to positive rates for genes included on the MGPT. Positive rates for RNA analysis were analyzed and compared. Carriers of mono-allelic P/LP variants were excluded from the analysis.**Results:** The total number of individuals with a BC indication who underwent testing was 16,092. Limiting the analysis to a BC guidelines-based panel, the positive rate was 6.8% (1,089/16,092), compared to 9.6% (1,537/16,092) by MGPT testing. Only 7 individuals had a deep intronic P/LP variant or VUS that was upgraded with RNA analysis increasing the diagnostic yield by 0.14%. There were 4,738 individuals with an indication of CRC and/or polyposis who underwent testing. The diagnostic yield of a CRC guidelines-based panel was 6.8% (323/4,738) compared to 12.3% (582/4,738) with MGPT. RNA analysis increased the yield by only 0.25% (12/4,738). **Conclusion:** These data show that utilization of MGPT compared to limited analysis using guideline-based panels increases the diagnostic yield by 41% in individuals with a personal and/or FHx of BC and 81% in individuals with a personal and/or FHx of CRC and/or polyposis, however RNA analysis has a minimal impact on diagnostic yield.

Session Title: Cancer Poster Thursday Session

Board 8002T: A highly accurate risk factor-based multiethnic XGBoost model for identifying patients with skin cancer

Authors:

M. D'Antonio[1], K. A. Frazer[2]; [1]Univ. of California, San Diego, La Jolla, CA, [2]Univ. of California, San Diego, LA JOLLA, CA

Abstract:

Understanding associations between genetics, lifestyle, social determinants of health (SDOH), and skin cancer onset holds profound significance for clinicians and patients. With the availability of the **All of Us** (AoU) dataset, elucidating these relationships to enhance clinical decision-making is increasingly feasible. We explored the relationships between skin cancer, genetic ancestry, lifestyle, SDOH, and PDE5A inhibitor use in ~400,000 AoU participants. We developed an XGBoost machine learning model for predicting skin cancer across diverse populations and show how multiple factors influence skin cancer: **Genetics**: Skin cancer predominantly affects individuals of European ancestry, but non-Europeans often present with more advanced stages and poorer survival rates. These disparities are attributed to variations in skin pigmentation and vitamin D production, which are genetically determined. Our study incorporated genetic data into the prediction model, revealing that Admixed and Hispanic individuals with skin cancer have a greater fraction of European ancestry, suggesting that different genetic factors may influence skin cancer risk across populations. **Lifestyle and SDOH**: Assessing the impact of non-genetic factors on skin cancer risk presents inherent complexity, given the intricate web of interconnected variables at play. Wealthier individuals show lower incidence of skin cancer due to better health awareness and early detection. Socioeconomic factors profoundly impact skin cancer outcomes, highlighting the need for strategies to address disparities and enhance early detection. Use of **PDE5A inhibitors**: Multiple studies associate PDE5A inhibitors with increased skin cancer risk, particularly among those with a cancer history. However, a causal link is difficult to establish due to the interconnectedness of lifestyle and socioeconomic factors. Our model shows that PDE5A inhibitor use is independently associated with skin cancer risk in Europeans, contributing to predictive accuracy. This study integrates genetic and non-genetic determinants to stratify individuals based on skin cancer risk. By combining genetics with lifestyle and socioeconomic factors, we provide valuable tools for early detection and personalized healthcare interventions.

Session Title: Cancer Poster Thursday Session

Board 8003T: A multi-omic approach to genetic variants in breast cancer: Insights from the Norwegian Breast Cancer Study

Authors:

I. Brorson[1], D. O'Mahony[2], S. D. Bos-Haugen[3], A. Sundaram[1], P. Wernhoff[2], G. I. Grenaker Alnæs[2], V. Kristensen[1]; [1]Univ. of Oslo, Oslo, Norway, [2]Oslo Univ. Hosp., Oslo, Norway, [3]Cancer Registry of Norway, Oslo, Norway

Abstract:

Breast cancer is the most frequently diagnosed cancer globally, with approximately 2.3 million new cases and about 685,000 deaths recorded in 2020. Genetic susceptibility accounts for five to ten percent of breast cancer cases, making it one of the most significant risk factors for the disease. Most genetic predisposition is believed to result from the combined effects of single nucleotide polymorphisms (SNPs). Individually, these SNPs increase risk by less than 1.5-fold but collectively contribute significantly to breast cancer susceptibility. Genome-wide association studies (GWAS) have identified 210 breast cancer susceptibility loci that cumulatively explain approximately 18.3% of the familial relative risk of breast cancer, with a notable overrepresentation in non-coding regions. Analyzing expression data, DNA methylation information, and regulatory RNA elements can help prioritize associated variants into quantitative trait loci (QTLs) to delve deeper into the functional mechanisms of these variants. This project aims to further investigate breast cancer genetic variants and their intricate connections to gene regulation, shedding light on their roles in breast cancer. We aim to predict causal germline variants located in genetic regions known to harbor a higher proportion of variants with regulatory impacts in association to breast cancer disease. Specifically, we will integrate multi-omic QTL layers to assess interactions in relation to breast cancer variant causality.

The Norwegian breast cancer study (NBCS) dataset is based on 2,980 individuals and comprises of germline genotyped and somatic data, characterized of several finely-curated omic layers. These molecular layers include DNA methylation, gene expression, copy number alteration, micro RNA (miRNA) and long non-coding RNA (lncRNA). Cis-acting quantitative trait loci (QTLs) have been predicted using germline genotypes analysed with the different molecular layers from tumour. We have identified 1,175 expression-, 222 lncRNA-, 116 miRNA- and 30,413 methylation QTLs with a false discovery rate (FDR) < 0.05 using the QTL mapper TensorQTL. We have replicated well-known breast cancer QTLs as well as identified novel ones. The somatic QTLs will be combined with the corresponding germline QTLs measured in blood to detect putative 'super' QTLs. We intend to model the QTL interactions within our dataset to analyse

the complex interactions between genetic variants and multi-omics expression in a comprehensive manner.

Session Title: Cancer Poster Thursday Session

Board 8004T: Analytical Evaluation of Whole Genome Sequencing for Acute Myeloid Leukemia

Authors:

W. Gong[1], G. M. Tagliazucchi[1], M. Ghildiyal[1], M. Chavez[1], K. Nobuta[1], E. Sayyari[1], S. Truong[1], V. Teng[1], G. Kim[1], B. Quyen[1], D. Spencer[2], S. Kim[1], J. Bernd[1], E. de Feo[1], S. Comer[1]; [1]Illumina, Inc., San Diego, CA, [2]Washington Univ. in St. Louis, St. Louis, MO

Abstract:

Background
Acute myeloid leukemia (AML) is the most common leukemia in adults. Current methods described by World Health Organization 2022 and the European Leukemia Net 2022 rely on cytogenic and molecular profiling for AML classification. Recent advances have demonstrated that next generation sequencing may improve prognostic prediction and risk stratification of AML patients. A tumour-only high coverage whole genome sequencing method was developed to assess the analytical performance (limit of detection, sensitivity, specificity and reproducibility) in identifying somatic small variants, structural variants (SV) and copy number alterations (CNA) specific to AML.
Method
Seracare® myeloid panel, NOMO-1, Kasumi-1, HCC1187 and AML clinical specimens were used to assess the analytical performance of this novel method. Extracted genomic DNA (~350ng) from contrived specimens, cell-lines, mononuclear cells from whole blood or bone marrow aspirate were manually processed by Illumina DNA PCR-Free Prep, Tagmentation and sequenced at 4-plex on S4 flowcells using NovaSeq™ 6000. The sequence data was analysed by DRAGEN™ Somatic v4.2 with a custom virtual 156 AML specific gene panel variant classification algorithm.
Results
Although the target sequence was on average ~200X, samples were in silico down sampled to 140X to establish the following performance specifications. Limit of detection was established for small variant (SNVs and indels) and SV to be 5% and 7.3%, respectively. CNA were detected with copy number fold change of 1.085 and 0.87 for duplication and deletion events, respectively. Further, loss of heterozygosity was detected with purity of 17%. Overall, we

observed accuracy >95% across clinical samples with sensitivities >95%, >99% and >95% for small variants, SV and CNA, respectively and specificity of >99.9%.

Conclusion

We have developed a new tumour-only whole genome sequencing method for genomic profiling of AML patients with an expected turn-around-time of 5 days. Like previous findings, the analytical performance shown here indicates high accuracy in identifying variants indicative for AML. In addition, the higher sequence depth of this method provides improved sensitivity in detecting prognostic and risk stratification biomarkers compared to other genome sequencing or cytogenetic methods.

Session Title: Cancer Poster Thursday Session

Board 8005T: Beyond *RB1* : analysis of somatic copy number alterations in retinoblastoma

Authors:

P-s. Lai, G. Tan, C. Liu, T. Quah; Natl. Univ. of Singapore, Singapore, Singapore

Abstract:

Background: Retinoblastoma (RB) is a rare childhood retinal cancer caused by biallelic loss of *RB1* gene. Although the somatic mutation burden appears low in RB compared to other cancers, observations of other molecular alterations such as copy number changes, translocations, promoter hypermethylation, etc. suggest that there may be other genes beyond *RB1* involved in tumor progression and/or prognosis. In this study, we assessed genome-wide copy number changes in RB tumors and characterized them for patterns of allelic loss and imbalance. We also investigated whether suppressor of tumorigenesis and/or proliferation (STOP) genes and Copy-number alterations Yielding Cancer liabilities Owing to Partial loss (CYCLOPS) genes are associated with genomic copy-number loss using a high-density SNP BeadChip array. Results: Homozygous Copy Number Variations (CNV) losses, Loss of Heterozygosity (LOH) and copy neutral LOH changes were detected via the Allele-Specific Copy number Analysis of Tumors (ASCAT) algorithm. Average tumor ploidy across the cases was found to be 2.3, with mean tumor purity of 78% across the same cohort. To identify regions with a statistically high frequency of copy number aberrations over background aberration, we used the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm. Recurrent large scale copy number variations were observed in chromosomes 1, 2, 3, 4, 6, 12, 13, 15, 16 and 19, whereby focal copy number losses (<3Mb) were located at 2q, 4q, 5p, 6p, 7q, 8q, 9q, 10q, 12q, 15q, 16q & 17q. Overall, copy number losses in bilateral tumors (57.6%, 152/264) were slightly

higher than in unilateral tumors (42.4%, 112/264). Average incidence of copy number loss was 22 per tumor with 25.3 losses in bilateral tumors and 18.6 losses in unilateral tumors. Patterns of uniparental disomy or copy neutral LOH in chromosome 13 were observed in 15% of tumors. The minimal common region of genomic loss (<3Mb) was restricted to 19q, 16p and 16q. The STOP genes subject to copy number loss in these regions
were *USPL1*, *NUDL1*, *KIAA0513*, *GCSH* and *ARID3A* while two CYCLOPS genes,
namely *RPS15* and *ALOX5AP* were identified. Our results show that the genomic landscape of retinoblastoma may be more complex than originally thought and the recurrent presence of deleted STOP and CYCLOPS genes suggest that they may contribute towards development of RB. Clinical significance of these genes needs to be further evaluated with clinicopathologic correlations. Conclusions: Somatic copy number alterations in RB appears to be quite distinct from other cancers. Some CYCLOPS and STOP genes may be involved in RB and could be potential biomarkers of cancer drivers.

Session Title: Cancer Poster Thursday Session

Board 8006T: Biologically-targeted discovery-replication scan identifies G×G interaction in relation to risk of Barrett's esophagus and esophageal adenocarcinoma

Authors:

M. Buas[1], Q. He[2], S. P. Verma[1], X. Zhang[1], G. Ann-Sophie[3], C. Maj[3], E. Naderi[1], K. Graz[1], J. Chen[4], P. Gharahkhani[5], X. Shu[1], K. Offit[1], S. MacGregor[5], T. Vaughan[6], J. Schumacher[7], L. Yan[4]; [1]Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Fred Hutch Cancer Ctr., Seattle, WA, [3]Univ. Hosp. of Marburg, Marburg, Germany, [4]Roswell Park Comprehensive Cancer Ctr., Buffalo, NY, [5]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [6]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [7]Univ. of Marburg, Germany, Marburg, Germany

Abstract:

Inherited genetics represents an important contributor to risk of esophageal adenocarcinoma (EAC), and its precursor Barrett's esophagus (BE). While genome-wide association studies (GWAS) have identified ~30 genetic variants associated with BE/EAC susceptibility, genetic interactions have remained unexamined. To address long-standing analytical challenges, we combined knowledge-guided filtering for 'search space' reduction and penalized regression for efficient variable selection. Discovery scans were restricted to genes with (A) known/plausible links to BE/EAC pathogenesis (n=493) or (B) prior evidence of biological interactions (n=4,196). ~75 x 106 pairwise SNP×SNP interactions were screened via the hierarchical group lasso method

glinternet using BEACON GWAS data (5,810 BE/EAC cases; 3,207 controls). The top ~2000 interactions retained in each scan were prioritized using LRT P values (P_{int}) from single logistic models. Identical scans were repeated among males only (78%), with two independent GWAS datasets used for replication (5,100 BE/EAC cases; 6,326 controls). In the overall and male-specific primary replication, 11 of 187 and 20 of 191 interactions satisfied Pint<0.05, respectively. The strongest evidence for secondary replication was for rs17744726×rs7218480 among males (P_{int}=0.051), with consistent directionality of interaction across all cohorts ($P_{meta}=2.19\times10^{-8}$). These variants map to an intron of the kinase gene BLK and the 3'UTR of CDKN2B, respectively, with strong eQTL relationships reported in esophagus or other cell types. Prior evidence suggests that the CDKN2B variant may influence CDKN2B expression levels via modulation of micro-RNA binding. At the BLK locus, we identified a nearby proxy variant (r^2=1) for rs17744726 within a putative enhancer region in GI tract tissues, based on informatics-driven prioritization of likely functional/causal SNPs. An ~800-bp genomic fragment encompassing this variant was found to exhibit allele-specific enhancer activity in luciferase reporter assays in esophageal cell lines. Existing enhancer-gene maps based on computational predictions and chromatin interaction profiles implicated BLK, FAM167A, and GATA4 as plausible gene targets. This study represents the first systematic investigation of genetic interactions in BE/EAC susceptibility. Our results indicate differential risk associations for CDKN2B variation, by BLK genotype, suggesting a novel dependency between two established loci in the genetic architecture of BE/EAC.

Session Title: Cancer Poster Thursday Session

Board 8007T: Case report: a deep-intronic and splice-altering variant in the *ATM* gene

Authors:

L. Elemans[1], E. Kroon[1], T. van Ham[1], R. Meijers[2], R. van Minkelen[1], T. van Dijk[1], E. Korpershoek[1]; [1]Dept. of Clinical Genetics, Erasmus Med. Ctr., Rotterdam, Netherlands, [2]Dept. of Pathology, Erasmus Med. Ctr., Rotterdam, Netherlands

Abstract:

Background and Objective: Deep-intronic non-coding variants can have significant biological effects, yet they are typically not detected using exome sequencing and classification of these variants remains challenging. We present a case of a 68-year-old male, diagnosed with pancreatic cancer and liver metastases. A multi-platform approach was utilized to interpret the effect of a deep-intronic *ATM* variant, ultimately to pursue a genetic diagnosis for this patient

and meet the criteria for inclusion in a national targeted anti-cancer treatment study (DRUP). **Methods:** Next-Generation Sequencing (NGS) of a panel of genes associated with hereditary pancreatic cancer was performed on germline DNA. The detected variants were filtered and annotated with Alissa Interpret software. Variants were evaluated and classified using Alamut Visual and the ACMG/AMP Variant Interpretation Guidelines for *ATM* (v1.1). RNA from blood (PAXgene) was analyzed by RT-PCR and Sanger sequencing. Additionally, NGS and Sanger sequencing was performed on tumor DNA isolated from liver tissue, to evaluate loss of the wild type allele. **Results:** Initial investigation of germline DNA showed heterozygosity for a deep intronic variant in *ATM*, NM_000051.3(ATM):c.3994-160T>C, p.?, which could be associated with the patient's medical diagnosis of pancreatic cancer. *In silico* splice site predictors indicated loss of a cryptic donor splice site for this *ATM* variant and gain of a SF2/ASF exon splicing enhancer. According to the Clinvar archive, the *ATM* variant was previously found to induce aberrant mRNA splicing. We confirmed this finding by RT-PCR and Sanger sequencing, which showed the formation of a pseudo-exon through cryptic intron retention (NM_000051.3(ATM):r.3993_3994ins3994-190_3994-79,p.(Ile1332fs)) and partial intron retention (NM_000051.3(ATM):r.3993_3994ins3994-190_3994-1,p.(Ile1332fs)), leading to a frameshift and a premature stop codon. Lastly, NGS and Sanger sequencing of tumor DNA showed loss of the wild type allele. **Conclusion:** Based on our findings, the deep-intronic *ATM* variant was classified as likely pathogenic. This genetic diagnosis supported the patient's eligibility for treatment within the DRUP study. This case report highlights the strength of combined panel-NGS and targeted RNA analysis, to achieve accurate genetic diagnosis of deep-intronic variants and support personalized cancer treatment.

Session Title: Cancer Poster Thursday Session

Board 8008T: Characterization of 397T/C, 351A/G polymorphisms of ERα gene and 331G/A polymorphism of PR gene as genetic markers of breast cancer predisposition in Cameroon

Authors:

N. Nguedia Kaze; Covenant Univ., Ota, Nigeria

Abstract:

Background: Breast cancer is a real public health problem in Cameroon, where more patients with this cancer usually die a year after diagnosis, as it is still based on histological examination, mortality due to that cancer is far from decreasing. Since cancer is an accumulation of molecular changes, the +331 *G/A* polymorphism of *PR* gene and 397*T/C* and

351*A/G* polymorphisms of *ER* alpha gene could be investigated as etiological factors. **Objective**: We studied the association between polymorphisms in the estrogen receptor alpha (*ER* alpha) gene and Progesterone Receptor gene (PR) and breast cancer. **Methods**: We carried out a case control study, in which 16 cases diagnosed positive for breast cancer at the Yaoundé general hospital was recruited by consulting the patient file and Blood samples was also collected from them and 22 healthy women who were recruited using a questionnaire and an inform consent signed each of them. The +331 *G/A* polymorphism in the *PR* gene was identified using NIaIV endonuclease and by PCR-RFLP and direct molecular haplotyping was used to determine the relationship between the 2 polymorphisms in the *Er* alpha gene, 397*T/C* and 351*A/G* by restriction fragment-length polymorphisms. The data were analyzed using Statistical Package Social Sciences v20. **Results**: The study population was mostly postmenopausal with an invasive ductal carcinoma with an average age of 53+/-11,08 years. A mutant *A//A* genotypic frequency of the *PR* gene was high in both, the case and control groups 100% and 94% respectively. We obtained an Odds Ratio of 1.067 with 95% Confidence Interval of 0.940-1.211, making the mutation associated to the risk of developing breast cancer. In addition, we identified the mutant *CG* haplotype allele of the *ER* alpha gene with a predominance in the cancer group exclusively with a frequency of 56.25% with an OR of 2.022 with a 95% CI 0.432 - 9.461 associated with breast cancer. However, the *CG//CG* mutant genotype of the *ER* alpha gene (31.25%) in the diseased population was significantly associated with breast cancer with an OR 1.455, CI 1.045-2.024 and a p-value of 0.05. **Conclusion**: This shows that the polymorphisms studied would predispose to a risk of developing breast cancer.

Session Title: Cancer Poster Thursday Session

Board 8009T: Clinicogenomics data guide targeted cancer therapies for improved clinical outcomes

Authors:

R. Sasik, C. Nasamran, D. Chilin-Fuentes, T. Whisenant, S. B. Rosenthal, K. Fisch; UC San Diego, La Jolla, CA

Abstract:

Clinicogenomics is the integration of real-world clinical datasets with genetics or other omics data, from clinically indicated tests performed during care. Clinicogenomics data are used for determining appropriate targeted therapeutics based on observed genomic alterations, for determining disease subtyping, and for molecularly guided treatment in general.
The large amount of clinical genetics data accumulated at the UCSD Moores Cancer Center

motivated the development of ARMOR: the Aggregated Registry for Molecular Oncology Research. The latest version of ARMOR from 2024 contains clinical genetics data from over 8,000 patients from four vendors (Foundation Medicine, Tempus, Caris and Guardant). Access is provided under a data use agreement for approved research use of the ARMOR database through a cBioPortal interface.
Clinicogenomics data in ARMOR allows discovery of significant correlations between patient survival and covariates including cancer type, somatic genotype or subtype of tumors, tumor mutational burden, and most importantly, therapeutics. For example, cancer patients of all types treated with immune checkpoint inhibitors (N=1,414), whose tumor mutational burden is in the top 20 percentile, have significantly longer survival time than those in the lower 80 percentile, with nominal p-value < 10^(-14). On the other hand, colorectal cancer patients treated with immune checkpoint inhibitors (N=193) have significantly shorter survival time (nominal p<0.02) compared to those given traditional chemotherapy (N=1083). Also, glioma patients with somatic mutations in IDH1 or IDH2 genes (N=65) have significantly longer survival (p<10^(-6)) than those without (N=185), regardless of treatment type. As these are observational studies and the patients were not randomized in the two study arms, the nominal p-values may not be interpreted as probabilities applicable to the general population. However, this observational evidence may nevertheless be used to inform clinical decision making and inform research studies using real-world clinicogenomics data.

Session Title: Cancer Poster Thursday Session

Board 8010T: Cmai: predicting antigen-antibody interactions from massive sequencing data

Authors:

B. Song[1], K. Wang[2], S. Na[3], J. Yao[1], F. J. Fattah[1], M. S. von Itzstein[1], D. Yang[1], J. Liu[1], Y. Xue[1], C. Liang[1], Y. Guo[3], I. Raman[1], C. Zhu[1], J. Dowell[1], J. Homsi[1], S. Rashdan[1], S. Yang[1], M. Gwin[1], D. Hsieh[1], Y. Gloria-Mccutchen[1], P. Raj[4], X. Bai[1], J. Wang[5], j. conejo-garcia[6], X. Yang[1], G. David[1], J. Huang[3], T. Wang[1]; [1]Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [2]Southern Methodist Univ., Dallas, TX, [3]Univ. of Texas at Arlington, Arlington, TX, [4]UT Southwestern Med Ctr, Dallas, TX, [5]New York Univ. Grossman Sch. of Med., New York, NY, [6]Duke Sch. of Med., Durham, NC

Abstract:

The interaction between antigens and antibodies is the key step underlying the function of the humoral immune system in various biological contexts. However, current experimental approaches for profiling antibody-antigen interactions are costly and time-consuming, and can

only achieve low-to-mid throughput. On the other hand, bioinformatics tools in the field of antibody informatics mostly focus on optimization of antibodies given known binding antigens, which is a related but very different research question. In this work, we developed an Artificial Intelligence tool, Cmai, to address the prediction of the binding between antibodies and antigens. Cmai achieved an AUROC of 0.91 in our validation cohort, and recognizes amino acids of the antibodies and antigens that are in contact. Cmai found that, during immune-related adverse events (irAEs) caused by immune-checkpoint inhibitor (ICI) treatment, the humoral immunity is preferentially responsive to intracellular antigens from the organs affected by the irAEs. In contrast, extracellular antigens on malignant tumor cells are inducing B cell infiltrations, and the infiltrating B cells have a greater tendency to co-localize with tumor cells expressing these antigens in the tumor microenvironment, according to Cmai's predictions. We further found that the abundance of tumor antigen-targeting antibodies is predictive of ICI treatment response. Overall, this work provided a powerful tool to profile the landscape of antigen-antibody interactions in biological samples, which reveals novel insights at unprecedented levels and will yield powerful tools for translational development.

Session Title: Cancer Poster Thursday Session

Board 8011T: Comprehensive analysis of rare genetic variants in hepatocellular carcinoma across diverse population groups

Authors:

A. Garofalo[1], P. Chotiprasidhi[2], J. Ma[2], K. Sato-Espinoza[2], L. Guare[1], K. Cardone[1], T. Odia[1], S. Antwi[3], K. Wangensteen[2], A. Verma[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Mayo Clinic, Rochester, MN, [3]Mayo Clinic, Jacksonville, FL

Abstract:

Liver cancer is the third leading cause of cancer death worldwide, with 906,000 new cases and 830,000 deaths in 2020. Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer. Several genome-wide association studies have identified common variants associated with HCC. Previous work has reported a high rate (11-14%) of pathogenic or likely pathogenic (P/LP) germline variants in cancer predisposition genes in patients with HCC. However, the contribution of rare genetic variants remains unclear. To better understand the association and effect of rare variants on the risk of HCC, we conducted an exome-wide analysis of rare coding variants in 329 cases and 22,965 cancer-free and cirrhosis controls of European and African ancestry in Penn Medicine Biobank (PMBB). Our approach used a gene burden analysis framework, in which we aggregated rare coding variants within genes. These variants

were categorized based on several predictive algorithms into four groups: 1) predicted loss of function, 2) damaging or deleterious missense variants, 3) other missense variants and 4) synonymous variants. We utilized SAIGE-GENE+ to perform a gene-burden regression analysis for each variant category, using different allele frequency thresholds to assess the burden of rare and ultra-rare variants independently. All models were adjusted for sex, age, and the first 10 principal components to control for population stratification. Among the European population group, *MS4A12* (OR=1.17 [1.09-1.25], P=7.638 x10^{-6}) was the most significant gene associated with HCC when compared to cancer-free controls. Among the African population group, *RHOU* (OR= 2.33 [1.62-3.34], P=4.413 x10^{-6}) was the most significant gene associated with HCC when compared to cancer-free controls. These findings lay a foundation for subsequent investigations to confirm these associations and explore their potential roles in HCC pathogenesis. Our future direction is to expand this analysis to other biobanks, such as Mayo Clinic, All of Us, and UK Biobank, and conduct a meta-analysis, which will improve the power to detect associations.

Session Title: Cancer Poster Thursday Session

Board 8012T: Cosegregation of *NF1* mutation and *PMS2* Structural Variant with Hodgkin Lymphoma- A Rare Event

Authors:

I. Akalin[1], I. Sunar[1], I. Kahramanoglu[2]; [1]Metagentech Ctr. for Genetic Diseases, Istanbul, Turkey, [2]Ilker Kahramanoglu Obstetrics and Gynecologic Oncology Ctr., Istanbul, Turkey

Abstract:

Hodgkin lymphoma (HL) is a rare curable lymphoid neoplasm which mainly affects lymphatic system of the body. There are different types of Hodgkin lymphoma, but the classical type accounts for 95% of the cases.
PMS2 is one of the most important genes involved in mismatch repair mechanism, thus disrupted function of this gene is associated with Constitutional Mismatch Repair Deficiency (CMMRD) disease. As a result of DNA mismatch repair disorders resulting from changes in the *PMS2* gene, "cafe-au-lait" spots, axillary freckles and other neurofibromatosis-like appearances and adenomatous colonic polyps may occur in affected individuals. Individuals may be predisposed to many malignancies including lymphoma. Here we report a 39 year-old male patient with Hodgkin lymphoma in remission with no relapse who is one of the collaborating authors. He was first diagnosed with HL two years ago. He was born to a non-consangineous family without history of tumors other than his mother with glioblastoma. He was admitted to

genetic counseling for possible germline susceptibility for cancer. Whole Exome Analysis (WES) with Copy Number Variation (CNV) analysis was carried out. WES revealed c.3883A&gtG:p.T1295A heterozygous mutation on *NF1* gene which is associated with Neurofibromatozis type 1. Copy number analysis revealed a deletion at 7p22.1 (chr7:5973249-5973692) that is containing exon 15 of *PMS2* gene with zero copy number. In a recently published article, coexistence of both *PMS2* and *NF1* mutations in individuals was reported to have glioblastoma, anaplastic astrocytoma and T-lymphoblastic lymphoma. Here we report a novel association of *PMS2* structural variant and *NF1* point mutation with Hodgkin lymphoma for the first time up to the literature.

Session Title: Cancer Poster Thursday Session

Board 8013T: Deciphering the genetic architecture of lung cancer survival: GWAS meta-analysis with 10K patients and integrative multi-omics study ⭐

Authors:

M. Du[1,2], J. Xin[2], L. Su[1], D. Christiani[1]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Nanjing Med. Univ., Nanjing, China

Abstract:

Current genome-wide association studies (GWASs) underlying case-control design have identified over 40 risk loci associated with lung cancer susceptibility, and its corresponding polygenic risk score (PRS) serves as a useful tool to identify high-risk individuals of developing lung cancer. However, the performance of PRS in predicting lung cancer prognosis remains limited, which highlight the importance to decipher the genetic architecture specific to lung cancer survival. Leveraging the genetic and clinical data of over 10,000 lung cancer patients from the Boston Lung Cancer Study, International Lung Cancer Consortium, The Cancer Genome Atlas, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial and UK Biobank cohorts of European ancestry, a large-scale survival GWAS meta-analysis was conducted to identify genetic variants associated with lung cancer overall survival, followed by functional annotation, fine-mapping, and pathway enrichment analysis. Furthermore, the transcriptome-wide association study (TWAS) approach was used to identify putative lung cancer prognosis-related genes, as well as characterized by single-cell and spatial transcriptomics. Besides, phenome-wide Mendelian randomization analysis (MR-PheWAS) was utilized to profile causal factors of lung cancer prognosis. We observed three independent common genetic loci associated with lung cancer overall survival, including rs77676649 (5q12.3; nearby *NLN*, hazard ratio [HR] = 1.26, *P* =

4.81×10^{-8}), rs12257527 (10p13; nearby *CAMK1D*; HR = 1.31, P = 1.28×10^{-8}) and rs79829102 (22q13.31; nearby *TBC1D22A*; HR = 1.91, P = 4.46×10^{-9}), for which, they were mainly enriched in pathways of selenocompound and lipoic acid metabolism, overlapped with enhancer histone modification, and functionally involved in lung tumor progression. Further, among six candidate lung cancer prognostic genes (P_{TWAS} < 0.001) provided by TWAS, *ANXA13* was prioritized in epithelial cells ant its higher expression was associated with poorer prognosis. In addition, MR-PheWAS analysis identified 15 potential causal factors related to lung cancer survival, of which the education level showed the strongest effect that patients with higher education was causally associated with a decreased risk of death ($P_{\text{inverse variance weighted}}$ = 2.06×10^{-4}). This large-scale GWAS study provided additional genetic biomarkers involving lung cancer progression, which would advance the understanding of genetic architecture of lung cancer survival and throw light on the individualized therapy.

Session Title: Cancer Poster Thursday Session

Board 8014T: Deep learning variant annotation methods increase pathogenic variant detection in women with breast cancer and nominate candidate noncoding variants.

Authors:

M. Akamandisa[1], B. Wubbenhorst[1], D. Gutman[1], S. Jewell[1], F. Zhuang[1], D. Wu[1], K. D'Andrea[1], S. Khanna[1], H. Symecko[1], R. Gillette[1], P. Shah[1], A. Bradbury[1], D. Huo[2], K. N. Maxwell[1], Y. Barash[1], S. M. Domchek[1], K. L. Nathanson[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Chicago, Chicago, IL

Abstract:

Breast cancer (BrCa), the most common cancer in women, has a 30% heritability estimate. Pathogenic variants (PV) in high and moderate penetrance susceptibility genes increase the likelihood of developing BrCa. PV carrier status informs treatment with platinums and PARP inhibitors, prevention of second BrCa in patients, and screening and prevention in family members who may also carry PVs. Women with early onset BrCa (< age 40), family history of BrCa, and multiple primary tumors have a high prior probability (PP) for PVs. Thus, women with such features undergo multi-gene panel testing. However, more than 60% of women with a high PP of genetic risk do not have PVs identified following testing. It is possible that variants within known BrCa susceptibility genes that contribute to disease are missed as they are in gene-specific noncoding regions, and thus not evaluated. We had previously sequenced and evaluated 524 BrCa patients without PVs in known BrCa susceptibility genes. To determine

whether gene-specific noncoding variants contribute to BrCa risk in this cohort, we analyzed these sequencing data using deep learning models for splicing aberrations (SpliceAI, Pangolin, TrASPr), G-quadruplex creating/disruption (G4mer), and upstream open reading frame (uORF) detection. We identified 14 splicing variants in 20 (3.8%) patients. Of these, 9 variants in 12 (2.3%) patients had not previously been identified. 5 out of 9 variants are known PVs whereas 3 intronic variants in *ATM* and *RAD51C*, and 1 exonic *TP53* variant have as yet unknown significance. We also found 6 (1%) 5'/3'-UTR variants in *ATM, CDKN2A, PTEN*, and *RAD51D* predicted to alter uORFs or to break G4s, both of which are predicted to regulate gene expression. All noncoding variants had frequency <0.5% in gnomAD.
The *RAD51C* c.145+246A>T intronic variant was found exclusively in women of African descent (AFR) and impacts splicing as verified by RT-PCR and amplicon-based sequencing of RNA from carriers. It was not associated with BrCa, irrespective of ER-status, in 32,040 AFR women in the AABCG. However, there was a small (7%) increase in variant frequency in women with ER-negative versus ER-positive BrCa suggesting that it could modify risk of ER-negative BrCa, prevalent in AFR women. Validation of impact on splicing, gene expression levels, and BrCa association studies are underway for the remaining variants. In summary, analysis of noncoding regions with deep learning methods increases identification of PVs within known genes and nominates candidate variants that could contribute to BrCa risk.

Session Title: Cancer Poster Thursday Session

Board 8015T: DeepSomatic: A SNV and Small-Indel Somatic Caller Using Deep Neural Networks

Authors:

K. Shafin; Google LLC, Mountain View, CA

Abstract:

DeepVariant is a highly accurate germline variant caller that applies deep dearning to classify germline variants with high accuracy. Here, we present DeepSomatic, which applies DeepVariant's convolutional neural network (CNN) to accurately call somatic mutations from paired tumor-normal sequencing. To develop DeepSomatic, we adapted the input pipeline to accommodate tumor-normal pairs. DeepVariant model inputs consist of a set of pileup images, also referred to as "channels", that represent features extracted from sequence data at candidate sites. Features include base, base quality, mapping quality, and haplotype. We modified this input for DeepSomatic by stacking the channels for tumor-normal pairs into a single input. In this way, our model can compare tumor and normal reads and learns to

distinguish germline variants from somatic mutations.
For training, DeepSomatic uses sequence data from the SEQC2 consortium, which has established a benchmarking dataset for HCC1395 (a triple-negative breast cancer cell line) across a variety of sequencing technologies. Additionally, we have supplemented the SEQC2 data with additional sequencing from cell lines (H2009, H1437, HCC1954, HCC1937, Hs578T) to further improve our training datasets and enable benchmarking across different cancer types. Using these data, we have trained models on whole-genome data from Illumina , PacBio, and Oxford Nanopore (ONT).
We report that DeepSomatic outperforms existing somatic callers across sequencing technologies. For example, we observe that our DeepSomatic Illumina model has a SNP F1 score of 0.983 on held-out chr1 HCC1395 data, which outperforms ClairS (F1=0.969), and Strelka2 (F1=0.952). Similarly, our PacBio model has a SNP F1=0.95 compared to ClairS (F1=0.935). Finally, our ONT model achieves an F1=0.869, which is an improvement over ClairS (F1=0.863). We similarly observe that DeepSomatic outperforms on F1 scores for indels as well. We plan to investigate tumor-only models and additional training approaches to further improve DeepSomatic.

Session Title: Cancer Poster Thursday Session

Board 8016T: Development and application of *MUTYH*-specific ACMG/AMP variant classification criteria

Authors:

E. Nadeau[1], C. Rioja[2], M. Pineda[2], M. Thet[3], A. Laner[4], D. Buchanan[5], J. Del Valle[2], E. Borras[6], F. Hernandez[7], C. Heinen[8], I. M. Frayling[9], K. Mahmood[10], D. Goldman[11], J. Montgomery[12], J. Ngeow Yuen Yin[13], M. Johnson[7], P. Georgeson[10], R. Urban[14], S. David[15], S. Farrington[16], T. Hansen[17], W. Shen[14], X. Shi[18], X. Yin[12], D. Ritter[19], A. Latchford[20], F. Macrae[12], S. Tavtigian[21], G. Capellá[2], M. Greenblatt[1]; [1]Univ. of Vermont, Burlington, VT, [2]Catalan Inst. of Oncology, Barcelona, Spain, [3]Univ. of Melbourne, Parkville, Australia, [4]MGZ, Munich, Germany, [5]The Univ. of Melbourne, Parkville, Australia, [6]Invitae, San Francisco, CA, [7]Ambry Genetics, Aliso Viejo, CA, [8]Univ Connecticut Hlth.Ctr., Farmington, CT, [9]Inst. of Med. Genetics, Cardiff, United Kingdom, [10]Dept. of Clinical Pathology, Univ. of Melbourne, Parkville, Australia, [11]Constantiam Bioscences, Houston, TX, [12]Royal Melbourne Hosp., Parkville, Australia, [13]Natl. Cancer Ctr., Outram, Singapore, [14]Mayo Clinic, Rochester, MN, [15]Univ. of California Davis, Davis, CA, [16]Univ. of Edinburgh, Edinburgh, United Kingdom, [17]Copenhagen Univ. Hosp., Roskidle, Denmark, [18]GGC, Greenwood, SC, [19]Baylor Coll. of Med., Houston, TX, [20]St Mark's Hosp., London, United Kingdom, [21]Huntsman Cancer Inst., Salt Lake City, UT

Abstract:

The ClinGen InSiGHT Hereditary Colon Cancer/Polyposis Variant Curation Expert Panel (VCEP) was formed in 2021 to curate variants in genes definitively associated with these conditions. The *MUTYH* subVCEP created gene-specific ACMG/AMP variant classification rules for *MUTYH*-associated polyposis (MAP), an autosomal recessive disorder caused by inherited biallelic pathogenic variants in *MUTYH*. MUTYH is a base excision repair glycosylase that excises adenines mispaired with 8-oxo-deoxyguanosine, which if left unrepaired lead to somatic G>T transversions. We developed ACMG criteria for *MUTYH* based on literature review of clinical, *in vitro*, and *in silico* data. Gene-specific classification criteria were developed on monthly calls of 3 expert subgroups: clinical, computational, and functional, supplemented by calls with the full subVCEP to define the specifications of all ACMG criteria for *MUTYH*. We used the upcoming points-based ACMG guidance in our rules. Currently, there are 2797 *MUTYH* variants in ClinVar. Of these, 1268 are missense, of which 88.6% (1124) are Variants of Uncertain Significance (VUS). Our gene-specific rules should resolve 7.3% of missense variants that currently have conflicting assertions, and upgrade 43.9% of missense variants with a single submitter to 3-star Expert Panel status. Several codes such as PM2, PM3, and PM4 were converted to points-based systems with strengths updated accordingly. We also established specific points-based phenotype rules for variant carriers to account for MAP's diverse phenotypes: variable colon polyp burden, age of onset, extracolonic features, and cancer association. We assigned the strongest level of evidence to somatic *KRAS* and *APC* G>T transversions which provide strong support for a MAP diagnosis, but test results are rarely available. *In vitro* MUTYH functional assays were thought to be promising data to classify variants. However, we could not apply criteria PS3/BS3 for any functional assay due to inadequate use of controls and lack of statistical analyses to establish evidence strength. Using allele frequencies from gnomAD v4.1.0, 7 variants will be classified as Benign based on ACMG code BA1, and 3 more receive "benign points" using code BS1. We will also use the newly updated *in silico* PP3/BP4 criteria to determine which tools are best for scoring *MUTYH* missense variants. In our pilot phase, we are testing our rules on select variants to validate that they will lower the number of VUS in ClinVar and ultimately, improve MAP management. The development of well calibrated MUTYH functional assays and improved clinical phenotype data will improve the classification and curation of *MUTYH* variants.

Session Title: Cancer Poster Thursday Session

Board 8018T: Development of Cross-ancestry Polygenic Risk Scores for Breast Cancer in Women with African Ancestry ⭐

Authors:

J. Li[1], H. Zhang[2], J. McClellan[1], G. Gao[3], K. Wang[4], J. Palmer[5], W. Zheng[6], O. Olopade[1], C. Haiman[7], D. Huo[1], AABCG Consortium; [1]Univ. of Chicago, Chicago, IL, [2]Natl. Cancer Inst., Cambridge, MA, [3]Univ. of Chicago, CHICAGO, IL, [4]Natl. Cancer Inst., Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, [5]Slone Epidemiology Ctr., Boston, MA, [6]Vanderbilt Univ Med Ctr., Nashville, TN, [7]Univ. of Southern California, Los Angeles, CA

Abstract:

Introduction: Polygenic risk scores (PRSs) predict breast cancer (BC) risk and have shown utility for screening in European ancestry women (EUR). However, PRSs developed for non-European ancestry, particularly African American (AA) women, have limited predictive accuracy.
Methods: We analyzed data from 17,454 BC cases and 19,093 controls in the African American Breast Cancer Genetics (AABCG) consortium, split into training (70%), tuning (10%), and testing (20%) sets. Genome-wide association analyses for overall BC, ER-positive BC, ER-negative BC and triple-negative BC (TNBC) were performed in the training dataset, controlling for age, genotyping platform, and the top 10 principal components. We built single-ancestry PRS models using forward stepwise selection (FSS), PRSice-2, PLINK C+T, Lassosum2, and LDpred2-auto. These were combined with published EUR 313-SNP models (330-SNP for TNBC) in the tuning set to obtain cross-ancestry models. Genome-wide cross-ancestry PRS methods, including CT-SLEB and PRScsx, were also used. We created an ensemble model by combining different PRS models across subtypes using ensemble-based approaches. All developed PRS models were then evaluated in the independent testing set. Additional validation was performed using the All of Us (AoU) study with 834 BC cases and 29,122 controls from the AA population.
Results: The observed-scale SNP-based heritability was estimated at 0.29 (SE: 0.12) for overall BC risk, 0.43 (SE: 0.18) for ER-positive BC, 0.44 (SE: 0.24) for ER-negative BC, and 1.13 (SE: 0.40) for TNBC. The best models showed good discriminating capacity: ensemble-based models for overall (AUC=0.626) and ER-positive BC (AUC=0.626), and combined cross-subtype and cross-ancestry FSS models for ER-negative (AUC=0.647) and TNBC (AUC=0.654). The odds ratio per standard deviation was 1.63 (95% CI: 1.55, 1.71) for the ensemble-based PRS of overall BC risk, with an estimated life risk of 28% for the top 1%. The cross-ancestry model using CT-SLEB combined with the EUR 313-SNP model was validated (AUC=0.61) among AA women in the AoU study.
Conclusion: We developed PRSs with improved discriminating ability for BC risk in women with African ancestry. Cross-ancestry and subtype learning enhanced model performance. These

findings indicate common variation has a larger role in BC in AA women than previously reported, warranting further validation in external AA populations and clinical evaluation.

Session Title: Cancer Poster Thursday Session

Board 8019T: Effect of Inactivation of MAPT Through Gene Knockdown on Tumor Invasion and Microtubule Stability in Prostate Cancer Cells

Authors:

L. Chen[1,2]; [1]Academy of Sci., Leesburg, VA, [2]J&C Hlth.care, Fairfax, VA

Abstract:

Prostate cancer is a common disease that affects about 13 out of every 100 American men and androgen independent prostate cancer, which makes up around one third of prostate cancer has a life expectancy of around 12 to 18 months (**1**). Tumor invasion, the amount in which a tumor can invade into surrounding tissue, causes spreading in the cancer cells, leading to more damage and difficult treatment due to metastasis, or the development of new tumors around the original tumor. Finding the tumor invasion is important to assess the amount of damage a tumor can cause; thus, it is important to test its significance in prostate cancer tumors in order to properly determine if it is viable to treat prostate cancer using shRNA to knockdown tau expression. In prior studies, it has been found that knockdown of tau led to increased effectiveness of chemotherapies such as docetaxel (**2**). The current data has supported, using an acid phosphatase assay, that there is a significant decrease in cell viability in the prostate cancer cells following the transfection of the shRNA supporting prior research (**2**). For tumor invasion, it was found using a transwell invasion assay, that there isn't a statistically significant difference between the transfected and non-transfected cells, though this could be due to lack of trials. Further research on the tumor invasion will be necessary to properly determine the relationship between tumor invasion and tau expression in prostate cancer cells.
References: 1. Current oncology (Toronto, Ont.) (2010) vol. 17,2: 24-9. **2.** Research and reports in urology, (2017) 9, 71-77.

Session Title: Cancer Poster Thursday Session

Board 8020T: Functional characterization of the 1p36.33 pancreatic cancer GWAS locus ⭐

Authors:

K. Connelly[1], K. Hullin[1], E. Abdolalizadeh[1], J. Zhong[1], D. Eiser[1], J. Hoskins[1], A. O'Brien[1], A. Jermusyk[1], I. Collins[1], Pancreatic Cancer Cohort Consortium, Pancreatic Cancer Case Control Consortium, S. J. Chanock[1], R. Stolzenberg-Solomon[1], A. Klein[2], B. Wolpin[3], J. Shi[4], J. Smith[5], S. Antwi[6], L. Amundadottir[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Johns Hopkins Med Sch, Baltimore, MD, [3]Dana-Farber Cancer Inst., Boston, MA, [4]Natl. Cancer Insititute, Rockville, MD, [5]Georgetown Univ., Washington, DC, [6]Mayo Clinic, Jacksonville, FL

Abstract:

Pancreatic ductal adenocarcinoma (PDAC), the third leading cause of cancer-related deaths in the US, is often difficult to diagnose. A 2018 meta-analysis of PanScan I-III and PanC4 genome-wide association studies (GWAS) identified 1p36.33 (rs13303010, P = $7.3x10^{-7}$; OR= 1.20; 95% CI = 1.12-1.29) as a suggestive PDAC risk locus. We performed a new meta-analysis with an additional 1066 PDAC cases and 9399 controls from the UK Biobank and found that the 1p36.33 signal surpasses the GWAS significance threshold (OR=1.24, 95% CI = 1.16-1.32, P=$2.09x10^{-10}$). Using fine mapping, we identified seven candidate functional variants for functional studies. Two of the seven (rs13303327 and rs13303160) demonstrated allele-specific binding and regulatory activity in *in vitro* binding and luciferase assays, respectively. *In-silico* transcription factor (TF) binding analysis predicted the ELF and AP1 TF motifs to be disrupted by the risk alleles of rs13303327 and rs13303160, respectively. We confirmed that ELF2 preferentially binds the rs13303327 protective allele *in vitro* but did not observe an allelic preference in cell lines using chromatin immunoprecipitation (ChIP)-qPCR. For rs13303160, JunB and JunD preferentially bound the protective allele *in vitro* and *in vivo*, implicating this polymorphism as a functional variant underlying risk at 1p36.33. The GWAS signal strongly colocalized (PP=0.99) with a significant *KLHL17* expression quantatitve trait locus (eQTL) signal in the pancreas (P=$2.3x10-^{10}$) where the risk allele is associated with lower *KLHL17* expression. Additionally, the eQTL effect direction corresponded with our *in vitro* regulatory results. The KLHL family of proteins are substrate recognition proteins for the Cullin3 E3 ubiquitin ligase complex and are implicated in cancer. We confirmed that KLHL17 is associated with Cullin3 complex members in the pancreas. In PDAC-derived cell lines, knockdown/overexpression of KLHL17 did not exhibit a growth phenotype. Thus, we performed an *in silico* differential gene expression analysis between the top and bottom quartiles of *KLHL17* expression in GTEx Pancreas and TCGA PAAD samples. Differentially expressed genes in both datasets were enriched for inflammation-related pathways, suggesting lower *KLHL17* is associated with a pro-inflammatory environment. We hypothesize that KLHL17 may mitigate inflammation by recruiting pro-inflammatory proteins for ubiquitination and degradation. As such, lower expression of KLHL17, which associates with PDAC risk, would result in increased un-degraded proteins and unresolved inflammation, a precursor in PDAC pathogenesis.

Session Title: Cancer Poster Thursday Session

Board 8021T: Functional genetic variant rs59133000 of gastric cancer regulating *PTGER4* in response to *Helicobacter pylori* infection and affecting NK cell functions ⭐

Authors:

Y. Chen[1,2], X. Wang[1], H. Ma[1], G. Jin[1], C. Yan[1]; [1]Nanjing Med. Univ., Nanjing, China, [2]Southeast Univ., Nanjing, China

Abstract:

Background: Over the last decade, genome-wide association studies (GWASs) have identified nearly 20 gastric cancer risk loci. Among them, the susceptibility genetic variation at chromosome 5p13.1 has been found to interact with Helicobacter pylori (*H. pylori*). However, the specific interaction mechanism and its impact on the role of susceptibility genes have not yet been elucidated. **Methods:** Based on the transcriptome and genotype data of 645 normal stomach tissue, we conducted expression quantitative trait locus (eQTL) analysis and co-localization analysis with gastric cancer GWAS to identify susceptibility genes at 5p13.1. After functional annotation, we performed CUT&RUN of histone H3K27ac and H3K4me1, as well as luciferase reporter gene assay, chromatin conformation capture (3C) and CRISPRa to validate the function of rs59133000 in regulating the expression of susceptibility gene *PTGER4* by influencing the enhancer activity, with or without *H. pylori* infection. We then conducted a series of in vitro and in vivo experiments and single cell RNA sequencing (scRNA-seq) to explore the role of *PTGER4* in gastric cancer development and emphasized on NK cells. Finally, we establish a co-culture system between gastric cancer cells and NK cells to explore the mechanism of *PTGER4* in affecting NK cell function. **Results:** We found that the genetic variation at 5p13.1 significantly co-localized with the susceptibility gene *PTGER4*, and confirmed that the enhancer region around the functional variant rs59133000 was spatial interacted with the promoter of *PTGER4*. The risk allele C of rs59133000 could increases binding of NF-κB transcription factor activated by *H. pylori* and results in up-regulation of *PTGER4* expression. Although *PTGER4* does not affect tumor phenotype in vitro, but in vivo experiments and scRNA-seq analysis confirmed that knockdown of *PTGER4* in cancer cells could significantly inhibited the proliferation of cancer cells, along with significantly decreased proportion of NK cells. We further found that *PTGER4* increased the expression of NK cell inhibitory ligand HLA-E, which activated NK-inhibitory receptor NKG2A and finally repressed NK cell functions. **Conclusions:** The risk allele C of rs59133000 at 5p13.1 could elevate the expression

of *PTGER4* by increasing binding of p50/p65 under *H. pylori* infection. *PTGER4* increasing the HLA-E expression in gastric cancer cells and inhibiting NK cells functions through NKG2A receptor and ultimately promoting the development of gastric cancer.

Session Title: Cancer Poster Thursday Session

Board 8023T: Genetic predisposition to pancreatic adenocarcinoma in Czech patients - *BRCA2, ATM* and *BRCA1* are the most frequent mutant genes

Authors:

M. Koudová, S. Chvojka, L. Cerna, M. Bittoova, M. Urbanova, M. Hodúlová, F. Zembol, B. Honysová, S. Šuhajová, H. Dvořáčková, M. Kovář, A. Puchmajerova, M. Pokorná, L. Horáková, P. Klempt, D. Stejskal; GENNET, Prague, Czech Republic

Abstract:

Introduction: Pancreatic adenocarcinoma (PAC) is the devastating malignancy with poor prognosis, late diagnosis and with increasing incidence. We present results of hereditary cancer (HC) predisposition testing of Czech patients diagnosed with PAC and their families. Methods: Between 2016 and 2024 we conducted genetic testing of 11,186 individuals in Czechia fulfilling clinical criteria for HC, 125 patients with PAC and 1084 healthy relatives of PAC cases among them. Presented results were evaluated from standard short-read NGS data (NextSeq 550, NovaSeq Xplus, Illumina) obtained from two panels: a) the community-developed, targeted hereditary cancer panel CZECANCA (Roche, PMID: 29649263) and from the customized, exome (WES) panel (Twist). The NGS data from both panels were alongside to Franklin (Genoox) platform analyzed using our proprietary bioinformatic pipeline using local installation of Ensembl genomic database for annotation and own variant database CheckBase for data handling and clinical reporting. Clinical interpretation was prioritized according to the primary indication, review based on IACR* standards, ACMG guidelines (PMID: 25741868), and HGVS genetic nomenclature. The exon deletions/duplications and major gene rearrangements were detected utilizing our coverage analysis alongside the Rainbow pipeline within the Franklin (Genoox). Positive results were verified by Sanger sequencing, while the CNV findings were confirmed by the MLPA (MRC Holland). Results: Clinically relevant variants (class 4 and 5) were detected in 23% of PAC patients and 17% of their healthy relatives. BRCA2 (20,7%), ATM (17,2%), MUTYH (10,3%) and BRCA1, FANCA, FH, RECQL (6,9%), were the most frequently mutated genes in patients, while mutations in BRCA2 (15,5%), BRCA1 (12,8%), ATM (9,1%) CHEK2 (6,9), FANCI (5,8%) and MUTYH, PALB2 (5,3%) were most common in relatives. The most

frequent pathogenic variant in the BRCA1 gene was c.1687C>T(p.Gln563Ter) and in the BRCA2 gene c.3847_3848del (p.Val1283LysfsTer2). Conclusion: These findings underline the importance of widespread genetic testing for HC predisposition in patients with PAC and their families. The results are important for setting up clinical follow-up to prevent PAC and other cancers in the family and for the possibility of targeted biological treatment in patients with gene mutations. (e.g. PARPi in BRCA positiv).

Session Title: Cancer Poster Thursday Session

Board 8024T: Genetic testing of healthy family members of Hereditary Breast and Ovarian Cancer patients in routine clinical practice

Authors:

T. Kahre[1,2], M. Tooming[1,2], K. Rekker[1], K. Toome[1], O. Fjodorova[1], L. Roht[1,3], P. Laidre[1], U. Šamarina[1], H. Roomere[1], Ü. Murumets[1], K. Ounap[1,2]; [1]Genetics and Personalized Med. Clinic, Tartu Univ. Hosp., Tartu, Estonia, [2]Genetics and Personalized Med. Clinic, Inst. of Clinical Med., Univ. of Tartu, Tartu, Estonia, [3]North Estonia Med. Ctr., Tallinn, Estonia

Abstract:

Genetic testing of breast and ovarian cancer patients, specifically BRCA1/2 and other related genes, has become standard practice in clinical diagnostics. Extending genetic testing to healthy family members (HFM) of individuals with Hereditary Breast and Ovarian Cancer (HBOC) enables the implementation of preventive measures and timely referral for screening. The aim of this study was to characterize the cohort of HFM referred from four major tertiary hospitals in Estonia to genetic testing at Tartu University Hospital, Genetics and Personalized Medicine (GPM) Clinic, from 2007-2023, and to evaluate the diagnostic efficacy of genetic testing. HFM were tested using Sanger sequencing if a known pathogenic variant had previously been identified in a family. Alternatively, Illumina TruSight Cancer/Hereditary Cancer gene panel testing was utilized in cases where no family member had undergone testing or was unavailable for testing.RESULTS: A total of 3722 HFM, comprising 3195 females and 527 males, were tested. Significant increase in testing of HFM was observed during the study period from 3 HFM in 2007 to 741 individuals in 2023. The average age of individuals referred to testing was 41 years (range 1-89), with an average of two relatives in a family affected by cancer. Specifically, 1207 individuals had one relative with cancer, 1158 had two, 688 had three, and 580 had four or more cancer-affected relatives. Altogether, 718 likely pathogenic or pathogenic variants [(L)PV] associated with elevated cancer risk were identified in 702/3722 individuals (18.8%). The majority of (L)PVs were identified in BRCA1/2 genes (404/718; 56.3%). Furthermore, remaining

314/718 (L)PVs (43.7%) were identified in non-BRCA genes such as CHEK2, FANCM, ATM, BRIP1, PALB2, RAD51C/D, BARD1, BLM, FANCC, PMS2, MSH2, MLH1, MSH6, NBN and TP53.In conclusion, the diagnostic efficacy of testing HBOC-related HFM was 18.8%. These individuals can effectively benefit from referral to preventive screening programs following genetic counselling. Furthermore, if affected family member is not available for genetic testing, the utilization of a comprehensive NGS gene panel in HBOC-related HFM is recommended, as it reveals ~44% more genetic alterations compared to testing BRCA1/2 genes solely.Funding: Estonian Research Council grants PRG471, PRG2040.

Session Title: Cancer Poster Thursday Session

Board 8025T: Genetically adjusting prostate-specific antigen strengthens associations with risk of upgrading and prostate tumor features in prostate cancer patients on active surveillance

Authors:

L. Goss[1], M. Liu[1], Y. Zheng[1], C. de la Calle[2], C. C. Pritchard[3], L. Kachuri[4], W. J. Catalona[5], J. S. Witte[6], D. W. Lin[2], L. F. Newcomb[1], B. F. Darst[1]; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Dept. of Urology, Univ. of Washington, Seattle, WA, [3]Dept. of Lab. Med. and Pathology, Univ of Washington, Seattle, WA, [4]Stanford Univ., San Francisco, CA, [5]Dept. of Urology, Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [6]Stanford Univ., Stanford, CA

Abstract:

BACKGROUND: Although prostate-specific antigen (PSA) screening reduces prostate cancer (PCa)-specific mortality, many PSA-detected cancers are indolent and unlikely to result in ill health effects. Accounting for genetic variation in PSA has been shown to improve prediction of aggressive PCa. However, it is unknown whether genetic adjustment of PSA could improve the prediction of upgrading in PCa patients on active surveillance, the strategy recommended for the >50% of PCa patients diagnosed with low- or favorable intermediate-risk PCa. In a cohort of PCa patients on active surveillance, we investigated the implications of genetically adjusting PSA and PSA density (PSAD) on prostate tumor features, particularly the risk of upgrading.
METHODS: This study included 1220 PCa patients on AS in the Canary Prostate Active Surveillance Study (PASS). A previously developed 447 variant PSA polygenic score (PGS_{PSA}) was applied in PASS. Genetically adjusted PSA (PSA_g) and PSAD ($PSAD_g$) were calculated for each participant by normalizing values relative to the average PGS_{PSA} in non-upgraders. Regression models were used to evaluate associations between PSA, PSAD, PSA_g, and $PSAD_g$, and prostate tumor outcomes, including upgrading at follow-up and percent of biopsy cores with cancer,

American Urological Association (AUA) symptom score, and quality of life urinary score at diagnosis. Models were adjusted for age and principal components of ancestry.
RESULTS: Of the 1220 patients, 470 had upgrading during the median 5.3 years of follow-up. Each unit increase in log-transformed PSA was associated with a 1.2-fold increased probability of upgrading (95% CI=1.01-1.38, P=0.03), while PSA_g was associated with a 1.4-fold increased probability of upgrading (95% CI=1.19-1.62, $P=3.2x10^{-5}$). Comparatively, each unit increase in PSAD was associated with a 7.7-fold increased probability of upgrading (95% CI=3.44-17.08, $P=6.4x10^{-7}$), while $PSAD_g$ was associated with a 12.3-fold increased probability of upgrading (95% CI=5.85-25.71, $P=3.3x10^{-11}$). PSA_g, PSAD, and $PSAD_g$ were also associated with a higher percentage of cores with cancer, while PSAD and $PSAD_g$ were associated with decreased AUA symptom score and quality of life urinary score. Consistent with these findings, each SD unit increase in PGS_{PSA} was associated with a 20% decreased probability of upgrading (95% CI=0.73-0.88, $P=2.3x10^{-6}$).
CONCLUSION: In PCa patients on active surveillance, genetically adjusting PSA and PSAD strengthened associations with upgrading compared to measured values by increasing the proportion of PSA variation that is reflective of factors related to PCa progression.

Session Title: Cancer Poster Thursday Session

Board 8026T: Genetics of Familial Uveal Melanomas

Authors:

M. Abdel-Rahman[1], P. Johansson[2], M. Hussein[1], L. Gonzalez[1], E. Schreiner[1], L. Byrne[1], F. Davidorf[1], N. Hayward[2], C. Cebulla[1]; [1]The Ohio State Univ., Columbus, OH, [2]QIMR Berghofer Med. Res. Inst., Herston, Australia

Abstract:

Purpose: Familial Uveal Melanomas (FUM), defined as two or more uveal melanomas (UM) in the same bloodline, are rare representing less than 1% of all UM patients. BAP1 mutations are detected in ~20% of FUM so other genes likely contribute to etiology. This study was performed to identify additional candidate Genes in FUM. **Methods**: A cohort of 54 FUM presented or referred to Department of Ophthalmology, The Ohio State University were first tested for *BAP1* mutations and those negatives were then tested using a multigene panel or whole exome sequencing (WES). A total of 64 subjects were tested including 54 probands and 10 family members. **Results**: Genetic testing of our cohort detected BAP1 germline pathogenic/likely pathogenic (P/LP) variants in 12/54 (22%). In addition germline P/LP in *MLH1*, *POLI*, *PALB2, MBD4, SMARCE1, RAD51B, XPC, ATM, RECQL4, TP53AIP1, DLEC1 &*

MMS19 were identified in one patient each. Interestingly, *MBD4* didn't segregate with the disease in another family member with UM. Excluding *MBD4*, we identified genetic alterations in known cancer genes in 23/54 (42.6%) of patients enrolled. **Conclusion**: Germline pathogenic variants in *BAP1* is the most common genetic alterations in FUM. Other genes in particularly those in DNA damage repair pathway could contribute to the tumorigenesis of this rare cancer

Session Title: Cancer Poster Thursday Session

Board 8027T: Genomic and Clinical Insights from a Large-Scale Study of 172 Retinoblastomas from Saudi Arabia

Authors:

M. AlThgafi (Abedalthagafi)[1], A. Maktabi[2], S. Almesfer[2], Y. Liu[3], D. Edward[2], R. Hoehndorf[3]; [1]Emory Univ., Atlanta, GA, [2]King Khalid Specialist Eye Hosp., Riyadh, Saudi Arabia, [3]King Abdullah Univ. of Sci. and Technology, Thuwal, Saudi Arabia

Abstract:

Retinoblastoma, a rare childhood cancer primarily driven by RB1 mutation or MYCN amplification, remains a significant challenge in pediatric oncology. In this study, we performed exome sequencing on 172 retinoblastomas along with matched blood DNA, utilizing clinical data from 171 patients. Of these cases, 119 had unilateral retinoblastoma, 49 had bilateral, and only three had trilateral occurrence. Our comprehensive analysis included single nucleotide variants and copy number alterations, revealing RB1 pathogenic variants in 76 patients and MYCN duplication in 79 cases, highlighting their critical roles in disease pathogenesis. Additionally, mutations in RPTOR and TSC2, part of the mTOR signaling pathway known to impact various cancers, were identified in two retinoblastoma patients. Notably, our copy number alteration analysis uncovered recurrent arm-level CNAs, including gains in 1q (23.3%), 2p (20.9%), and particularly 6p (41.3%), as well as other alterations such as gains in 5p (8.1%), 17q (8.1%), and loss in 16q (5.8%). Our study uniquely identified associations between genetic alterations and clinical features in retinoblastoma. Specifically, gains on chromosome 6p were linked to Rubeosis, Subretinal seeds, and Tumor LC beyond, while gains on 17q were associated with Rubeosis. Furthermore, gene collapsing identified RB1, MATR3, and CYP4F2 as potential retinoblastoma-related genes, while module identification analysis revealed retinoblastoma-associated modules such as HLA-DRB1-CBL, LPAR2, RB1, and TRIM28.Overall, this study presents one of the largest genomic cohorts of retinoblastoma to date, demonstrating the potential of next-generation sequencing (NGS) in uncovering convergent evolutionary mechanisms underlying gene mutations and copy number changes. Our findings underscore the promise of

individualized testing and analysis for validating emerging candidates, paving the way for improved therapeutic strategies and outcomes for retinoblastoma patients.

Session Title: Cancer Poster Thursday Session

Board 8028T: Global multi-ancestry meta-analysis uncovers pleiotropic genetic associations with thyroid cancer and other thyroid diseases

Authors:

S. White[1], B. O. Åsvold[2], E. Bertucci-Richter[3], M. Brasher[1], B. Brumpton[2], S. Chapman[3], S. Chavan[4], J. Chen[5], J. Cole[1], N. Cox[6], K. Crooks[4,7], A. EdrisMohammed[8], B. Fennessy[9], L. Fishbein[10], M. Fisher[4], L. Guare[11], J. Hirbo[6], T. Karaderi[12], E. Kenny[9], M. Lin[1], M. Moksnes[2], S. Morris[8], S. Namba[13], Y. Okada[13], R. Peculis[14], M. Preuss[9], N. Rafaels[4], V. Rovite[14], J. Shortt[1,4], P. Straub[6], M. Tobin[5], A. Verma[11], R. Walters[8], Y. Wang[15], W. Zhou[15], A. Williams[5], S. Zoellner[3], Colorado Center for Personalized Medicine, Global Biobank Meta-analysis Initiative, B. Haugen[10], C. Gignoux[1,4], N. Pozdeyev[1,4,10]; [1]Dept. of BioMed. Informatics, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]HUNT Ctr. for Molecular and Clinical Epidemiology, Dept. of Publ. Hlth.and Nursing, NTNU, Norwegian Univ. of Sci. and Technology, Levanger, Norway, [3]Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI, [4]Colorado Ctr. for Personalized Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Genetic Epidemiology Group, Dept. of Population Hlth.Sci., George Davies Ctr., Leicester, United Kingdom, [6]Dept. of Med., Div. of Genetic Med. & Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]Dept. of Pathology, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [8]China Kadoorie Biobank Collaborative Group, Oxford, United Kingdom, [9]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]Div. of Endocrinology, Diabetes and Metabolism, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [11]Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [12]Ctr. for Hlth.Data Sci., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, [13]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [14]Latvian BioMed. Res. and Study Ctr., Riga, Latvia, [15]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA

Abstract:

Thyroid cancer is the most common endocrine malignancy. Despite clinical and radiographic risk assessments, more than 90% of thyroid nodule biopsies produce benign, indeterminate, or inadequate results, frequently requiring additional testing and diagnostic surgery. We hypothesize that understanding the genetic risk of developing thyroid cancer and its interaction

with other thyroid diseases will help in diagnosing thyroid malignancies while preventing unnecessary procedures for benign thyroid nodules. In this study, we performed the world's largest multi-ancestry meta-analysis of genome-wide association studies for thyroid cancer, benign nodular goiter, and other thyroid diseases from 12 biobanks participating in the Global Biobank Meta-analysis Initiative. The meta-analysis aggregated data from 9,039 thyroid cancer patients, 36,527 individuals with benign thyroid nodules, 11,647 with Graves' disease, 105,825 with hypothyroidism, and 1,062,533 controls. An increased statistical power in multi-population analyses led to the confirmation of most established associations and the discovery of many novel genetic associations for four thyroid diseases. We identified (1) variants associated with thyroid cancer but not benign nodules (malignant variants, *PCNX2, DIRC3, EPB41L4A, MSRB3-AS1, and ZNF257-ZNF676*), (2) variants that are associated with both thyroid cancer and benign nodules (neoplastic variants, *PTCSC2, TERT, NRG1, INSR*, among others) and (3) benign variants (*FAM227B, MICOS10, TPO*, among many others). LD score regression in European-descent individuals showed a statistically significant genetic correlation between thyroid cancer and benign thyroid nodules (r=0.60, SE=0.21, p=0.004). In addition, we found a significant genetic correlation between thyroid cancer and Graves' disease (r=0.27, SE=0.07, p=0.0004), a mechanistically plausible observation due to the known pro-growth effect of thyroid-stimulating hormone receptor activation. Our findings indicate a significant shared genetic architecture between a subset of benign thyroid nodules and thyroid cancers, suggesting that some thyroid nodules diagnosed as benign may be premalignant lesions. This challenges the current clinical paradigm that benign and malignant nodules are distinct diseases with separate causal pathways.

Session Title: Cancer Poster Thursday Session

Board 8029T: Haplotype-resolved assemblies for somatic mutation benchmark development for a GIAB tumor-normal pair reference sample

Authors:

C. Xiao[1], J. McDaniel[2], J. Wagner[2], N. Olson[2], V. Patel[2], F. Thibaud-Nissen[1], A. Liss[3], J. Zook[2]; [1]NIH, Bethesda, MD, [2]NIST, Gaithersburg, MD, [3]Dept. of Surgery, Massachusetts Gen. Hosp., Boston, MA

Abstract:

NIST-led Genome-in-a-Bottle (GIAB) consortium has generated and released a wide variety of sequencing data, including multiple short- and long-read technologies, HiC, and optical mapping

for a new broadly-consented tumor-normal pair (HG008) reference sample derived from the same individual, aiming for development of high-quality somatic mutation benchmarks to advance the development of sequencing technologies and analytical methods. Here we report that we used the above data to generate chromosomal scale haplotype-resolved assemblies for both normal tissue and tumor cell lines using the latest Verkko and hifiasm assemblers, and used various evaluation tools including QUAST, Merqury and BUSCO to assess assembly quality. Our analysis showed that both the tumor and normal genomes were assembled with high quality (QV 64+). Most chromosomes were assembled into one single scaffold for the normal cell line even without ultralong nanopore sequences at this moment. The total bases of the Haplotype-1 and Haplotype-2 of the normal assembly are 2.99Gb and 2.98Gb, with N50s 132.8 and 125.3 Mb, respectively. Interestingly, when compared to T2T-CHM13v2.0, the tumor assemblies from both Verkko and hifiasm clearly revealed chromosomal- or arm-scale large deletions involving several chromosomes, including chr1p, chr3p, chr4, chr5q, chr6 (excluding centromere region), chr7q, chr8p, chr9, chr12q, chr16p, chr17p, chr18q, chr19p, chr20p, chr21q, and chr22q, with approximately 12% missing from the Haplotype-1 and 35% missing from the Haplotype-2. These large deletion events were supported by the data from other orthogonal technologies such as BioNano optical mapping and karyotyping. When the haplotype-resolved assemblies of the normal sample were used as a reference genome for short- and long-read mapping and subsequent somatic mutation discovery, we found read mapping error rates were reduced as compared to when GRCh38 and T2T-CHM13v2.0 were used as references, while the mapping rates were similar. We found that the vast majority of somatic SNVs was consistent across different reference assemblies, but we also identified a fraction of somatic SNVs being reference (GRCh38 or T2T-CHM13v2.0) specific, indicating potential false positives. While we continue to improve the qualities of the assemblies of this tumor-normal pair with ongoing data collection, we can already affirm that such de novo assembly-based analyses will provide more insights into the ongoing refinement of high-quality somatic mutation benchmarks for this tumor-normal pair.

Session Title: Cancer Poster Thursday Session

Board 8030T: Head and Neck Paraganglioma Study: Whole Genome & Transcriptome Analysis

Authors:

J. Geng, F. Bu, Y. Lu, H. Yuan; West China Hosp., Chengdu, China

Abstract:

Objective: This study comprehensively analyzes clinical characteristics, germline genetic variations, somatic mutations, and transcriptome expressions in Chinese head and neck paraganglioma (HNPGL) patients to uncover pathogenesis and aid clinical diagnosis and treatment. Using whole-genome and transcriptome sequencing, we systematically examine germline variations, somatic mutations, and gene expression patterns, aiming to provide a theoretical basis for precise treatment.Materials and Methods: Clinical data from 36 HNPGL patients were collected, including carotid body and jugular bulb tumors. Whole-genome sequencing identified germline variations, while tumor tissue samples underwent sequencing to analyze somatic mutations and gene expression. Bioinformatics and statistical analysis were applied to interpret the data.Results:Germline genetic variation analysis identified 14 pathogenic variations, with a 44.44% diagnostic rate. SDHB gene variations were predominant.Clinical characteristics linked to germline diagnosis showed SDHB variations associate with postoperative recurrence and multiple tumors, while SDHD variations are linked to bilateral onset and genetic imprinting.Somatic mutation analysis revealed 73 variations in CGC-listed genes, with ATRX mutations correlating with earlier diagnosis and faster tumor growth.T ranscriptome analysis indicated HNPGL development is associated with overexpression of genes like KLK4 and ARHGAP36, with differentially expressed genes involved in synapses, extracellular matrix, and actin proteins, and in signaling pathways PI3K-Akt, MAPK, and Rap1.Conclusion: The study reveals the pathogenesis of HNPGL and identifies potential therapeutic targets. SDHB and SDHD gene variations play a key role in HNPGL pathogenesis, and chronic hypoxic environments may contribute to its development. Somatic ATRX mutations and overexpression of tumor-related genes are closely related to tumor characteristics. The findings provide a basis for precise and molecular treatments of HNPGL, with future exploration aimed at uncovering the biological mechanisms behind these associations for more effective clinical guidance.

Session Title: Cancer Poster Thursday Session

Board 8031T: HiChIP-based epigenomic footprinting identifies a promoter variant of *UXS*1 that confers genetic susceptibility to gastroesophageal cancer.

Authors:

A. Gnanapragasam[1], E. Kirbizakis[1], A. Li[1], K. White[1], K. Mortenson[2], J. Cavalcante de Moura[1], W. Jawhar[1], Y. Yan[1], R. Falter[2], C. Russett[1], B. Giannias[1], S. Camilleri-Broët[1], N. Bertos[1], J. Cools-Lartigue[1], L. Garzia[1], V. Sangwan[1], L. Ferri[1], X. Zhang[2], S. Bailey[1]; [1]McGill Univ., Montreal, QC, Canada, [2]Huntsman Cancer Inst., Salt Lake City, UT

Abstract:

Genome-wide association studies (GWAS) have identified more than a hundred single nucleotide variants (SNVs) associated with the risk of gastroesophageal cancer (GEC). The majority of the identified SNVs map to noncoding regions of the genome. Uncovering the causal SNVs and the genes they modulate could help improve GEC prevention and treatment. Here, we used HiChIP against histone 3 lysine 27 acetylation (H3K27ac) to simultaneously annotate active promoters and enhancers, identify the interactions between them, and detect nucleosome free regions (NFRs) harboring potential causal SNVs in a single assay. Application of H3K27Ac HiChIP in GEC relevant models identified 61 potential functional SNVs that reside in NFRs and interact with 49 genes at 17 loci. The approach led to a 67% reduction in the number of SNVs in linkage disequilibrium at these 17 loci, and at seven loci a single putative causal SNV was identified. One SNV, rs147518036, located within the promoter of the UDP-glucuronate decarboxylase 1 (*UXS1*) gene appeared to underlie the GEC risk association captured by the rs75460256 index SNV. The rs147518036 SNV creates a GABPA DNA recognition motif, resulting in increased promoter activity, and CRISPR-mediated inhibition of the *UXS1* promoter reduced viability of GEC cells. These findings provide a framework that simplifies the identification of potentially functional regulatory SNVs and target genes underlying risk-associated loci. In addition, the study implicates increased expression of the enzyme UXS1 and activation of its metabolic pathway as a predisposition to gastric cancer, which highlights potential therapeutic avenues to treat this disease.

Session Title: Cancer Poster Thursday Session

Board 8032T: High Depth Allele-Specific Expression Across Healthy Tissues and Ovarian Cancer

Authors:

N. Vetr[1], N. S. Abell[1], E. Tsang[1], N. Cheng[2], S. Linder[1], K. S. Smith[1], J. Uzunović[3], F. Lamaze[4], K. Ardlie[5], P. Awadalla[4], J. B. Li[1], S. B. Montgomery[1]; [1]Stanford Univ., Stanford, CA, [2]Natl. Univ. of Singapore, Singapore, Singapore, [3]Univ. of Toronto, Toronto, ON, Canada, [4]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [5]Broad Inst., Cambridge, MA

Abstract:

Allele-specific expression (ASE) from RNA-sequencing (RNA-Seq) is a powerful modality for quantifying haplotype-specific gene expression across healthy and diseased human tissues. However, read length and sequencing depth limit ASE's power as a biomarker to detect germline- and somatic-driven regulatory events. To overcome this limitation, we applied

multiplexed microfluidic PCR and sequencing (mmPCR-Seq) to selectively amplify a set of target loci in a carefully curated list of 372 cancer-related genes, resulting in high-depth coverage of 1,056 select heterozygous sites (median Nx). Abnormal deviations from allelic balance (AB), including Loss of Heterozygosity (LoH), are key indicators of cancer development. Beyond leveraging mmPCR-Seq to estimate ASE across 52 healthy GTEx tissues in 79 individuals in comparison to RNA-Seq, we further contrasted a sample of 96 ovarian tumors against healthy controls to identify cancer-related genes most prone to allelic imbalance (ASE and LoH). For this analysis, we developed a novel beta-binomial mixture model with components corresponding to AB, ASE, and LoH, incorporating multilevel effects within and between components corresponding to gene, individual, and cancer stage. In the pan-tissue results, we identify numerous multi-tissue genes with elevated rates of ASE and LoH, including 38 that were consistently among the top 10% of elevated genes in 20 or more tissues and 88 genes at this tail proportion in 5 or fewer tissues. Comparing ovarian tumors to healthy controls, meanwhile, yielded 10 genes with over 95% posterior mass of elevated probability of ASE or LoH in Grade I cancer samples, as well as confident monotonic increasing trends in average probability of ASE and residual variance in ASE with tumor grade. To identify aberrant ASE in healthy and cancer samples and facilitate broader application, we also developed flexible software to support this and other Bayesian workflows. Combined, this work advances the Enhancing GTEx project to present a high-resolution survey of ASE in key cancer-associated and genetically regulated genes across human tissues. It also identifies numerous ovarian cancer-affiliated ASE genes, informing the impact of inherited and somatic mutations on gene expression across ovarian cancer stages.

Session Title: Cancer Poster Thursday Session

Board 8033T: Highly accurate classification of prostate cancer using a random forest classifier built on 3' cfDNA fragment ends

Authors:

T. Groth[1], C. Schwartz[1], A. Mishin[1], C. Troll[1], V. Rao[1], J. Siddiqui[2], A. Chinnayyan[2], R. Green[3]; [1]Claret BioSci., Scotts Valley, CA, [2]Univ. of Michigan, Dept of Pathology, Ann Arbor, MI, [3]Univ. of California Santa Cruz, Santa Cruz, CA

Abstract:

Prostate cancer (PCa) is one of the most prevalent cancers among men worldwide, and there is a growing demand for a noninvasive approach to early diagnosis. Here we present a method to classify between healthy and PCa individuals using a random forest classifier based on short K length sequences (kmers) obtained from the fragment ends of the sequenced cell-free DNA

from plasma samples. Here, we used plasma cfDNA samples from 43 prostate cancer patients and 83 healthy individuals that were prepared using SRSLY®, a single-stranded library preparation method that retains the native sequence on the 5' and 3' end unlike traditional library preparation methods that lose the 3' end sequence during the prerequisite end-repair step. The libraries were sequenced to a depth of 5-300 million reads per sample. We retrieved kmers from the 3' fragment ends of both reads - specifically all K length sequences between K bases into the read, and K bases into the genome. This generated one read kmer, one genomic kmer and K-1 junction spanning kmers for each terminus. For each kmer location, we then created a position weight matrix using kmer counts. We performed classification using a random forest machine learning model that was trained on 70% of the data and tested on the remaining 30%. We found the strongest performance with kmers of 3 bases, with in an accuracy of 0.932, sensitivity of 0.867 and specificity of 0.965 in classifying healthy vs PCa. We identified that the read2 read kmers had the strongest influence on classification. We subsampled our libraries to different depths and observed that we maintained our classification metrics below 100,000 reads. Through these results, we have identified a robust machine learning based signal that is localized to the 3' end of plasma cfDNA fragments that can distinguish healthy and prostate cancer samples with high accuracy, at low depth of sequencing. Interestingly, when we subset the fragments based on different genomic regions we found no distinguishable difference in the classification results; suggesting that this signal is derived from global, genome-wide changes in cfDNA fragmentation that are specific to prostate cancer.

Session Title: Cancer Poster Thursday Session

Board 8034T: Identification of Gene Fusions Using RNA-Seq in a Large Cohort of CLL Tumors. ⭐

Authors:

N. Fadra[1], D. O'Brien[1], C. McCabe[1], S. Brown[1], T. Ma[1], E. Braggio[2], S. Slager[1]; [1]Mayo Clinic, Rochester, MN, [2]Mayo Clinic, Scottsdale, AZ

Abstract:

Chronic Lymphocytic leukemia (CLL) is a B-cell malignancy characterized by a high degree of clinical variability. Gene fusion events, resulting from the combination of two independent genes in the genome, are known to drive tumor progression in over 15% of cancers by creating defective oncogenic proteins or by dysregulating gene expression. These events are novel potential biomarkers or therapeutic targets but remain poorly characterized in CLL. In this study, we applied a comprehensive bioinformatics method to RNA sequencing data from 676 samples

to identify and characterize gene fusions events in CLL. Paired end RNA-seq reads were aligned to the human genome reference build hg38 using the STAR aligner, followed by identification of gene fusions using the STAR fusion caller. Fusion candidates were then annotated and filtered using a custom tiered bioinformatics approach to include fusions present within annotated RefSeq/Ensemble genes along exon-exon boundaries, calculated to be in-frame, and supported by a total of at least 5 reads. We excluded read-through events, fusions known to be recurrent artifacts in cancer specimens, and those present in normal tissues from the GTEx consortium. Candidates present in <10% of the cohort were then evaluated for changes in gene expression, and further validated by RT-PCR and Sanger sequencing. We identified 17 distinct gene fusion events occurring in relevant genes in B-cell biology. Among these, notable fusion candidates include *B2M-IGHM, BIRC3-RNF34*, and *IGH1-ZFP36L1*. The *IGH1-ZFP36L1* fusion results in threefold overexpression of *ZFP36L1*, a gene involved in mRNA decay machinery and the negative regulation of hematopoietic cell differentiation. The *BIRC3-RNF34* fusion was observed to silence *BIRC3*, resulting a twofold decrease in gene expression. Low expression of *BIRC3* in CLL is known to be associated with a more aggressive clinical course and chemotherapy refractoriness. RNA-seq is a reliable method for detecting gene fusions, yet its application in CLL has predominantly focused on splicing and gene expression analyses. In this study, we identified 17 distinct gene fusions in a large cohort of CLL tumors, highlighting key genes relevant to the disease. These findings underscore the importance of mining gene fusions in CLL, as they hold significant potential as therapeutic targets and biomarkers. Further investigation is essential to fully understand the impact of these fusions on CLL pathogenesis and disease progression. The validation of these events using orthogonal techniques underscores the robustness of our findings and the effectiveness of the computational approaches used

Session Title: Cancer Poster Thursday Session

Board 8035T: Identification of rare non-coding variants in familial non-medullary thyroid carcinoma

Authors:

T. T. Nieminen[1], D. F. Comiskey[2], S. Liyanarachchi[2], J. Wu[2], M. Sheikh[2], I. Hendrickson[2], P. Brock[2], M. Ringel[2]; [1]Univ. of Helsinki, Helsinki, Finland, [2]The Ohio State Univ., Columbus, OH

Abstract:

Background: Familial non-medullary thyroid carcinoma (FNMTC) occurs when three or more family members are affected by usually papillary thyroid carcinoma (PTC), the most common form of NMTC. While the genetic predisposition to NMTC is among the highest of all cancers,

the genetic determinants among NMTC families is not well understood. Here we aim to understand the contribution of rare non-coding germline variants in the etiology of FNMTC. **Methods:** We performed whole-genome sequencing (WGS) and linkage analysis in 17 PTC families in whom prior studies failed to disclose a culpable variant. In total, we found 41 protein coding region variants in 40 genes that co-segregated with family members affected with PTC. We were unable to find putative pathogenic protein coding variants in 6 of the 17 PTC families. Therefore, we further leveraged our WGS data to include non-coding variants in our analysis for all 17 families. To study tens of thousands of non-coding variants we adopted a unique filtering strategy to identify variants that occurred in known DNA elements and transcription factor binding sites, near regions that have a significant impact on gene expression or splicing in thyroid tissue, and/or in characterized thyroid enhancers. We annotated variants by using two analyses (ENCODE and transcription factor binding site) within BasePlayer software. We separately analyses (1) expression quantitative trait loci (eQTL), (2) splicing quantitative trait loci (sQTL), and (3) thyroid enhancers. We then ranked variants according to predicted pathogenicity and performed Sanger sequencing in all individuals of each family. **Results:** In total, 121 variants were selected based on *in silico* prediction and our custom ranking analysis in each pedigree. Finally, we identified 56 variants that showed co-segregation among all PTC-affected individuals and were absent from unaffected individuals, including the five of six PTC families for whom no protein coding variants were previously found. **Conclusion:** Our hypothesize-driven approach accounted that the most likely place for pathogenic non-coding variants was in either in theoretical (ENCODE) or empirical-determined (eQTL, sQTL, thyroid enhancer) regulatory regions that demonstrate at a minimum, basal thyroid expression, a positive family linkage score, and co-segregation among PTC affected individuals. Our data affirm that non-coding variants are important in the etiology of FNMTC and provide a framework for identifying pathogenic non-coding germline variants. Further studies are needed to functionally characterize these variants to better understand the molecular mechanism of their pathogenicity.

Session Title: Cancer Poster Thursday Session

Board 8036T: Identifying patterns in testing and prevalence of gBRCA mutations among women with breast cancer who underwent hereditary cancer germline testing. A cross-sectional observational study

Authors:

R. Pinninti[1], S. A. Syeda[2], K. Madamchetty[1], S. Kodandapani[1], S. J. Rajappa[1]; [1]Basavatarakam Indo-American cancer Hosp. & research Inst., HYDERABAD, India, [2]Indian Inst. of Publ. Hlth., Hyderabad, India

Abstract:

BackgroundPathogenic mutations in BRCA genes can heighten the risk of several cancers among carriers. About 10-15% of all breast cancers can have a familial link with a majority from mutations in the BRCA genes. Patterns for germline BRCA mutation testing in breast cancer have recently evolved to facilitate therapy decisions with PARP inhibitors.**Methods**We aim to assess the germline testing patterns of physicians in our hospital by examining the clinical profile of breast cancer patients who underwent testing between Jan'21 & Sept'23. Patients who underwent hereditary cancer multi-gene (30 gene panel) mutation testing using next-generation sequencing were included in the study. A simultaneous analysis was performed with multiplex ligation-dependent Probe Amplification for BRAC1 & BRCA2, deletions & duplications. The classification of the variants as pathogenic & variants of uncertain significance (VUS) was determined by the guidelines presented by the American College of Medical Genetics & Genomics.**Results**In the study period, 303 patients had undergone germline testing. The median age of the cohort was 48 years (SD 10.1, range 20 - 75). The majority were young, 178 (58.7%) were < 50 years old, & 175 (57.8%) had a family history of cancer. Specifically, a family history of breast cancer was reported in 92 (30.4%). Bilateral breast cancer was noted in 15 (5%), while ovarian cancer was reported in 8 (3%) patients. At testing, 59(22%) had metastatic disease. Triple-negative (TNBC), hormone receptor-positive, & HER 2 positive were reported in 45.5%, 34.9%, & 15.5% respectively. Pathogenic germline BRCA mutations were detected in 45 (14.8%) patients (BRCA1 9.2% & BRCA2 5.6%), & most had TNBC (32/45, 71.1%; p=0.001). The highest prevalence was seen among TNBC (32/138, 23.2%), "TNBC plus < 50yrs" (22/73 - 30.1%), & "TNBC plus <50yrs plus family history of cancer" (14/40, 35% p<0.000). VUS was noted in 26 (8.5%, 4 BRCA1 & 22 BRCA2). Non-BRCA germline mutations detected were 2 each in RAD50 & MUTYH, & 1 in MRE1 gene.**Conclusions**The majority of those tested were young, had TNBC & had a family history of cancer, reflecting the prevailing guidelines for germline testing in the study period.

Session Title: Cancer Poster Thursday Session

Board 8037T: Identifying variance in pathway activity across grade and stage in Human Hepatocellular Carcinoma ⭐

Authors:

A. Garrison[1], K. Buetow[1,2], J. Brill[1]; [1]Arizona State Univ. Sch. of Life Sci., Tempe, AZ, [2]Sch. of Complex Adaptive Systems, Tempe, AZ

Abstract:

While only the sixth most common cancer globally, liver cancer is the third most deadly. Despite the importance of accurate diagnosis and effective treatment, standard diagnostic tests for most solid organ neoplasms are not required for the most common type of liver cancer, Hepatocellular Carcinoma (HCC). In addition, major discrepancies in the practices currently in place limits the ability to develop more precise oncological treatment and prognosis. This study aimed to identify biomarkers, with potential to more accurately diagnose disease progression and determine prognosis. Pathway centric analysis was utilized to calculate activity metrics for 1,324 biological pathways in 374 The Cancer Genome Atlas (TCGA) hepatocellular carcinoma donors. Further statistical analysis was done on two datasets, formed to identify grade or stage at time of diagnosis. The datasets were evaluated individually. Based on the variance and normality of each pathway's activity levels in the respective data sets analysis of variance, Tukey-Kramer, Kruskal-Wallis, and Mann-Whitney-Wilcox tests were performed, when appropriate, to determine any statistically significant differences in pathway activity levels. 20 pathways were identified as showing significant activity level across stage only, 77 across grade only, and 23 pathways were shown to have significant patterns across both stage and grade. One pathway, role of ran in mitotic spindle regulation biocarta, showed significant increase in activity as both stage and grade increase. 22 other pathways, that show a significant decrease in activity as both stage and grade increase, are known to regulate the immune system, metabolism, and transportation of small molecules. Notably, the primary bile acid biosynthesis pathway shows a decreasing trend across all stages and grades with no discrepancies. While some overlap is seen, there was a significant number of pathways unique to either stage or grade. These pathways are known to affect the cell cycle, cellular transport, disease, immune system, and metabolism regulation. The biological pathways named by this research depict prospective biomarkers for progression of hepatocellular carcinoma per subdivision within both stage and grade. Notably, the number of significant pathways that differ across grade surpass that of stage, which supports the ability of pathway analysis to better predict cancer treatment outcome over traditional evaluation based on stage alone. Furthermore, the pathways identified herein have potential to aid in the development of new methods of early and more accurate diagnosis, as well as targeted treatment based on the affected biological processes.

Session Title: Cancer Poster Thursday Session

Board 8038T: Impact of Exposure to Fire Smoke on miRNA Expression in Firefighters' Blood and Skin Samples

Authors:

C. Xu[1], J. Gipson[1], R. J. Agnew[2], W. M. Freeman[3,4], X. Xu[5], J. Hwang[6]; [1]Univ. of Oklahoma Hlth.Sci., Oklahoma City, OK, [2]Oklahoma State Univ., Stillwater, OK, [3]Oklahoma Med. Res. Fndn., Oklahoma City, OK, [4]Oklahoma City VA Med. Ctr., Oklahoma City, OK, [5]Shanghai Univ. of Sport, Shanghai, China, [6]Univ. of Texas Hlth.Sci. Ctr., Houston, TX

Abstract:

Firefighters are recognized as working in high-risk environment exposures. These high-risk exposures increase the firefighters' risk in disease, including physical exposures, such as heat and fallen debris, as well as chemical exposures, like the byproducts of incomplete combustion - Polycyclic Aromatic Hydrocarbons (PAH). Recent study confirms that firefighters are at an increased risk for all types of cancer when compared with the general population. Exposure to PAH has been associated with an elevated incidence of cancer in firefighters. Significant levels of PAH have been found on skin, a surface of personal protective equipment following fire activities. Along with inhalation, skin absorption is a primary route of PAH exposure. Previous studies have shown changes in microRNA (miRNA) expression between experienced firefighters and new recruits in whole blood samples. In our current study, we aim to investigate the direct impact of the fire smoke exposure on miRNA expression in both blood and skin samples of firefighters, which may help explain the elevated health risks of firefighters from a genetic perspective. We followed 25 firefighters and collected blood and skin samples before and after fire activities, resulting in a total of 82 samples. The paired miRNA sequencing data were analyzed using DESeq2. We modeled sample type (Skin vs Blood) and exposure status (Post- vs Pre-fire suppression) simultaneously, controlling for the top two principle components to adjust for unobserved confounders and batch effects. Based on the significance level of q-value <0.05, we identified 5 miRNAs (*miR-125a-3p*, *miR-548a-5p*, *miR-1272*, *miR-362-5p*, and *miR-421*) that were significantly differentially expressed post-fire exposure compared to pre-exposure, and 6 miRNAs (*miR-451a*, *miR-216a-5p*, *miR-1286*, *miR-28-5p*, *miR-346*, and *miR-421*) that were significantly differentially expressed between skin and blood samples. There was no significant interaction between sample type and exposure status. Further, we examined the functions of the significant miRNAs and their target messenger RNAs (mRNAs) using gene set enrichment analysis and annotation from Ingenuity Pathway Analysis (IPA). The downstream bioinformatics analyses revealed several experimentally validated target mRNAs are known biomarkers for multiple diseases, including acute myeloid leukemia, melanoma, pulmonary hypertension, and acute respiratory distress syndrome. In general, our study revealed miRNA expression change

due to fire smoke exposure in both blood and skin samples, advancing our understanding on the impact of firefighters' environmental exposures on their health.

Session Title: Cancer Poster Thursday Session

Board 8039T: Insights into the Development of Metastasizing Uveal Melanoma by BAP1 knockdown of a Primordial Eye Organoid Model ⭐

Authors:

P. Begani, A. Z. Eriksen, T. A. Blenkinsop, A. M. Bowcock; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Uveal melanoma (UM) arises from neural crest cells within the eye and has a mortality rate of 80%. Almost all UMs harbor oncogenic mutations in *GNAQ, GNA11* or *CYSLTR2*. However, the metastatic competency of UM is due to the presence of secondary mutations. Alterations usually arise in *SF3B1* or *EIF1AX*, conferring low metastatic risk, or *BAP1*, conferring high metastatic risk. BAP1 is a ubiquitin carboxy-terminal hydrolase regulating multiple cellular processes including chromatin modification, DNA repair and the cell cycle. It is hypothesized that loss of *BAP1* in uveal melanocytes leads to a more stem cell like phenotype, leading to tumorigenesis and metastasis. However, this pathway is poorly understood. We investigated this in a primitive eye organoid model termed Self-formed Ectodermal Multi-Zone (SEAM). The ocular SEAMs arise when hiPSCs spontaneously differentiate to form four concentric zones. Cells within each SEAM zone reflects early states within the eye. BAP1 localized to Zones 1 and 2 which are associated with neural/neural-derived cells. Zone 2 developed dense melanin pigmentation at around Day 20 of differentiation reflective of neural crest cell differentiation into melanocytes. We downregulated *BAP1* in hiPSCs with the CRISPR-Cas9 system and differentiated them into ocular SEAMs. Bulk RNA-sequencing from days 0 to 50 (5-day intervals) revealed significant alterations in various pathways such as melanogenesis, morphogenesis (WNT signaling), ECM assembly and degradation, chromatin modeling and DNA repair resulting from *BAP1* knockdown. This also led to reduced melanin pigmentation in Zone 2, consistent with a role of *BAP1* in melanogenesis. Single-cell RNA sequencing of SEAMs at day 23 revealed gene clusters corresponding to different tissues of the eye including cornea, sclera, oligodendrocyte precursors, nerves, and fibroblasts. There were also three melanocyte clusters that corresponded to choroidal melanocytes (uveal melanocytes), pigmented epithelial cells of the ciliary body and retinal pigment epithelium (RPE). *BAP1* knockdown led to downregulation of common pigmentation and melanocyte markers in the uveal melanocyte cluster, but not in

the other melanocyte or pigmented epithelial cell clusters. Ocular SEAMs therefore provide important insights into the role of *BAP1* in the developing eye and in tumorigenesis by revealing BAP1 targets in specific melanocyte subpopulations that may correspond to progenitors of UM.

Session Title: Cancer Poster Thursday Session

Board 8040T: Intrinsic subtype distributions across inherited breast cancer genes: an opportunity to refine treatment?

Authors:

T. Pal[1], S. Reid[1], E. Teslow[2], M. Huang[2], B. Yilma[2], S. Fragkogianni[2], T. Ahmed[2], J. Whisenant[1], A. Weidner[1], P. Rajagopal[3], M. Stoppler[2], C. Chao[2], A. Kurian[4], A. Arafa[5], E. Antonarakis[5], S. Yadav[6], F. Couch[6]; [1]Vanderbilt-Ingram Cancer Ctr., Nashville, TN, [2]Tempus AI, Chicago, IL, [3]Natl. Cancer Inst., Bethesda, MD, [4]Stanford Women's Cancer Ctr., Palo Alto, CA, [5]Univ. of Minnesota Masonic Cancer Ctr., Minneapolis, MN, [6]Mayo Clinic, Rochester, MN

Abstract:

Introduction: Approximately 5-10% of breast cancers (BC) are inherited, primarily due to germline pathogenic/likely pathogenic variants (GPV) in high and moderate penetrance genes. We sought to compare intrinsic BC subtypes from females with GPVs in these genes and conduct subgroup analysis among hormone receptor positive (HR+, estrogen and/or progesterone positive); HER2 negative (HER2-) BC, comparing distribution of intrinsic subtypes compared to sporadic BC.
Methods: A retrospective cohort of female BC patients with GPV in *BRCA1*, *BRCA2*, *PALB2*, *ATM* or *CHEK2* detected incidentally by tumor-normal matched Tempus xT assay or via a validated germline test, and tumor testing including whole transcriptome RNA expression analysis via the xR assay underwent PAM50 subtyping to determine intrinsic subtypes (i.e., Luminal A, Luminal B, Basal, HER2-enriched). Intrinsic subtype distribution was compared across inherited genes, and to sporadic cases. Subgroup analysis was conducted in the HR+/HER2- subgroup, comparing intrinsic subtype distribution for each gene to sporadic BCs.
Results: Among 97 *BRCA1*, 124 *BRCA2*, 68 *PALB2*, 52 *ATM*, 78 *CHEK2*, and 4599 sporadic BC patients, the median age was 47, 49, 53, 52, 57 and 57, respectively. Distribution of the 4 intrinsic subtypes showed that basal subtype encompassed a majority in *BRCA1* carriers (74.4%) and was less common in *ATM* (6.7%) and *CHEK2* (8.5%) carriers. In contrast, the Luminal A subtype encompassed a higher proportion in *ATM* (60.0%) and *CHEK2* (75%) carriers, compared to *PALB2* (52.5%), *BRCA2* (48.6%) and *BRCA1* (11.0%). Among the HR+ subgroup, basal and

Luminal B subtypes were over-represented among 21 *BRCA1* (43%; n=9 and 33%; n=7 respectively) compared to sporadic (11%; n=207 and 22%; n=398 respectively). Among HR+ Luminal A subtype, *CHEK2* tumors (n=41) were over-represented (80%; n=33) while *BRCA1* tumors were under-represented (24%; n=5), compared to sporadic tumors (60.4%; n=1093).

Conclusion: Our findings demonstrate significant differences in the distribution of intrinsic subtypes across inherited BC genes, with the basal subtype seen predominantly in *BRCA1*, and under-represented in both *ATM* and *CHEK2*. Among the HR+ subgroup, basal subtypes remained over-represented in *BRCA1* carriers; and Luminal B subtype was also over-represented. Identification of non-Luminal A tumors based on intrinsic subtyping may be of both prognostic and predictive importance, with consideration of more aggressive treatment. Consequently, our findings highlight the importance of intrinsic tumor subtyping to identify aggressive tumors over-represented among females with inherited BC due to *BRCA1*, *BRCA2*, and *PALB2* GPVs.

Session Title: Cancer Poster Thursday Session

Board 8041T: Leveraging machine learning to resolve biological confounding and expedite lymphoma diagnosis

Authors:

N. Xiong[1], T. Chen[1], L. McBride[1], M-X. Wu[1], E. Cohen[1], R. Redd[1], L. McGrath[1], D. Cheddie[2], M. van der Schyff[2], C. Slome[2], M. Tolstorukov[1], M. A. Murakami[1], M. Merrill[1], K. Antel[1,2]; [1]Dana-Farber Cancer Inst., Boston, MA, [2]Univ. of Cape Town, Cape Town, South Africa

Abstract:

Machine learning can augment methodological advances in sequencing and bioinformatics to overcome diagnostic challenges in cancer, including the need for invasive tissue biopsy, highly specialized pathology expertise, and confounding by disorders with similar clinical and molecular features. This is particularly significant in HIV- and TB-endemic regions such as sub-Saharan Africa, where diagnosing lymphoma, which is potentially curable if detected early, is hindered by resource constraints and overlapping symptoms with highly prevalent infectious diseases. We developed a machine learning model that uses the outputs of a targeted cell-free DNA sequencing assay, including recurrent genetic alterations and B-cell receptor clonotypes, to detect lymphoma from blood plasma samples. The model was trained on data from 60 cancer-free individuals, 792 individuals with various solid tumors, and 852 individuals with several subtypes of B-cell lymphoma. The cohort was split into training (80%) and testing (20%) sets and trained with several machine learning algorithms including logistic regression, random forest, k-

nearest neighbors (KNN), support vector machine (SVM), and extreme gradient boosting (XGBoost) to distinguish lymphoma from non-lymphoma samples. We achieved the best results with the SVM classifier, obtaining an ROC-AUC of 0.86 and sensitivity and specificity values of 0.84 and 0.75, respectively. The SVM model was then evaluated on 10 blood plasma samples from patients with known diagnoses (8 lymphoma, 2 healthy) from Groote Schuur Hospital at the University of Cape Town. The model achieved an ROC-AUC of 0.94 and sensitivity and specificity values of 0.88 and 1 on the validation set. Our results, while preliminary, demonstrate the potential of incorporating data from multiple cell-free DNA sequencing modules and using machine learning to diagnose lymphoma in a manner that is timely, minimally invasive, and applicable to low-resource settings.

Session Title: Cancer Poster Thursday Session

Board 8042T: Long non-coding RNA expression profiling of breast cancer in Moroccan women ⭐

Authors:

S. Ayoubi[1], P. Ostano[2], M. Mello-Grand[2], I. Gregnanin[2], M. Elkarroumi[3], S. Nadifi[1], G. Chiorino[2]; [1]Genetics and Molecular Pathology Lab, Med. Sch., Univ. Hassan II of Casablanca, Casablanca, Morocco, [2]Cancer Genomics Lab, Fondazione Edo ed Elvo Tempia Valenta, Biella, Italy, [3]Mohammed VI Ctr. for Cancer Treatment, Univ. Hosp. Ctr. Ibn Rochd, Casablanca, Morocco

Abstract:

INTRODUCTION: Long non-coding RNAs (lncRNAs) are non-protein coding genomic transcripts longer than 200 nucleotides. LncRNAs have recently emerged as a new heterogeneous class of non-coding RNAs, involved in gene regulation either at the transcriptional or the post-transcriptional level. The role of lncRNAs in breast carcinogenesis has been widely studied throughout the world, but there is no study performed in developing countries like Morocco regarding the involvement of lncRNAs in breast cancer. Hence, we conducted this transcriptomic study to analyze the implication of lncRNAs in breast cancer molecular subtypes in the Moroccan population. EXPERIMENTAL MODEL: We analyzed lncRNAs differential gene expressions of 30 fresh breast tumors and 13 normal matched breast tissues by Agilent Human Genome Microarrays. Functional analysis of differentially expressed lncRNAs was carried out for the four molecular subtypes (luminal A, luminal B, HER2 over-expressed, and triple-negative breast cancer).RESULTS: A total of 159 lncRNAs were differentially expressed in all molecular subtypes combined (logFC>1; P value<0.01). Interestingly, HER2 and luminal A subtypes involved the highest number of differentially expressed lncRNAs, followed by luminal B and

triple-negative subtypes. MALAT1 lncRNA was strongly up-regulated in all molecular subtypes. While MIR497HG lncRNA was commonly under-expressed in the four breast cancer types. CONCLUSION: This study is to our knowledge the first to analyze the lncRNAs differential expression in Moroccan women. Focusing on the lncRNAs expression machinery may help to understand the mRNAs-lncRNAs interactions better and to explore new possibilities for breast cancer targeted therapy.

Session Title: Cancer Poster Thursday Session

Board 8043T: Machine learning-inferred family history of cancers correlated with smoking behavior and tumor characteristics in lung cancer

Authors:

S. Mukherjee[1], A. Schroeder[2], S. Chatterjee[1], J. Cadley[1], Y. Kemel[1], J. Jee[1], C. Fong[1], J. Woo[1], C. Falcon[1], M. Mehine[1], C. Bandlamudi[1], D. B. Solit[1], M. Ladanyi[3], M. F. Berger[1], Z. Stadler[2], M. G. Kris[1], D. Jones[4], D. R. Jones[1], A. Schoenfeld[1], F. Polubriaginof[2], J. Mak[1], A. Chatterjee[1], M. Donoghue[1], N. Schultz[1], C. M. Rudin[1], K. Offit[2]; [1]Mem. Sloan Kettering Cancer Ctr., NEW YORK, NY, [2]Mem. Sloan Kettering Cancer Ctr., New York, NY, [3]Mem Sloan Kettering Cancer Ctr, New York, NY, [4]

Abstract:

While smoking is considered as the primary risk factor of lung cancer, family history also plays a significant role with 2-3-fold increased risk. Our study aims to extract patients' characteristics including family history of cancers and smoking behavior accurately from the Electronic Health Record (EHR) using weak supervision -a machine learning (ML) method and correlate with their germline and somatic mutational profile, tumor mutation burden (TMB), chromosomal instability and somatic mutational signature. We analyzed 7150 patients who underwent targeted clinical sequencing interrogating approx. 500 genes. The weak supervision method used a combination of generative model for intermediate labels from the unstructured clinical notes extracted from EHR and subsequently tuned by a ML classifier to generate the final class labels. A randomly sampled set of 500 patients was manually curated to determine the performance of ML method. The weak supervision classifier had almost perfect performance for 2-label classes (ever smokers and never smokers) with macro F1-score: 97.7%; for 3-label classes (never smoker, former smoker, and current smoker) it was 79.8% and for estimating family history of patients with close relatives (1st and 2nd degree) with lung cancers, F1-score was 85.98% and balanced accuracy 88.4%. Analyzing genomic data, we observed that smoking

status was associated with TMB and the smoking mutational signature 4 in a dose-dependent manner (p-value < 2e-16). Among 6594(92%) patients with ML-inferred family history of cancers, 30% of them had at least 1 close relative diagnosed with lung cancer and 4.6% had > 2 close relatives with lung cancer. Prevalence of family history of lung cancer was significantly higher in females diagnosed with lung adenocarcinoma compared to males (32.4% females vs 27% males, p-value < 2e-5) and younger patients (p-value &lt 0.05). *KRAS* and *STK11* driver mutations were significantly enriched in patients with family history of lung cancer supporting the strong correlation observed between smoking behavior and the burden of cancer in the family. Significantly higher proportion of patients with family history of cancers consented for germline genetic testing and harbored germline pathogenic/likely pathogenic variants in cancer predisposing genes compared with patients without any family history (20% vs 8%, p-value < 0.01). In conclusion, we demonstrated high performance of ML approach for automated curation from EHR that significantly reduced the time and cost. Our study revealed distinct genomic patterns and smoking behaviors associated with family history of cancers in lung cancer patients.

Session Title: Cancer Poster Thursday Session

Board 8044T: MC4R rs79783591 Founder Mexican Variant Related to Aggressive Phenotype in Testicular Germ Cell Tumors

Authors:

S. Harari Arakindji[1], B. Cuevas-Estrada[2], J. Ríos-Rodríguez[3], P. Ostrosky-Wegman[4], R. González-Barrios[2], T. Wegman-Ostrosky[5]; [1]Univ. Anáhuac México Norte, Mexico City, Mexico, [2]Laboratorio de Regulación de la Cromatina y Genómica, Inst. Natl. de Cancerología, Mexico City, Mexico, [3]Tecnológico de Monterrey, Escuela de Med. y Ciencias de La Salud, Mexico, Mexico, [4]Inst. de Investigaciones Biomédicas, Univ. Natl. Autónoma de México, Mexico City, Mexico, [5]Precision Med. Lab., Inst. Natl. de Cancerología, Cdmx, Mexico

Abstract:

Testicular germ cell tumors (TGCT) represent the most frequently malignant neoplasm in young males. In general, TGCT has a good prognosis, with evidence of >80% 5 year overall survival, and up to 80% cure rates after first line treatment. In Mexican patients the prognosis is notably poorer. Recognized prognostic factors include non-seminomatous (NS) histology, advanced stage at diagnosis, younger ages, and lower education levels. Germline variants of TGCT are largely unknown. The Mexican founder variant rs79783591 (c.806T>A; p.Ile269Asn) in the *MC4R* gene has been associated with child and adult obesity. *MC4R* is expressed in germ

cells in the testis. Studies reported that a point genetic variant in *MC4R* is responsible for experimental testicular teratoma formation in mice. **Objective:** The objective of this study is to investigate whether patients with TGCT who are carriers of the rs79783591 variant, present a more aggressive phenotype. **Method:** In 238 TGCT Mexican patients, the rs79783591 variant was detected in 9 of them through germline exome sequencing and genotyping. The frequency within our cohort was compared to that of the Mexican population (PAGE Study) using the Fisher 's exact test. **Results:** The difference in frequencies between our cohort and the Mexican population was statistically significant (<0.00001).The phenotype of the 9 patients who carried rs79783591 was as follows: The mean age at diagnosis for all carriers ranged from 18 to 28 years, with a mean age of 22. Furthermore, 88.8% of these patients exhibited a non-seminoma histological type. Eight patients received systemic therapy, including chemotherapy and, in some instances, radiotherapy. Additionally, seven patients presented with metastases, predominantly located in the lungs and retroperitoneum. Notably, metastases were observed in the lumbar and thoracic spine in two patients, while one patient displayed brain metastases. **Conclusion:** Within our cohort, an increased aggressiveness in phenotype was noted among patients harboring the variant, manifesting in earlier diagnostic ages, heightened metastasis rates, and chemoresistance. Despite the variant's affiliation with obesity, only one patient exhibited a BMI surpassing 30. The identification of this variant in TGCT patients marks a pioneering observation. Given its increased prevalence in the cohort compared to the Mexican frequency, further detailed investigation is required.

Session Title: Cancer Poster Thursday Session

Board 8045T: Modulation of Oncogenic Regulatory Networks by Exosomal miRNAs in Patients with Non-Small Cell Lung Cancers. ⭐

Authors:

A. Hassanin[1,2], K. Ramos[1]; [1]Texas A&M Inst. of BioSci.s and Technology, Houston, TX, [2]Dept. of Animal Wealth Dev., Faculty of Vet. Med., Suez Canal Univ., Ismailia, Egypt

Abstract:

Lung and bronchus cancers remain the deadliest cancers in the US and worldwide, accounting for 24% of all cancer deaths, with greater than 70% of these deaths attributed to lung cancer metastasis. In recent years, exosomes have received increasing attention as a sensitive, specific, and minimally invasive liquid biopsy of cancer. Previously, we assessed the diagnostic and prognostic utility of Long Interspersed Element-1 (LINE-1) and associated microRNAs (miRs) in plasma-derived exosomes from patients with non-small cell lung cancers (NSCLC). LINE-1 ORF-1

and ORF-2 mRNA and protein, along with a panel of eight microRNAs (miR-21-5p, miR-126-3p, miR-210-3p, miR-221-3p, Let-7b-5p, miR-146a-5p, miR-222-3p, and miR-9-5p), were detected in female and male patients with squamous cell lung cancer (SQCLC) and lung adenocarcinoma (LUAD) at all cancer stages. Differential levels of these analytes in exosomes correlated with sex-specific clinicopathologic characteristics. Little is known about how LINE-1 influences the expression of functional non-coding RNAs in NSCLC. To this end, studies were conducted to explore genetic interactions between genes within the LINE-1 regulatory network and exosomal miRs in lung cancer patients. Nine genes (ICAM1, AGL, PKIA, RPM39, RGS3, RGS13, VAMP3, VCAM1, and TGFβ1) within the LINE-1 regulatory network were identified as targets for exosomal miRs. These findings suggest that functional interactions between exosomal LINE-1 and miRs, may impact oncogenic signaling in NSCLC. To test this hypothesis, plasma exosomes from ostensibly healthy subjects or early and late-stage NSCLC lung cancer patients matched for demographic characteristics were co-cultured with non-tumorigenic BEAS-2B cells for 72h. Cells co-cultured with plasma exosomes from late-stage NSCLC patients exhibited high levels of LINE-1 ORF1 and ORF2 mRNAs and their corresponding proteins, as well as increased levels of exosomal miRNAs compared to untreated cells. In sharp contrast, exosomes from ostensibly healthy and early-stage NSCLC markedly inhibited LINE-1 and all genes within its genetic regulatory network. We conclude that exosomes readily impact regulatory control of the LINE-1 machinery in human bronchial epithelial cells, with late-stage NSCLC derived exosomes activating miR-regulated oncogenic signaling in non -tumorigenic recipient cells. Together, these findings raise important implications for lung cancer progression and metastasis.

Session Title: Cancer Poster Thursday Session

Board 8046T: Molecular analysis of patients with Richter's syndrome (RS) and progressive chronic lymphocytic leukemia (CLL)

Authors:

H. Yan[1], H. Wang[2], Y. Liu[1], H. Jin-Lee[1], E. Klee[1], S. Tian[1], W. Ding[1]; [1]Mayo Clinic, Rochester, MN, [2]Mayo Clinic Hlth.System, Mankato, MN

Abstract:

Richter's syndrome (RS) represents the development of an aggressive lymphoma in chronic lymphocytic leukemia (CLL), which remains an area of unmet clinical need. Also, effective therapies are lacking for progressive CLL who have developed resistance after targeted therapies. Hypothesizing that progressive CLL and RS possess distinct molecular abnormalities, our study aimed to dissect the genomic alterations in a small cohort of patients. RNA

sequencing was performed using node tissues from aggressive CLL (n=5) and RS (n=7) patients. Unsupervised clustering identified two clusters with distinct expression profiles, with five RS in cluster C1 and the other seven in C2 including two RS and all five CLL. Within C2, two of the five CLL patients who ultimately developed RS are more similar to the two RS in expression profile, suggesting pre-existence of transcriptomic signature for RS. The analysis identified 1,431 discriminatory genes from autosomes, including 840 genes up-regulated in the 3-sample subcluster within C1, 474 up-regulated in C1 and 117 up-regulated in C2. Overall, C1 showed elevated expression across majority of the differentially expressed genes between the two clusters. These genes are enriched in PI3K-AKT, NOTCH, and RAS/MAPK signaling pathways known to be involved in Richter transformation or aggressive clinical course in CLL. In addition, using deconvolution of bulk RNA-seq, we observed higher abundances of macrophages, activated CD4 memory and CD8 T cells in C1 compared to C2, supporting the diversity of immune cell composition between RS and CLL as previously reported. Through Tempus gene panel DNA sequencing, we further identified 51 genes that carried copy number alterations preferentially occurring in either of the clusters. Over 80% of them were enriched in C2, mainly 17q gains in the CLL cases. Together, our findings highlight noticeable differences in gene expression, cellular composition and genomic alterations among RS and progressive CLL.

Session Title: Cancer Poster Thursday Session

Board 8047T: mSigDetect: Homologous recombination deficiency detection using mutational signatures in targeted long read sequenced tumours

Authors:

W. Ng[1], L. Wong[1], S. Chong[1], J. Ye[1], J. Ngeow[2]; [1]Lee Khong Chian Sch. of Med., Singapore, Singapore, [2]Natl. Cancer Ctr. Singpore, Singapore, Singapore

Abstract:

Background: Homologous recombination deficiency (HRD) affects cancer types including breast, ovarian, pancreas and prostate and patients with HR deficient tumours can respond to poly ADP-ribose polymerase (PARP) inhibitors and platinum-based therapies. HRD is associated to the loss of either germline or somatic, *BRCA1* or *BRCA2* and other associated variants. However, a large proportion of tumours do not have an identifiable variant in *BRCA1*/*BRCA2* and other related genes. Patterns of mutations—Mutational Signatures—of HRD can be detected by identifying changes in DNA due to defective DNA repair and genomic instability in HRD deficient tumours. Current computational tools to detect HRD are based on whole genome sequencing

(WGS) and an assessment of robustness of these approaches for clinical implementation are lacking.
Methods: We developed mSigDetect to identify HRD associated mutational signature (SBS3) in tumours, sequenced on Oxford Nanopore Technologies (ONT) targeted adaptive sampling technology. We built our mutational signatures classifier based on mutational profiles of 567 breast cancers from the International Cancer Genome Consortium (ICGC) and tested mSigDetect on synthetic tumours (n=200) and two independent cohorts of breast patients sequenced on Illumina short reads (n=224) and on ONT long reads (n = 38). Using the simulated synthetic tumours, we determined the sensitivity and specificity of our tool. We tested the performance of mSigDetect under different scenarios of tumour cellularity by down sampling and combining HR deficient cell lines with paired normal lines to test a range of tumour cellularities. Last, we developed a computational workflow for ONT sequencing, to carry out real-time variant calling and mutational signature identification to reduce the turnaround time of the diagnostic test.
Results: With a target region of 5% of the human genome sequenced, mSigDetect could identify HR deficient tumours in both synthetic (sensitivity = 0.73, specificity = 1) and breast cancer patient tumours (sensitivity = 0.81, specificity = 0.99) compared to WGS. We designed mSigDetect to detect SBS3 in a tumour-only setting, without the need of a paired normal sample. Signature detection of SBS3 using mSigDetect greatly outperformed sequencing whole exome targeted regions. Based on cell line mixing simulations, we recommend sequencing parameters for ONT adaptive sampling to identify HR deficient tumours robustly in a clinical setting.

Session Title: Cancer Poster Thursday Session

Board 8048T: Multi-omic integrative analysis uncovers key transcription factors and target proteins associated with renal cell carcinoma and its subtypes

Authors:

S. Chhetri[1,2], T. D. Winter[3], M. J. Machiela[4], A. Battle[1,5], M. P. Purdue[6], S. Chanock[3], D. Dutta[4]; [1]Dept. of BioMed. Engineering, Johns Hopkins Univ., Baltimore, MD, [2]Ctr. for Computational Biology, Johns Hopkins Univ., Baltimore, MD, [3]Lab. of Genetic Susceptibility, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, [4]Integrative Tumor Epidemiology Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, [5]Dept. of Computer Sci., Johns Hopkins Univ., Baltimore, MD, [6]Occupational and Environmental Epidemiology Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD

Abstract:

Genome-wide association studies (GWAS) have identified risk loci for renal cell carcinoma (RCC) and its major subtypes. To comprehensively characterize the landscape of molecular dysregulation in RCC, it is important to identify key transcription factors (TFs) whose DNA binding is altered by genetic variants associated with RCC and the downstream molecular targets they influence.

Here we performed a series of mixed model-based integrative analyses to test the association between 449 TF ChIP-seq profiles across 9 kidney-related cell lines and summary statistics from a multi-ancestry GWAS of 29,020 RCC cases and 835,670 controls. We identified 96 unique TFs with SNPs near TF ChIP-seq peaks significantly associated (p-value < $1x\ 10^{-4}$) with their effect on RCC, including *EPAS1*, *ARNT* and *PAX8*, previously implicated in RCC pathogenesis. Of these, we identified a subset of 77 TFs whose binding sites are in active promoter/enhancer regions in RCC tumors, providing further functional evidence for RCC risk. Further, we found the co-occupancy of 71 pairs of RCC-related TFs to be associated with RCC risk beyond individual TF effects. Analyses restricted to RCC subtypes revealed substantial heterogeneity of TF associations, unique co-occupancy patterns, and epigenomic interactions, identifying 22 TFs linked to one subtype but not overall RCC.

Subsequently, we partitioned the RCC-associated variants according to TF binding sites proximity to each of the 96 identified TFs to construct *partitioned genetic scores* aggregating the effects of TFs on RCC. We performed a summary statistics-based association model with the *partitioned genetic scores* and the levels of 2,940 plasma proteins profiled in the UK Biobank, to identify proteins influenced by the RCC-related TFs. Overall, we identified 38 proteins associated (FDR < 5%) with TF partitioned genetic scores, of which 23 were distal (>5Mb) to the constituent variants of the partitioned score. For example, we identified TLR3 and ZP3 to be associated with the partitioned scores of the TFs *EPAS1* and *ARNT*, and TYMP with that of TFs *PAX8* and *HNF1B*, indicating these proteins are likely affected by RCC-related variants disrupting binding sites of the corresponding TFs. Using in silico analysis and data from tumor molecular profiling, we provide independent evidence for several of identified proteins to be relevant to RCC pathogenesis.

Our analysis provides an agnostic interrogation of the landscape of TF effects on RCC and its subtypes and identifies downstream protein targets through a genetic score partitioning approach, offering additional insights into genetic variation in TF binding sites underlying RCC susceptibility.

Session Title: Cancer Poster Thursday Session

Board 8049T: Multi-omic profiling of tumor-derived extracellular vesicles enables early cancer detection

Authors:

M. Kinnersley, K. Havranek, C. Seibold, T. Blohm, T. Seibold, R. Short-Miller, S. Lodmell, A. Mast, R. Kemp, K. Van Vaerenberghe; FYR Diagnostics, Missoula, MT

Abstract:

Early cancer detection and diagnosis is critical for early intervention and treatment, which can improve patient outcomes and increase survival rates. Liquid biopsies using tumor-derived cell-free DNA (ctDNA) have become an increasingly common tool for minimally-invasive diagnostic testing. Despite success in identification of late-stage cancers, ctDNA assays do not consistently perform well in early detection applications and provide limited insight into mechanisms driving disease. Blood plasma extracellular vesicles (EVs), which carry heterogeneous protein, nucleic acid, and metabolite cargos derived from various cell types including cancer cells, are an attractive alternative to ctDNA that provide a richer diagnostic analyte base. Because only a small fraction of circulating EVs are derived from diseased cells, current EV-based methods for cancer detection often suffer from low signal-to-noise ratios. To circumvent this issue, we developed an EV subpopulation enrichment technology that specifically targets tumor-derived EVs. Sub-populated EV-omic molecular profiles (protein/RNA/miRNA) interrogated using machine learning and gene network analyses distinguish early stage (I and II) lung and ovarian cancer patients, patients with benign tumors and healthy control patients from one another. Known tissue markers as well as novel features provide the basis for patient stratification, and all three analytes contribute meaningfully to group separation and our understanding of patient tumor biology. Coupled with state-of-the-art proteomics and transcriptomics workflows, our novel EV enrichment platform allows us to capitalize upon the powerful diagnostic information contained within tumor-derived EVs to identify and characterize early-stage cancers using standard blood plasma samples.

Session Title: Cancer Poster Thursday Session

Board 8050T: NEK2-driven pathways: A mutli-omic analysis to uncover molecular insights into drug-resistant myeloma ⭐

Authors:

M. Bauer, K. R. Ryan, N. Garcia Garcia, H. Xu, F. Zhan; Univ. of Arkansas for Med. Sci., Little Rock, AR

Abstract:

Multiple Myeloma (MM) is a plasma cell malignancy characterized by extensive structural and numerical chromosomal abnormalities. By comparing gene expression profiles (GEP) of samples at baseline, during intensive chemotherapy, and at relapse, we identified chromosomal instability (CIN) genes associated with poor prognosis in MM. Among these, NEK2, a serine-threonine kinase, is the most significantly upregulated gene in myeloma cells from patients in complete remission (CR) with minimal residual disease (MRD) and at relapse early during chemotherapy following tandem transplantations. Our research demonstrates that both pharmacological and genetic inhibition of NEK2 delay tumor growth and induce apoptosis in myeloma cells. High NEK2 expression is also linked to poor outcomes in various cancers.
NEK2 impacts a broad range of downstream targets and gene expression through: 1) direct interaction, 2) activation of methyltransferases, 3) interaction with splicing factors leading to alternative splicing, and 4) genomic changes due to chromosomal instability. To elucidate NEK2's mechanism of action, we employed a multi-omic approach. We sequenced the DNA and RNA of two myeloma cell lines: NEK2 knocked out (ARP1 shNEK2) and NEK2 overexpressing (ARP1 OENEK2), along with their respective controls (ARP1 scrNEK2 and ARP1 EV).
The RNA sequencing pipeline included alignment to the Gencode v43 reference genome using minimap2, and quantification and analysis with Bambu, IsoQuant, ESPRESSO, DESeq2, DEXSeq, and IsoformSwitchAnalyzeR. For DNA sequencing, we used minimap2 for alignment and Modkit for methylation analysis. This comprehensive analysis revealed significant genomic, transcriptomic, and epigenomic changes attributed to NEK2 manipulation.
Our analysis highlighted NEK2's profound influence on the cellular transcriptome, revealing genome-wide transcript isoform dysregulation in ARP1 cells overexpressing NEK2 compared to control cells. NEK2 overexpression altered numerous genes and transcripts, notably affecting cell cycle regulation, apoptosis, and mRNA splicing pathways. Differential methylation analysis revealed significant regions in OENEK2 vs. EV, particularly affecting pathways involving transcription factors, Rap1 signaling, and apoptosis. One significant methylated region near the MSH2 gene, which plays a crucial role in the DNA repair pathway, underscored potential epigenetic regulatory mechanisms linked to NEK2 overexpression. These findings shed light on the molecular pathways influenced by NEK2, highlighting its role in the emergence of drug-resistant myeloma and identifying potential therapeutic targets.

Session Title: Cancer Poster Thursday Session

Board 8051T: Novel non-coding promoter mutations in *BCL2L12* and *IRF3* impacting 5% of melanoma patients ⭐

Authors:

X. Janssens[1], G. M. Marrón-Liñares[1], B. Verbelen[1], K. Timcheva[1], V. Konstantakos[2], S. Aerts[2], D. Pepe[1], K. De Keersmaecker[1]; [1]Lab. for Disease Mechanisms in Cancer, KU Leuven, Leuven, Belgium, [2]Lab. of Computational Biology, VIB-KU Leuven, Leuven, Belgium

Abstract:

In the search for driver mutations in cancers, the majority of researchers have focused on DNA mutations that alter the amino acid sequence of proteins. However, mutations in non-coding elements of the DNA can be of interest, as they can influence mRNA and protein expression of cancer genes via various molecular mechanisms.
In this study, we describe the functional impact of a previously unidentified mutational cluster of non-coding point mutations in the overlapping promotor region of the *BCL2L12* and *IRF3* genes, occurring in 5% of melanoma patients. BCL2L12 regulates caspases 3 and 7 and p53 tumor suppressor function, but its exact role as pro- or anti-apoptotic factor seems cell type dependent. *IRF3* is essential for innate immunity, regulates interferon-stimulated genes in response to DNA damage and maintains chromosome stability in cancer cells.
The three non-coding variants (chr19: 49665847 C>T, 49665874 C>T, 49665875 C>T) occur significantly more in male patients (87%) compared to female melanoma patients (13%; $P<.001$). Patients carrying either of these mutations have significantly reduced *BCL2L12* and *IRF3* mRNA expression. Additionally, the fraction of patients that were not responding or had stable disease upon administration of immune checkpoint therapy was 19% higher in melanoma patients with *BCL2L12/IRF3* promoter mutations ($P=.19$).
To further validate our findings, we modeled these mutations in a human pre-malignant melanoma cell line (Mel-ST) using CRISPR-Cas9 knock-in at the endogenous gene locus. The mutant Mel-ST cells display a significant 15% downregulation of BCL2L12 ($P<.001$) and an 8% downregulation trend ($P=.10$) of IRF3 on mRNA and protein level.Further bioinformatic analysis revealed that all three mutations disrupt or create new binding motifs for transcription factors of the ETS family. Indeed, a dual-luciferase reporter system revealed significantly lower mutant *BCL2L12* promoter activity upon introduction of the mutations ($P<.001$). Interestingly, cells containing the 49665847 C>T mutant are also significantly more sensitive to doxorubicin. Our findings demonstrate that non-coding point mutations can have a significant impact on gene expression, influencing cancer drug sensitivity and immunotherapy outcomes in

melanoma patients. This highlights the potential clinical significance of non-coding promoter mutations in *BCL2L12* and *IRF3* in melanoma.

Session Title: Cancer Poster Thursday Session

Board 8052T: Optimizing NGS-based MSI Analysis: A Novel Microsatellite Loci Selection Model for Both Tumor-normal Pair and Tumor-only Applications

Authors:

Z. Pu[1], J. Zhou[2], F. Chang[1], W. Mu[3]; [1]Kingmed, Shanghai, China, [2]KingMed diagnostics, ShangHai, China, [3]Kingmed Diagnositcs, Guangzhou, China

Abstract:

Background: Microsatellite instability (MSI) is a crucial tumor-agnostic biomarker, reflecting deficiencies in the DNA mismatch repair system. The selection of appropriate microsatellite loci is pivotal for the accuracy of MSI detection using next-generation sequencing (NGS) data. To optimize NGS-based MSI analysis, we developed a sophisticated loci selection model that significantly improves the precision of MSI detection.
Methods: We created a distinctive loci selection model based on the Mean Absolute Error and the Mann-Whitney U test. Potential distinctive MSI loci were identified by using RepeatFinder to pinpoint A/T monomer loci, which were then cross-referenced with the sufficiently covered region in relevant panels. An indicator, denoted as Pd, was utilized to determine the instability score for each microsatellite locus, with higher Pd values indicating greater instability. The Mann-Whitney U test identified loci exhibiting significant differences in Pd distribution between MSI-high (MSI-H) and MS-stable (MSS) samples in the training dataset. Loci with stable Pd in MSS samples and consistent coverage across all samples were further selected as distinctive loci.
Results: We evaluated our loci selection model using 196 tumor-normal sample pairs with known MSI status, sequenced using a 700-gene NGS panel. Of these, 35 samples were used as training data for MSI loci selection, and 161 samples were used for validation. In the validation dataset, the distinctive loci showed superior discriminatory power in separating MSI and MSS samples compared to unselected loci. Furthermore, to assess the model's performance in tumor-only samples, we applied our model to 75 tumor samples sequenced with an 88-gene NGS panel with known MSI status and used MSIsensor-pro in a tumor-only mode to determine microsatellite status. The results demonstrated that the distinctive loci consistently outperformed unselected loci in discriminating MSI and MSS samples in tumor-only MSI

detection using a smaller panel. Additionally, simulation studies confirmed that a cancer NGS panel with a minimum of 30 selective loci is required to ensure reliable MSI status discrimination. In conclusion, our model provides a robust framework for enhancing the accuracy of MSI determination through distinctive loci selection.

Session Title: Cancer Poster Thursday Session

Board 8053T: Optimizing sequencing depth for reliable ctDNA analysis in plasma samples using down-sampling strategies

Authors:

Y. Gu[1,2], J. Wong[2], S. M. Landers[3], A. Bhalla[3], K. Torres[3], P. Scheet[2]; [1]The Univ. of Texas M.D. Anderson Cancer Ctr. UTHlth.Graduate Sch. of BioMed. Sci., Houston, TX, [2]Dept. of Epidemiology, The Univ. of Texas M.D. Anderson Cancer Ctr., Houston, TX, [3]Dept. of Surgical Oncology, The Univ. of Texas M.D. Anderson Cancer Ctr., Houston, TX

Abstract:

Goal: to evaluate in silico repeatability and accuracy of low coverage copy number alteration (CNA) calling in circulating DNA. Background: Circulating tumor DNA (ctDNA) comprises short DNA fragments released by tumor cells into the bloodstream. Traditional cancer diagnosis methods, such as surgical biopsies, are invasive and painful. Liquid biopsy, utilizing ctDNA and cell-free DNA (cfDNA), presents a less invasive alternative for cancer diagnosis and metastasis monitoring. Detecting circulating tumor DNA (ctDNA) is inherently challenging due to its fragmentation, degradation, and typically low and variable concentration in the bloodstream. These factors demand methods with robust sensitivity and minimal false positives to ensure high reproducibility. The assumptions underlying ichorCNA suggest that it can estimate tumor fraction from 0.1x sequencing coverage data with high reproducibility by analyzing somatic copy number alterations (SCNAs; Adalsteinsson, 2017, Nat Commun). Method: High-depth sequencing data from seven plasma samples of NF-1 patients were downsampled to different sequencing depths (1x, 5x, 10x, 30x, 50x, and 100x) across four independent replicates at each depth. This downsampling aimed to evaluate sequencing quality, estimate the hypothetical tumor content, and investigate the variability in ichorCNA results. The consistency of the ichorCNA algorithm was assessed using the Jaccard index for pairwise comparisons at both segment and arm levels. Result: Quality control of downsampled sequencing data showed high-quality libraries with low duplication rates and a high percentage of properly paired sequences. Enriching for shorter fragments, setting a ctDNA threshold at 145bp, revealed significant variations in tumor cell fractions across samples. Samples with higher tumor concentrations, as

defined by ichorCNA and fragment size, showed consistent event callings across various depths, even at 1X, and their Jaccard indices increased with depth. In contrast, samples with lower tumor concentrations struggled with reproducibility, particularly in distinguishing amplification from neutral events, and showed no specific pattern in Jaccard indices. These observations highlight ichorCNA's depth-sensitive accuracy, influenced by tumor cell fraction. To explore this relationship, we propose subsetting sequencing data to simulate varying proportions of low and high fragment sizes across depths by mimicking changes in tumor proportion. This study will clarify how tumor fraction and sequencing depth affect ichorCNA's ability to detect genomic events and inform ctDNA experiment design.

Session Title: Cancer Poster Thursday Session

Board 8054T: Performance of various cell free DNA extraction kits in healthy plasma samples

Authors:

M. Alberts, M. Klemmensen, S. Borrowman, B. Chandra, **M. Tung**; Univ. of Iowa, Iowa City, IA

Abstract:

Precision oncology is becoming a reality with liquid biopsies being approved by the U.S Food and Drug Administration (FDA) for detecting various tumor types. Liquid biopsies can detect freely circulating cell-free DNA (cfDNA). The success of the cobas® epidermal growth factor receptor (EGFR) mutation test, which was approved in 2016 is a great example. This test leverages plasma as a companion diagnostic test for the detection of hotspot mutations in the *EGFR* gene. The success of this test paved the way for development of a wide array of blood-based cancer screens. However, relatively low concentration and the short half-life of cfDNA in plasma poses a significant challenge to extraction strategies. We sought to compare the efficiency of three different commercially available cfDNA extraction kits. Two of these kits utilized a spin column with a silica membrane for cfDNA binding, extraction, and elution: QIamp circulating nucleic acid kit, (Qiagen), and *Quick* cfDNA/RNA serum and plasma kit (Zymo). The third kit utilized magnetic beads for binding and elution in the extraction protocol: MagMAX™ cell free DNA isolation kit (Applied Biosystems). We performed this study using 12 mls of whole blood collected from two healthy donors in an EDTA tube. The primary blood collection tube was centrifuged within 30 minutes at 1,900G for 10 minutes. Plasma was carefully removed without disturbing the buffy coat, transferred to a new 15 ml tube, and centrifuged again at 3,000G for 15 minutes. The clear supernatant was transferred to a new 15 ml tube without disturbing the pellet. Plasma supernatant recovery was between 4 to 5 mLs,

and subsequently 1 mL of plasma was used for each of the spin column kit. 2 mL of plasma was used for the magnetic bead-based extraction kit. cfDNA extraction was performed as per the manufacturer's specification. The processing time was the fastest for the QIamp kit, followed by the MagMAX and finally, the Zymo kit. Of all the tested kits, QIamp kit retrieved the highest amount of cfDNA. We also evaluated the impact of freezing of plasma prior to cfDNA extraction using three frozen plasma samples from the same healthy control within a 6 to 8-months period. We found that frozen plasma did not impact the cfDNA extraction yield. Notably, we evaluated the performance of our extraction protocol, and our results were consistent with the reference standard for cfDNA recovery from Horizon (Multiplex I cfDNA in synthetic matrix II, Horizon Discovery Ltd). In conclusion, we observed that the QIamp circulating nucleic acid spin-column kit is a robust cfDNA extraction kit, with significantly higher extraction yield compared to other kits for downstream genomic analysis.

Session Title: Cancer Poster Thursday Session

Board 8055T: Phenotypic characterization of individuals with germline *POT1* variants: A retrospective analysis

Authors:

B. Schultz, A. Raper, T. Hartman, Regeneron Genetics Center, Penn Medicine Biobank, K. Nathanson; Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Introduction: The highly conserved POT1 protein binds and protects telomere ends. Although germline *POT1* variants have been associated with several cancer types, the full spectrum of *POT1*-related tumors remains incompletely described and prevalence is unknown. Herein, we describe the cancer types, genotypes, and clinical characteristics of individuals with germline pathogenic and likely pathogenic variants (P/LPVs) ascertained clinically and from the Penn Medicine Biobank (PMBB). **Methods:** Exome data were reviewed for *POT1* P/LPVs from 43,731 PMBB participants. A retrospective chart review was performed on 48 individuals with *POT1* P/LPVs, 37 identified through PMBB and 11 identified through clinical testing. **Results:** The frequency of *POT1* P/LPVs in PMBB was 1:1,151. Among the 48 individuals, thirty unique *POT1* P/LPVs were identified. High rates of cancer were observed in the clinical and PMBB cohorts (63.6% and 59.5%, respectively), with individuals in the clinical cohort more likely to be diagnosed with multiple primary cancers (54.5% vs 18.9%, p=0.03). The 48 patients were diagnosed with a combined total of 50 solid malignant neoplasms (13 types), 14 hematologic neoplasms (nine types), and 28 benign neoplasms (16 types). Along with previously

described associated cancer types (chronic lymphocytic leukemia, melanoma, thyroid cancer, and hematologic malignancies), we found high rates of prostate cancer (34.8 % of men over age 50) and non-small cell lung cancer (10.4%) in individuals with germline *POT1* P/LP variants. **Conclusions:** *POT1* P/LPVs are found at a frequency comparable to *NF1* in an unselected population, and predispose to a broad spectrum of neoplasms, which now includes prostate and non-small cell lung cancers. These findings expand our understanding *POT1* associated disease and will impact clinical management recommendations for *POT1*-tumor predisposition syndrome.

Session Title: Cancer Poster Thursday Session

Board 8056T: Populations and methods used to define prostate cancer polygenic risk score categories impact the interpretation of risk

Authors:

B. Guo[1], A. Sahimi[2], X. Sheng[3], F. Chen[3], C. A. Haiman[3], D. V. Conti[3], B. Darst[1]; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]New York Univ. Grossman Sch. of Med., New York, NY, [3]Ctr. for Genetic Epidemiology, Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California/Norris Comprehensive Cancer Ctr., Los Angeles, CA

Abstract:

Introduction: Prostate cancer (PCa) is the most diagnosed cancer in males in the US and is highly heritable. Polygenic risk scores (PRS) hold the potential to identify individuals at higher risk of PCa and inform screening strategies. However, the interpretation of PRS relies heavily on the choice of comparison population. Here, we investigated how different approaches to determine PRS risk categories impact PRS interpretation. **Methods**: We evaluated a multi-ancestry PCa PRS in 5,288 cases and 12,754 controls from self-identified European (EUR), African (AFR), East Asian (EAS), and Hispanic (HIS) populations from the Multiethnic Cohort (MEC). The PRS was constructed using 451 variants identified in a previous PCa GWAS of 156,319 cases and 788,443 controls across four continental ancestries. We categorized MEC participants into PRS deciles within their respective populations using four methods to determine decile cutoffs: 1) PRS distribution in MEC controls; 2) PRS distribution in MEC controls after adjusting the PRS for the first 10 principal components (PCs); 3) PRS distribution in external reference populations matched by continental ancestry; and 4) personalized PRS decile cutoffs that reflect each participant's unique ancestral background, calculated as a linear combination of their ancestral proportions from ADMIXTURE and decile cutoffs in external reference populations. The reference populations used for methods 3 and 4 consisted of 1,890 EUR, AFR,

EAS, and AMR individuals from 1KGP and the PAGE Global Reference Panel, limiting to those who most closely resembled each respective population based on PCA. Logistic regression models adjusted for age and the first 10 PCs were used to evaluate the association between each PRS decile and PCa risk. **Results**: We observed high variability in the predictive ability of PRS across methods and populations, with no single method consistently having the strongest OR. Compared to the 40%-60% PRS category, the OR for men in the top 90%-100% PRS decile ranged from 3.12 (95% CI: 2.13-4.57; method 4) to 4.25 (95% CI: 3.48-5.19; method 2) in EUR, 2.72 (95% CI: 1.31-5.64; method 3) to 5.51 (95% CI: 2.07-5.51; method 4) in AFR, 2.05 (95% CI: 1.32-3.16; method 3) to 4.35 (95% CI: 3.68-5.14; method 1) in EAS, and 2.06 (95% CI: 1.44-2.94; method 3) to 3.93 (95% CI: 3.27-4.72; method 1) in HIS. **Conclusions**: These findings demonstrate that the PRS interpretation and whether an individual is categorized as high, intermediate, or low risk based on their PRS relies on the populations and methods used to define PRS risk categories. These choices will have clinical implications and are crucial to consider prior to clinically implementing PRS.

Session Title: Cancer Poster Thursday Session

Board 8058T: Reclassification of VUS shows the value added of using ClinGen´s ENIGMA Expert Panel Specifications for BRCA1 and BRCA2 over the standard ACMG classification system

Authors:

A. Laner[1], A. Benet-Pages[2], T. Wohlfrom[1], L. R. Nassar[3], M. Haeussler[3], V. Steinke-Lange[1], E. Holinski-Feder[4]; [1]MGZ, Munich, Germany, [2]MGZ- Med. Genetics Ctr., Munich, Muenchen, Germany, [3]Genomics Inst., Univ. of California, Santa Cruz, Santa Cruz, CA, [4]MGZ - Medizinisch Genetisches Zentrum, München, Bavaria, Germany

Abstract:

Abstract Purpose: Variants of uncertain significance (VUS) are considered one of the biggest bottlenecks in the translation of genetic test results into precise clinical recommendations. The 2015 ACMG/AMP classification guidelines by Richards et al. established a general framework for the assessment and classification of genetic variants, yet gene-specific rules are needed to enable better variant classification to reduce the number of VUS. The process of gene-specific adaptations of the ACMG codes is led and accompanied by ClinGen and implemented by Variant Curation Expert Panels (VCEP). The ENIGMA BRCA1 and BRCA2 VCEP recently published its specifications for the *BRCA1* and *BRCA2* genes. Methods: We set out to evaluate the results of reclassifying 121 VUS identified in these genes in a clinical setting with the latest annotation

data using the standard ACMG/AMP- and ENIGMA VCEP-classification systems. To simplify the re-evaluation process we have created a special UCSC genome browser track hub which displays the exact data points required for variant classification using the ENIGMA VCEP rules at exon and variant level. Results: By comparing the codes used and their different weighting in the two approaches, we were able to demonstrate the superiority of the application of ENIGMA VCEP specifications which resulted in a dramatic reduction of VUS (81% ENIGMA VCEP vs. 17% ACMG/AMP standard). Conclusion: For the diagnostic analysis of the *BRCA1* and *BRCA2* genes, the use of the ENIGMA VCEP specifications is without alternative to achieve the best possible result in the clinical translation of genetic variations.

Session Title: Cancer Poster Thursday Session

Board 8059T: Shifting the lens: analysis of genomics cohorts beyond discrete ancestry categories reveals overlooked somatic-germline interactions across cancer types

Authors:

H. Mohsen, T. Perea-Chamblee; Mem. Sloan Kettering Cancer Ctr., New York, NY

Abstract:

A growing body of research highlighted a strong presence of continuums that describe human genomic diversity. Historical scholarship has also extensively investigated the use of discrete classification systems in human genomics further questioning their validity. To investigate the limiting effect of pre-defined stratification, we performed a wide comparative driver GWAS analysis at the trans-population (i.e. without stratification) vs continent-based population levels across 16 cancer types. Results demonstrate that in the MSK-IMPACT cohort (n = 56,826 samples), the analysis of the trans-population genomic similarity continuum captures significant somatic-germline interactions ($p < 5 \times 10^{-8}$) across cancer type-specific and pancancer cohorts, including a susbet of interactions overlooked by the discrete continent-based paradigm. We demonstrate that the allele frequencies of overlooked germline SNPs tend to have similar directional relationships with somatic alterations across pre-defined discrete categories, and that earlier stratification prevents their detection due to lowered statistical power. PheWAS analysis of detected germline variants revealed associations with cancer and non-cancer phenotypes reported in previous studies as well as novel hits. To expand the scope of the trans-population lens, we carried out a comparative analysis on 17 whole-exome TCGA cohorts to identify potential cancer drivers based on mutational frequencies. Results similarly revealed known and potential cancer drivers at the trans-population level that would be overlooked

when pre-defined continent-based stratification is applied. The lens we adopt further highlights the necessity for new quantitative perspectives to comprehensively detect genomic patterns across the variation continuum and to underscore more complex understandings of human genomic variation beyond pre-defined, discrete stratification schemes.

Session Title: Cancer Poster Thursday Session

Board 8061T: Synergistic effect of microRNA enhances target specificity with reduced cellular toxicity: A potential tool in cancer therapeutics

Authors:

R. Chatterjee, S. Mazumder; Indian Statistical Inst., Kolkata, India

Abstract:

Therapeutic application of microRNAs has emerged as a promising tool in cancer management. However, most of the clinical trials were not successful due to the cytotoxic effect imposed by larger doses of a single miRNA required for effective treatment. In this study, we have investigated the factors mediating the miRNA synergism in enhancing target specificity while reducing the off-target effects. Co-expression of miRNA-mimic pairs showed a synergistic reduction in gene expression of the common target genes having miRNA responsive elements (MREs) at a distance between 200-800bp. Deletions of either of the miRNA seed sequences resulted in a loss of synergism indicating miRNA-mediated synergistic reduction in gene expression. To further study the effect of miRNA synergism, oral cancer cells were co-transfected with reduced doses of miRNA-mimic pairs. A synergistic reduction in the gene and protein expression of the common target genes were observed while, the unique target genes did not show significant reduction in their expression. Co-expression of miRNAs also showed significant reduction in cell proliferation, migration and spheroid formation in cancer cells but no significant change was observed in the normal cells. Proteomic and metabolomics profiling showed synergistic reduction of the common oncogenic proteins and their downstream metabolites that are involved in cancer related pathways in miRNA mimic-pair compared to individual miRNA mimic treatment. A reduction in inflammatory and cellular stress induced pathways were also observed indicating an overall reduction in off-target effect and cellular toxicity potentiating the importance of this approach in effective cancer therapeutics.

Session Title: Cancer Poster Thursday Session

Board 8062T: Synergistic interactions of genetic alterations and immune microenvironmental features predict patient survival and aggressive cancer subtypes

Authors:

J. Reimand[1], M. Slobodyanyuk[1], A. T. Bahcheli[1], J. To[2], Z. P. Klein[1], K. Guevara-Hoyer[1], C. McIntosh[3], M. Bhat[2], M. Bayati[1]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Univ. Hlth.Network, Toronto, ON, Canada, [3]Princess Margaret Cancer Ctr., Toronto, ON, Canada

Abstract:

Driver mutations and genomic copy-number alterations activate oncogenes and disrupt tumor suppressor genes and thereby provide hallmark cancer properties to cells. Hundreds of known driver genes have been identified in major cancer types. However, the tumor immune microenvironment is also major contributor to tumor progression, patient outcomes and therapy responses, as evidenced in the recent advent of immunooncology and novel therapeutic approaches. A myriad of cell types involved in inflammation and immune function are known to have complex tumor-promoting and anti-tumor roles that are increasingly characterised for mechanistic understanding and therapy development. Yet we have only a limited understanding how specific genomic driver alterations interact with the activities of immune cells and other cells of the tumor microenvironment, and how these interactions affect patient outcomes.

To study the synergistic and antagonistic interactions of the cancer genome and its microenvironment, we performed an integrative machine learning analysis of more 10,000 cancer samples to identify pairs of genomic alterations and immune cell infiltration profiles whose co-occurrence in matched tumor genomes and transcriptomes was highly predictive of patient survival. Our analysis spanned hundreds of positively selected driver mutations and recurrent copy-number alterations identified in cancer genomes combined with 22 major types of immune cells predicted from bulk patient transcriptomes. We present a catalogue of 43 immunogenomic interactions (IGX) in 15 cancer types in which the combination of a specific driver mutation and the profile of a particular immune cell type was informative of tumor progression or overall survival, and as reference, significantly exceeded the prognostic potential of the genomic alteration and immune cell profile alone. For example, in luminal-A breast cancer, reduced levels of neutrophils combined with chromosomal deletions of 11q13 were indicative of worse prognosis and tumor progression as observed in two independent cancer cohorts. At the transcriptomic level, this IGX associates with the modulation of proliferative and immune system signatures in luminal-A breast cancer. Further, IGXs often correlate with differential inflammation and immune responses as well as differential expression of

immunotherapy targets. This highlights previously uncharacterised subsets of high- or low-risk subsets of tumors that are jointly defined by their genetic and immune microenvironmental profiles, demonstrating avenues for risk stratification and therapy development.

Session Title: Cancer Poster Thursday Session

Board 8063T: The impact of cigarette smoke on the transcriptome across human tissues

Authors:

L. Copeland-Hardin[1], L. Tong[1], K. Demanelis[2], B. Pierce[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Cigarette smoke (CS) contains toxicants including carcinogens that adversely affect the health of many organs. Prior studies have not comprehensively examined CS-driven transcriptional responses across non-diseased tissues to understand mechanisms that mediate or defend against the pathogenesis of cancer and other diseases. We examined the association of CS with transcriptome-wide measures of RNA (1) expression, (2) splicing variation, and (3) editing events previously estimated using RNA-seq data from human tissue samples collected by the Genotype-Tissue Expression (GTEx) project. We examined 31 different tissues collected from 878 postmortem donors, with sample sizes ranging from n=183 to n=753 per tissue type. Overall, 67% of donors were cigarette smokers. Regression models adjusting for participant/sample characteristics and surrogate variables were used to identify CS-associated transcriptome features. CS was associated (FDR 0.05) with gene expression levels in 29 tissue types. The number of CS-associated genes was highest in lung (4,104 genes), thyroid (1,174), esophagus - mucosa (368), whole blood (337), pancreas (293), and heart - atrial appendage (293). Over 80 CS-associated genes were detected (FDR 0.05) in at least four tissue types, including *AHRR* (21 tissue types), *CYP1A1* (19), *CYP1B1* (17), and *GPR15* (12), which were upregulated by CS. The response of immunoglobulin gene clusters to CS was prominent in stomach, skin, whole blood, and heart-atrial appendage, with dozens of immunoglobulin (Ig) Kappa (chr2), Heavy chain (chr14), and Lambda genes (chr22) showing clear association with CS in these tissues. Gene set enrichment analyses of tissue types with >250 CS-associated genes using hallmark gene sets showed enrichment of genes in MYC targets and oxidative phosphorylation gene sets in most tissues analyzed. Other gene sets showing enrichment in multiple tissues included xenobiotic metabolism, inflammatory response, TNF alpha signaling, and P53 pathway. CS-associated splicing events were identified in six tissue types, including lung

(27 events), whole blood (10), and thyroid (7). The most prominent splicing associations were observed in pancreas (*CD36*, $p = 2x10^{-16}$), thyroid (*CD36*, $p = 4x10^{-8}$), and lung (*ITGAE*, $p = 9x10^{-7}$). No clear associations between CS and RNA-editing events were observed. This work provides a comprehensive description of the impact of CS for a wide variety of human tissues, many of which are prone to smoking-induced cancers. It provides a novel resource to further study how the transcriptome facilitates or protects against CS-driven diseases in human tissues.

Session Title: Cancer Poster Thursday Session

Board 8064T: Unraveling high-grade gliomas: Enhancing cytogenomic aberration assessment through subtelomeric MLPA analysis

Authors:

V. Rodriguez-Machuca[1], J. Ramírez-Corona[1], L. Bustos-Rodríguez[2], H. Velázquez-Santana[3], S. Núñez-Velasco[4], M. González-Rodríguez[1], J. Corona-Rivera[5], L. Bobadilla-Morales[6], A. Corona-Rivera[6]; [1]Centro Univ.rio de Ciencias de la Salud. Univ. de Guadalajara, Guadalajara, Mexico, [2]Departamento de Anatomía Patológica, Nuevo Hosp. Civil “Dr. Juan I. Menchaca”, Guadalajara, Mexico, [3]Servicio de Neurocirugía, Nuevo Hosp. Civil “Dr. Juan I. Menchaca”, Guadalajara, Mexico, [4]Servicio de Neurocirugía, Antiguo Hosp. Civil “Fray Antonio Alcalde”, Guadalajara, Mexico, [5]Univ. de Guadalajara, Centro Univ.rio de Ciencias de la Salud, Guadalajara, Mexico, [6]Univ Guadalajara, CUCS, Guadalajara, Mexico

Abstract:

High-grade gliomas (HGGs) (WHO grade 4) are the most aggressive primary brain tumors that confer a poor prognosis despite current treatment strategies. At the molecular level, HGGs are difficult to study due to the great heterogeneity of genomic alterations they harbor, however, some of these alterations, being highly recurrent, have been considered as diagnostic and prognostic biomarkers. Some known molecular biomarkers include cytogenomic aberrations such as whole or partial chromosome gains and/or losses (e.g. chr 7 gain, chr 10 loss, co-deletion of 1p and 19q, etc.). Since there is no gold standard for assessment of these aberrations, diverse methods have been used, some of which include MLPA assays. Particularly, subtelomeric MLPA methodology has proven to be a robust and sensitive form of segmental aneuploidy profiling in oncology. In this study, we performed subtelomeric MLPA for segmental aneuploidy profiling as an alternative evaluation of recurrent cytogenomic aberrations found in HGGs at the chromosome level. A total of 83 tumor tissue samples from patients that met histopathological and molecular criteria for HGG were included in the study. MLPA assays were performed using the SALSA MLPA Probemix P036 from MRC Holland®. SALSA MLPA Probemix

P070 was used as a confirmatory assay. Probe gains or losses were considered true when there was concordance between both assays. Complete chromosome gains were identified in 17 cases, the most frequent gains were found on chromosome 7 (14/83), in addition, complete chromosome losses were identified in 56 cases, the most frequent losses included chromosome 10 (13/83) followed by chromosomes 16 (8/83) and 13 (5/83). On the other hand, partial chromosome gains and/or losses were identified in all samples. Recurrent gains in at least 50% of the samples were seen in the short arm of chromosome 7 (32/83) and 1 (28/83), respectively. Frequent losses in at least 50% of the cohort were seen in the short arm of chromosomes 10 (35/83), 13 (21/83), 16 (22/83), and X/Y(*PAR1*) (26/83) and in the long arm of chromosomes 3 (26/83), 4 (26/83), 16 (23/83) and 22 (20/83). Interestingly, MLPA found no losses in the 1p or 19q regions, which are associated with oligodendroglioma (WHO grade 3), and not considered as HGGs. According to these results, subtelomeric MLPA proves to be a useful and reliable technique to detect recurrent cytogenomic aberrations (at the chromosomal level) in HHGs. Our findings agree with previously reported frequencies of cytogenomic aberrations found in these tumors demonstrating that this technique can even detect some biomarkers recognized for their clinical value in neuro-oncology.

Session Title: Cancer Poster Thursday Session

Board 8065T: Uptake of Risk-Reduction, Surveillance and Therapeutic Interventions among Breast Cancer Patients with Pathogenic Germline Variants

Authors:

E. Russell[1], S. M. Nielsen[1], F. Facio[1], R. Ellsworth[1], E. Esplin[1], W. M. Korn[1], N. M. Tung[2], A. W. Kurian[3]; [1]Invitae Corp., San Francisco, CA, [2]Beth Israel Deaconess Med. Ctr., Boston, MA, [3]Stanford Univ., Stanford, CA

Abstract:

Background: Management of breast cancer (BC) patients with pathogenic germline variants (PGVs) can include risk-reducing bilateral mastectomy (BLM), PARP inhibitors (PARPi), enhanced breast imaging (MRI), risk-reducing bilateral salpingo-oophorectomy (RRBSO) and pancreatic cancer surveillance (endoscopic ultrasound, EUS/magnetic resonance cholangiopancreatography, MRCP). We assessed intervention uptake in >1,400 PGV-positive BC patients.

Methods: Germline genetic testing (GGT) data (Invitae Corp.) and insurance claims data (Komodo Healthcare Map™) were assembled for female patients with BC or DCIS diagnosed

2015-23, GGT <120 days from diagnosis, and ≥1 year of claims pre/post-GGT. Inclusion/exclusion criteria followed prior work (PMID 32027353). Genes analyzed were those conferring high (>50%; *BRCA1*, *BRCA2*, *CDH1*, *PALB2*, *PTEN*, *STK11*, *TP53*) or moderate (20-50%; *ATM*, *BARD1*, *CHEK2*, *NF1*, *RAD51C*, *RAD51D*) BC risk. Multivariable logistic regression models of intervention use included PGV risk, race/ethnicity, age, family cancer history, local metastasis, type of insurance, geographic region and time from diagnosis to GGT. Chi-square tests were used to compare intervention uptake rates by PGV risk.
Results: 1,432 patients had ≥1 PGV in a high- (51%) or moderate-risk (49%) gene: 68% White, 32% lymph node involvement, 69% BC family history, 28% Medicare/Medicaid, mean (SD) age at testing: 51 (12) and mean (SD) days from diagnosis to GGT: 40 (26). The intervention with the highest uptake was BLM (60% in all PGV carriers). BLM, RRBSO, and PARPi were significantly more frequent in high- (76%, 31%, 7%, respectively) vs. moderate-risk (42%, 6%, 0%) PGV carriers (p<0.001 for all). In the multivariable model, patients with BLM uptake were younger (odds ratio, OR 0.9, 95% confidence interval, CI 0.9-1), more often had family cancer history reported (OR 2, 95% CI 1-3) and had high-risk PGV (OR 5, 95% CI 3-7). Patients with RRBSO uptake had higher odds of pancreatic (OR 2, 95% CI 1-2) or ovarian (OR 2, 95% CI 1-3) cancer family history and high-risk PGV (OR 6, 95% CI 4-9) and lower odds of Medicare (OR 0.4, 95% CI 0.2-0.8). Less than one-third of high and moderate PGV carriers underwent surveillance measures (30% overall for breast MRI and 6% for EUS/MRCP) recommended for certain PGV carriers.
Conclusion: Overall, the uptake of interventions was higher in BC patients with high- compared to moderate-risk PGV. More research, however, is needed on the shared decision-making process, as uptake of BLM was relatively high in those with moderate-risk PGV, and uptake of surveillance measures was relatively low.

Session Title: Cancer Poster Thursday Session

Board 8066T: WGS as the new standard for cancer sequencing using the Ultima UG100

Authors:

J. Hess[1], E. Martin[1], I. Soifer[2], H. Benjamin[2], C. Cibulskis[1], B. Danysh[1], M. Coole[1], S. Gabriel[1], D. Juric[3], D. Lipson[2], G. Almogy[2], G. Getz[1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Ultima Genomics, Newark, CA, [3]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Comprehensive characterization of tumor evolution is essential for both retrospectively discovering the drivers of metastasis and treatment resistance, and also prospectively guiding therapy in response to emergent resistance mechanisms. Reconstructing the phylogenetic evolution of each patient's tumor over time and space requires sequencing multiple samples to fully capture the patient's clonal heterogeneity.
Deep whole genome sequencing (WGS) enables the creation of more robust phylogenies relative to whole exome or panel sequencing, since each clone is characterized by far more clone-specific mutations, making it much easier to deconvolve individual subclones within a single bulk tumor sample. Indeed, in low mutation rate tumor types, phylogenetic reconstruction is often impossible using exomes. However, sequencing costs have thus far been prohibitively high to perform WGS across the requisite number of samples, inhibiting our ability to discover and clinically survey drivers of resistance.
Recent development of more affordable WGS by Ultima Genomics has changed this. We developed a method to postprocess Ultima sequencing alignments to be turnkey compatible with any existing downstream analysis, including well-established accurate somatic mutation detection pipelines. We rigorously benchmarked the Ultima-derived calls using a ground truth set of somatic mutations from well-studied cell line tumor/normal pairs sequenced on a variety of platforms, and found that performance both in terms of variant calling sensitivity/specificity and overall sequencing accuracy was equivalent to and often superior than other platforms.
We then identified 32 patients from various cancer types (breast, lung, cholangiocarcinoma, and melanoma) treated with a variety of therapies. We collected between 4 to 14 samples per patient at autopsy (total of 303) and sequenced them with Ultima to an average coverage of 63x, applying our postprocessing method. This allowed us to identify mutations using a well-established best practices pipeline, and construct phylogenetic trees using our PhylogicNDT suite of tools. We then compared the results to trees previously built from whole exomes sequenced on Illumina. We demonstrate that the trees generated from Ultima WGS data are far more detailed than their whole exome counterparts.
The dramatically lower cost of Ultima sequencing will not only more accurately elucidate novel mechanisms of evolution and resistance, but also advance the use of real-time phylogenetic reconstruction in the clinic, benefiting patients as their physicians will be able to much more precisely guide the course of their treatment.

Session Title: Cancer Poster Thursday Session

Board 8067T: Integrating Germline and Somatic Data to Resolve Variants of Uncertain Significance (VUS) in Colorectal Cancer (CRC) Patientswho Self-Identified as Hispanic/Latino ⭐

Authors:

J. Amzaleg[1], J. O. Culver[2], C. N. Ricker[2], N. Gutierrez[2], Y. Jin[1], B. Waldrup[1], C. Chavez[2], L. Enriquez[2], J. Sanchez-Mendez[2], D. Hernandez[2], B. Salhia[2], L. Baezconde-Garbanati[2], M. C. Stern[2], E. Velazquez Villarreal[1], J. D. Carpten[1], H-J. Lenz[2], D. W. Craig[1]; [1]City of Hope, Duarte, CA, [2]USC Norris Cancer Ctr., Los Angeles, CA

Abstract:

Germline and somatic genetic testing are pivotal in cancer management, distinguishing between pathogenic and benign variants. Typically, these tests are conducted independently, but integrating both genomic data types can enhance the resolution of Variants of Unknown Significance (VUS), especially important in non-White and non-European populations where VUSs are more common. This integration is vital to extending the benefits of precision medicine and reducing health disparities.Our study evaluated 88 Hispanic colorectal cancer (CRC) patients, finding that 44% (39/88) harbored at least one germline VUS - significantly higher than the typically reported rate of less than 15%. By integrating tumor transcriptome data, mutational signatures, and copy number variations with traditional germline tests, we clarified VUSs and identified driver mutations. For instance, our integrated approach confirmed the biological significance of an AXIN2 VUS due to mutant-allele specific expression in tumor RNA. We also determined that 26% (10/39) of participants had likely benign VUSs, based on their tumor mutational profiles and RNA expression.Furthermore, among patients with VUSs in HRD-associated genes like BRCA2 and PALB2, none exhibited HRD via copy number analyses. Conversely, of the six patients displaying HRD, only one had a pathogenic variant in an HRD-associated gene (ATM). Additionally, we identified tumors with a DNA mismatch repair-like phenotype—high tumor mutational burden, high significant microsatellite instability, without mutations in Lynch Syndrome-associated genes. Methylation analysis revealed MLH1 promoter hypermethylation in these cases, affirming our method's capability to predict DNA mismatch repair deficiencies even in the absence of genetic alterations in mismatch repair genes.This study underscores the importance of integrating germline and somatic data to better interpret VUSs in CRC, enhancing variant classification and providing insights potentially missed by separate analyses. This approach improves the clinical utility of genetic testing and ensures more equitable genetic counseling across diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5001T: A CRISPR/Cas9 screen identifies several genes and non-coding regulatory elements involved in the regulation of red blood cell density

Authors:

N. Brosseau[1], T. Pincez[2], K. Sin Lo[1], M. Beaudoin[1], G. Lettre[3]; [1]Montreal Heart Inst., Montréal, QC, Canada, [2]Université de Montréal, Montréal, QC, Canada, [3]Université de Montréal, Montreal, QC, Canada

Abstract:

Red blood cell (RBC) traits may influence the severity of hematological diseases. For example, RBC density is associated with sickle cell disease (SCD) symptom occurrence. Genome-wide association studies have identified hundreds of sentinel variants associated with RBC density, as captured by routinely measured mean corpuscular volume (MCV), mean hemoglobin concentration (MHC), and mean corpuscular hemoglobin concentration (MCHC). Because of linkage disequilibrium, the causal variants and genes at these GWAS loci remain largely unknown. In this study, we used an immortalized $CD34^+$ erythroid cell line, HUDEP-2, expressing different Cas9 modalities to perform a screen targeting the different variants identified by GWAS. HUDEP-2 cells expressing either Cas9 or dCas9-KRAB were transduced with a library of 17,219 gRNA, then separated according to their density using a Percoll gradient. Cells at each interface of the gradient were collected and the abundance of each of the gRNA was quantified by Illumina sequencing. We expect sgRNA that reduce the density of the cells to be less abundant in higher density fractions, while sgRNA increasing the density will be more abundant in the higher density fractions. Twenty-one genomic regions were identified with the Cas9 system and 22 were identified with the dCas9-KRAB with a certain overlap at a false discovery rate <0.1. Based on the results of the first round, we selected and tested 150 gRNA in a second validation screen (including 30 new gRNA designed to act as controls). These two rounds of screening identified 15 coding genes impacting HUDEP-2 cell density, including genes known to be involved in RBC density regulation such as ATP2B4, SLC4A1, and CCND3, but also other genes such as OXSR1, CTU2, and MIDN. Our results also identified 11 non-coding regions. More than half of the non-coding hits are close to genes known to be involved in RBC density regulation or to coding genes also identified as having an impact on HUDEP-2 cell density during the screen.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5002T: A hot spot loss of function variant in*PLAU*confers risk for thrombotic events in severe COVID19 patients

Authors:

J. Nevado[1,2,3], A. Parra[1,2], J. Tenorio[1,2,4,3], M. Cazalla[1], A. Ahijado[1], S. Ramos[1], N. Gallego[1,2], C. Silvan[1], E. De la Morena[5], C. L. Flores[6], C. Ayuso[7], E. Guillen[8], S. Diz[9], R. Cruz[9], B. Sobrino[9], C. Carpio[10], R. Alvarez-Sala[11], E. López-Granados[12], J. Rivera[5], G. Pérez-Solero[13], S. Ferreiro[13], J. Blom-Dahl[13], A. Teper[14], J. Di Paola[15], J. Corral[5], V. L. Ruiz-Pérez[6], The Scourge Consortium, J. A. Riancho[16], C. Flores[17], A. Rojas-Martínez[18], Á. Carracedo[9], P. Lapunzina[1,2,3]; [1]INGEMM-HULP, Madrid, Spain, [2]CIBERER, Madrid, Spain, [3]ITHACA Res. Network, Madrid, Spain, [4]ADNTRO Genetics and BITGENETIC, Madrid, Spain, [5]IMIB, Murcia, Spain, [6]Inst. de Investigaciones Biomédicas, Madrid, Spain, [7]Fundación Jiménez Díaz, Madrid, Spain, [8]Hosp. Clínico Univ.rio Virgen de la Arrixaca, Murcia, Spain, [9]Univ. de Santiago de Compostela, Santiago de Compostela, Galicia, Spain, [10]Sº Neumología-HULP, Madrid, Spain, [11]Sº de Neumologia-HULP, Madrid, Spain, [12]Sº Inmunología-HULP, Madrid, Spain, [13]ADNTRO Genetics, Madrid, Spain, [14]Hosp. de Niños Ricardo Gutiérrez, Buenos Aires, CABA, Argentina, [15]Univ. of Washington in St Louis, St Louis, MO, [16]Univ. de Cantabria, Santander, Cantabria, Spain, [17]Hosp. Univ.rio N.S. de Candelaria, Tenerife, Canary Islands, Spain, [18]Tecnológico de Monterrey, Monterrey, Mexico

Abstract:

BACKGROUND COVID-19-associated coagulopathy (CAC) contributes to the morbidity and mortality of patients infected with SARS-CoV-2. *PLAU* gene encodes the secreted serine protease uPA (urinary plasminogen activator) that converts plasminogen to plasmin. Gain of function variants due to large duplication of this gene is associated with the Quebec platelet bleeding disorder. Knock-out mice of *PLAU* displaying loss of function variants develop thrombotic events including pulmonary thromboembolism. **METHODS** We performed whole genome sequencing and targeted resequencing in 300 patients who presented with severe COVID19 disease from the Spanish SCOURGE Consortium. We then replicated the findings in more than 1000 patients from different cohorts of patients with COVID19. We investigated whether loss of function variants in *PLAU* explain the deep venous thrombosis and pulmonary thromboembolism observed in some patients with severe COVID19. **RESULTS** We found four non-related patients with the same splicing variant in *PLAU:*NM_002658.3:c.970+1G>A from the SCOURGE Consortium and none patients from the other international cohorts. All patients suffered from a pulmonary thromboembolism in the context of their COVID19 infections. In these patients, the mutation disrupted the splice site of exon 9 of the gene leading to a diminished RNA expression and secretion of the protein and consequently a potential loss of

function of the protein. We also performed functional assays by RNAseq, minigenes, evaluated the platelet function and pharmacogenetics to deeply investigate the effect of this variant. **CONCLUSIONS** These are the first patients with loss of function variants in *PLAU* associated with a human disease. Like in mice, loss of PLAU function leads to increased peripheral thrombosis and pulmonary thromboembolism probably triggered by the COVID19 infection.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5003T: A multi-ancestral genetic association study of ischemic heart failure

Authors:

C. Liu[1], Q. Hui[1,2], A. K. Yadalam[3], J. Gaziano[4,5], P. W. F. Wilson[1,3], the Million Veteran Program, L. S. Phillips[1,6], J. Joseph[7,8], Y. V. Sun[1,2]; [1]Dept. of Epidemiology, Emory Univ. Rollins Sch. of Publ. Hlth., Atlanta, GA, [2]Atlanta VA Hlth.care System, Decatur, GA, [3]Div. of Cardiology, Dept. of Med., Emory Univ. Sch. of Med., Atlanta, GA, [4]Massachusetts Veterans Epidemiology Res. and Information Ctr. (MAVERIC), VA Boston Hlth.care System, Boston, MA, [5]Div. of Aging, Dept. of Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [6]Div. of Endocrinology, Dept. of Med., Emory Univ. Sch. of Med., Atlanta, GA, [7]VA Providence Hlth.care System, Providence, RI, [8]The Warren Alpert Med. Sch. of Brown Univ., Providence, RI

Abstract:

Heart failure (HF) has been a growing global health burden, with ischemic HF being a prevalent subtype associated with coronary artery disease (CAD). While CAD is a known precursor to HF in many cases, the genetic susceptibility to HF among patients with CAD remains poorly understood. Genome-wide association studies (GWAS) offer a powerful tool to elucidate the genetic architecture underlying complex diseases like ischemic HF. In this study, we conducted multi-ancestral GWAS meta-analyses to identify genetic variants associated with the development of all HF, HF with reduced ejection fraction (HFrEF, left ventricular ejection fraction [LVEF] ≤ 40%) and HF with preserved ejection fraction (HFpEF, LVEF ≥ 50%) in individuals with underlying CAD. Ancestry-specific GWAS of ischemic HF were conducted among participants from the Million Veteran Program, which has recruited over one million ancestrally diverse veterans. By leveraging data linkage to electronic health records, our study included 48,111 European (24,533 HFrEF; 18,625 HFpEF), 10,549 African (5,683 HFrEF; 3,781 HFpEF), and 3,403 Hispanic American (1,759 HFrEF; 1,301 HFpEF) ischemic HF cases, defined as those with CAD prior to or within 6 months of HF diagnosis, alongside 87,857 European, 15,177 African, and

6,810 Hispanic American patients with CAD but without HF. After data quality control, we performed GWAS using logistic regression models, with adjustment of age, sex, and top 10 principal components within each ancestry group, followed by multi-ancestry meta-analyses. We identified 4 genome-wide significant (GWS) loci for all ischemic HF: *CDKN1A*, *PGAP3*, *BPTF* and *ZBTB17*, which play important roles in various cellular processes, including cell cycle regulation, protein modification, immune response, and gene expression control. For ischemic HFrEF, we identified 12 GWS loci relevant to cell cycle regulation, ion transport, and muscle contraction, including *CDKN1A*, *CLCNKA*, *BAG3*, *FLNC*, *PGAP3*, *MAP3K7CL*, *TTN*, *SMARCB1*, *MLIP* and *E2F6*. No GWS loci were found for ischemic HFpEF. Using Genotype-Tissue Expression (GTEx) data, gene expression analysis indicated that the loci associated with all ischemic HF are enriched in muscle tissues, whereas those linked with ischemic HFrEF showed enrichment in prostate, muscle, and heart tissues. Our genetic association findings contribute to a deeper understanding of ischemic HF pathophysiology and may inform the development of personalized HF therapeutic strategies tailored to individual genetic profiles.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5004T: A multiancestry genome-wide association study of 108,477 individuals with gallstone disease identifies 37 novel genetic loci

Authors:

J. Lim, M. Vujkovic, M. G. Levin, D. Zhang, S. Abramowitz, S. Mukherjee, S. Damrauer, D. Rader; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: Gallstone disease is highly prevalent and a cause of significant morbidity in the United States, leading to one of the largest healthcare costs of gastrointestinal diseases. Our current understanding of the genetic architecture of gallstone disease is primarily based on large-scale genetic studies with a focus on populations of European ancestry. We aimed to expand our understanding of the susceptibility loci and the ancestry-specific genetic signals of gallstone disease. **Methods**: We conducted a large-scale genome-wide association study meta-analysis including 102,940 cases and 1,647,377 controls within 4 cohorts: UK Biobank (UKBB), FinnGen, Biobank Japan (BBJ), and Million Veteran Program (MVP). For MVP data, we used summary statistics available from the MVP genome-wide phenome-wide association study (gwPheWAS). For all cohorts, gallstone disease case was defined as Phecode 574.1. Ancestry-

specific analyses were performed for individuals of European, East Asian, Admixed American, and African ancestries using PLINK software. Subsequently, population-specific and multi-population meta-analyses were conducted using the inverse weighted, fixed effects model using METAL. The variants were designated genome-wide significant if P-value < 5×10^{-8}. We performed a replication study in the All of Us Research Hub consisting of 5537 cases and 166,333 controls. **Results**: A total of 91 (37 novel) loci were identified across all analyses, in which 85 (33 novel) loci were discovered in multiancestry meta-analysis, 68 (23 novel) loci among those of European ancestry, 15 (0 novel) loci among those of East Asian ancestry, and 4 (2 novel) loci among those of African ancestry. In the replication cohort (All of Us Research Hub), we found directional concordance among 67/84 loci (80%) within multi-ancestry, 54/68 loci (79%) within European ancestry, 11/15 loci (73%) within East Asian ancestry, and 3/4 loci (75%) within African ancestry. **Conclusions:** Through this large-scale multi-ancestry GWAS meta-analysis, we leveraged the genetic diversity across ancestry groups to uncover novel genetic variants of gallstone disease. The investigation of genetic basis of bile lithogenicity and other factors that affect gallstone development (e.g. gallbladder dysmotility) has the potential to advance pharmacogenomic approaches to medication development for primary and secondary prevention of gallstones and advance our understanding of regulatory mechanisms of cholesterol homeostasis that manifest in other gastrointestinal and cardiovascular phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5005T: A multi-ancestry, Genome-Wide Association Study (GWAS) of sepsis risk identifies a risk locus in TRIM22

Authors:

A. Biji, R. Shemirani, E. Kenny, S. Asgari; Inst. of Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Background: Sepsis occurs when infections elicit an extreme inflammatory reaction from the immune system. Its severity makes it the most expensive inpatient medical condition in the US and the second most common cause of in-hospital deaths. Hence timely risk assessment for potentially vulnerable patients is necessary as an intervention strategy.
Methods: We used BioMe, and the All of Us Biobank (AoU) to run the largest GWAS of sepsis to date. To obtain sepsis cases from patient health records in the BioMe biobank (53,913 participants, 65.4% non-European), we used PhecodeX to aggregate ICD9 and ICD10 billing codes. We used SAIGE for single variant association testing in 1,510 cases and 50,746 controls

separated by genetic ancestry groups (European, African, and Hispanic) and accounting for age, age2, sex, age.sex, age2.sex, genotyping chip, and 20 genotyping principal components. The results were meta-analyzed (Inverse-variance weighted) with available summary statistics from AoU European, African, and Hispanic subgroups (12,608 cases defined using PhecodeX and 204,837 controls). Results were filtered to include only variants with minor allele frequency (MAF) >0.1%.

Results: Test statistics from meta-analysis revealed no inflation (lambda GC = 1.01). A substitution on chromosome 11 (**rs149994054**) was significantly associated (beta=0.821, se=0.147, p=2.321e-08) with sepsis case-control status. rs149994054 is a low-frequency variant with an allele frequency of 0.8% in BioMe Europeans and 0.6% in All of Us Europeans (<0.5% in other populations). The variant's direction of effect was concordant in all populations; however, the signal was driven primarily by Europeans in both datasets. Indeed, a European-only meta-analysis improved the association (beta=0.908, se=0.164, p=3.162e-08). We are currently repeating these findings in the UK Biobank. rs149994054 lies in the intronic region of TRIM22, an interferon-inducible gene that has been primarily studied in the context of viral infections. However, variants in TRIM22 are also associated with other immune-related traits such as lymphocyte count and platelet volume. Further investigation is needed to establish the causal relationship between this variant and sepsis risk.

Conclusion: To our knowledge, this is the largest GWAS of sepsis. While only one locus reached genome-wide significance, several loci showed suggestive significance (p<1e-5) and may reach statistical significance with increased sample sizes, on which we are working.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5006T: A Novel Hypervariable Number Tandem Repeat in the Dopamine Transporter Gene is Associated with Multiple Dopamine-Related Phenotypes

Authors:

A. Apsley, D. Vandenbergh, Y. Veturi; The Pennsylvania State Univ., University Park, PA

Abstract:

Multiple genotype-phenotype analyses of the dopamine transporter gene (*SLC6A3/DAT1*) have generated conflicting results for dopamine-related phenotypes. In this pilot study, long-read genotypes of a hyper Variable Number of Tandem Repeats (hVNTR) in *SLC6A3* were used to test for association with dopamine-related phenotypes. Using a sample of 1,001 individuals (70.8% female, age = 53.0 ± 14.7 [23-97], and 92.6% African American) from the All of Us dataset with

long-read whole genome sequencing data, we extracted the hVNTR copy number alleles present for each individual. Using PacBio HiFi reads with an average coverage of 8x, 242 individuals had both hVNTR alleles, 505 individuals had one allele, and 254 individuals had no alleles present. Within the 242 individuals with both hVNTR alleles present, the average combined allele length was 41.9 ± 33.4 copies, with the longer alleles having 48.2 ± 24.5 copies and the shorter of the alleles having 16.9 ± 14.3 copies. Imputation of missing alleles for the 505 individuals with one allele present was performed using short-read count data across the region to estimate total hVNTR length (r=0.75, p<0.001). Two cohorts were constructed for analyses, one with short-read-imputed alleles (N=747) and one with only long-read-resolved alleles (N=242). Dopamine-related psychiatric conditions (substance use disorders, major depressive disorder and attention-deficit/hyperactivity disorder), and presence/absence of cocaine and amphetamines in blood, urine or saliva of study participants were used as phenotypes. In the larger cohort, combined hVNTR allele length was positively associated with depression diagnosis (p<0.05) with average hVNTR combined length in cases being 46.5 ± 37.9 and controls being 40.9 ± 32.2 copies. In the smaller cohort, alcohol use disorder (AUD) and presence of cocaine were negatively associated with longer hVNTR allele length (p<0.05). AUD cases had 34.5 ± 25.9 copies (cocaine cases had 37.9 ± 27.2 copies) and AUD controls had 43.6 ± 27.2 copies (cocaine controls had 43.7 ± 27.1 copies) of the longer allele. Although not significant, trends in AUD and presence of cocaine were in the same direction in the larger cohort. Our results suggest the importance of assessing the genotype of a novel hVNTR in *SLC6A3* for understanding the etiology of dopamine-related traits. We propose to validate our analyses in an independent cohort from Sun Yat-sen university (N= 405) that integrates long-read sequencing with clinical phenotypes. In addition to larger sample sizes, greater long-read sequencing depth (>8x) to avoid missing alleles is recommended for future long-read sequencing data releases in All of Us.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5007T: A VNTR containing multiple MEF2 binding sites is associated with the African origin haplotype protective for AD in *APOEε4* carriers

Authors:

L. Bertholim Nasciben[1], K. Nuytemans[1], F. Rajabli[1], D. M. Dykxhoorn[1], L. Wang[1], W. K. Scott[1], D. A. Davis[2], R. T. Vontell[2], K. McInerney[3], M. Cuccaro[1], G. Byrd[4], J. Haines[5], L. Adams[1], M. A. Pericak-Vance[1], Alzheimer's Disease Sequencing Project (ADSP), J. I. Young[1], A. J. Griswold[1], J. M. Vance[1]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami,

FL, [2]Brain Endowment Bank, Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL, [4]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC, [5]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH

Abstract:

*APOE*ε4 is the strongest genetic risk factor for Alzheimer's disease (AD). The effect size of *APOE*ε4 varies across genetic ancestries with African (AFR) local ancestry conferring a lower risk when compared to other ancestries. Recently, we identified an interaction of the A allele of rs10423769 with *APOE*ε4, which reduces the AD risk effect of *APOE*ε4/ε4 by an estimated 75%. The frequency of the protective allele is 12% in AFR populations but only 0.03% in European. rs10423769 is 2MB from *APOE* in a region of segmental duplications (SD), a cluster of pregnancy specific beta-1 glycoprotein genes (*PSGs*) and a lncRNA. To gain insights into the protective mechanism, we characterized the haplotype from AFR origin and investigated potential structural variation (SV) associated with the protective haplotype. **Method:** We characterized the haplotype using short read sequencing from the AD Sequencing Project (ADSP) and Haploview. Long read whole genome sequencing (LRS) with the Oxford Nanopore PromethION was performed on 18 samples of rs10423769_A carriers and 21 non-carriers. Two rs10423769_A carriers were also sequenced with PacBio Revio. SV calling was performed with Sniffles2 and TREAT was used for local assembly of Variable number tandem repeat (VNTR). **Result:** We extracted 1,962 rs10423769_A carriers from the ADSP database and determined a minimum shared haplotype of 21Kb (19:43099521-43120243) harboring the protective allele. Despite being in a SD area, LRS reads mapped uniquely to the region around the A allele, confirming that the haplotype is unique and not duplicated elsewhere in the genome. SV calling revealed a 140bp±78 insertion in a tandem repeat region 32Kb from the protective locus, with a frequency of 0.80 in homozygotes for the protective allele and 0.07 in non-carriers. The local assembly of the tandem repeat region revealed that the protective allele is associated with expanded VNTR alleles enriched in multiple MEF2D binding motifs. Initial studies demonstrate that the VNTR associated with the A protective allele is significantly larger than those on the G allele haplotype. **Conclusion:** We identified an AFR-specific haplotype protective against AD risk conferred by *APOE*ε4. This haplotype is associated with an expanded VNTR that contains MEF2 binding sites. MEF2 transcription factors are involved in neuronal development, synaptic plasticity, cognitive reserve and neurodegenerative diseases. Ongoing analyses include functional essays aiming to investigate the role of the protective locus and the VNTR region in *APOE*ε4 expression and MEF2 availability. LRS at higher coverage with the PacBio is being performed to assemble and resolve the complex SD region.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5008T: *ABCA7* deletion and *APOE4* allele lower the age-of-onset of Alzheimer Disease in African-Ancestry individuals

Authors:

B. Akgun[1], A. V. Gulyayev[1], K. L. Hamilton-Nelson[1], L. D. Adams[1], P. G. Whitehead[1], G. W. Beecham[1,2], B. W. Kunkle[1,2], A. J. Griswold[1,2], M. L. Cuccaro[1,2], R. O. Akinyemi[3], A. Ogunniyi[3], A. Caban-Holt[4], A. F. Zaman[1], P. R. Mena[1], I. M. Martinez[1], S. H. Blanton[1,2], K. F. McInerney[5], G. S. Byrd[4], J. M. Vance[1,2,5], J. L. Haines[6], African Dementia Consortium (AfDC), Alzheimer's Disease Sequencing Project (ADSP), F. Rajabli[1,2], M. A. Pericak-Vance[1,2,5]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Coll. of Med., Univ. of Ibadan, Ibadan, Nigeria, [4]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Sch. of Med., Winston-Salem, NC, [5]Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Inst. for Computational Biology, Dept. of Population & Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH

Abstract:

Background: The *ABCA7* and *APOE* genes are the strongest heritable risk factors for Alzheimer disease (AD) in African-ancestry (AA) populations. Both genes are involved in lipid metabolism processes associated with AD pathology. However, the impact of these genes on AD progression among AA remains unclear. This study assesses the independent and interactive effect of the African ancestry-specific *ABCA7* frameshift deletion and *APOE4* allele on the age at onset of AD.
Methods: Our dataset includes Whole Genome Sequencing (WGS) and phenotype data from 6352 individuals (2303 AD; 4049 cognitively unimpaired) of self-reported African ancestry (African American, and Black Hispanic) from three studies (ADSP, REAAADI, READD-ADSP). We extracted *APOE* genotypes (rs7412, rs429358) and *ABCA7* deletion (rs142076058) from the WGS. We used a Cox proportional hazards regression model (accounted for relatedness) to determine the association between age-of-onset and 1) *ABCA7* deletion, 2) *APOE* genotypes, and 3) *ABCA7-APOE* interaction. We also performed the same model by stratifying the cohort by sex.
Results: The survival time of the *ABCA7* deletion carriers (median=85 years) was significantly lower compared to non-carriers (median=90 years) ($P=1.4x10^{-6}$). Also, the survival time of individuals with *APOE* e3/e4 (median=83 years) was significantly lower than those with e3/e3 (median=90 years) and higher than those with e4/e4 (median=73 years). Interaction analysis showed that among individuals with *APOE* e3/e4 alleles, age at onset was lower in deletion carriers (median=77 years) compared to non-carriers (median=83 years) ($P=8x10^{-9}$). However,

among individuals with *APOE* e3/e3 or e4/e4, no significant differences in survival time were observed between deletion carriers and non-carriers. We didn't observe a notable change in the age-of-onset in sex-stratified analyses.
Conclusions: Our study showed that both *ABCA7* deletion and the *APOE* e4 allele independently lower the age-of-onset in AA populations. We also observed that *ABCA7* deletion lowers the age at onset among *APOE* e3/e4 carriers. These results suggest that this AA-specific *ABCA7* deletion and the *APOE* e4 allele may have synergistic effects on age-of-onset of AD. Elucidating the impact of *APOE-ABCA7* interactions on AD is critical for investigating the complex pathogenesis of AD and developing effective treatments in all groups. We will extend our study to include individuals of African ancestry from Africa, further investigating the *APOE* and *ABCA7* genes in these populations.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5009T: Absence of SCA27B and unique allelic distribution in Peruvian patients with undiagnosed cerebellar ataxia

Authors:

A. Saldarriaga-Mayo[1,2], I. Araujo-Aliaga[1,2], S. Echavarria-Correa[1,2], M-J. Dicaire[3], M. C. Danzi[4], S. Zuchner[4], D. Pellerin[4,5,3], B. C. Brais[3,6], M. Cornejo-Olivas[1,2]; [1]Neurogenetics Working Group, Univ. Científica del Sur, Lima, Peru, [2]Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurológicas, Lima, Peru, [3]Dept. of Neurology and Neurosurgery, Montreal Neurological Hosp. and Inst., McGill Univ., Montreal, QC, Canada, [4]Dr. John T. Macdonald Fndn. Dept. of Human Genetics and John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [5]Dept. of Neuromuscular Diseases, UCL Queen Square Inst. of Neurology and The Natl. Hosp. for Neurology and Neurosurgery, Univ. Coll. London, London, United Kingdom, [6]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada

Abstract:

Background and objectives: Spinocerebellar ataxia 27B (SCA27B) is a newly discovered form of hereditary ataxia characterized by the expansion of an intronic GAA repeat in the *FGF14* gene. The disease is characterized by a late-onset, slowly progressive cerebellar syndrome with frequent episodic symptoms. Although SCA27B has been described in different populations, its prevalence and genetic characteristics in the Peruvian population are poorly understood. The objective of this study is to estimate the frequency of SCA27B in a Peruvian cohort of patients with hereditary ataxias without a definitive diagnosis treated at the Neurogenetics Research Center (CIBN).

Methods: DNA samples were obtained from 166 Peruvian individuals with hereditary ataxia without definitive diagnosis, with informed consent. Genotyping of the GAA repeat in the *FGF14* gene was performed using a workflow including long-range PCR, bidirectional RP-PCR, and Sanger sequencing.
Results: No cases with an expanded GAA repeat above 250 were found. Following genotyping, we found no patients carrying a pathogenic expansion of at least 250 GAA-pure repeats. 27 cases (16%) were found to carry a complex [$(GAA)_n(GCA)_m$] expansion ranging in size from 310 to 332 triplet equivalents, which is thought to be non-pathogenic for SCA27B. Normal alleles were found in the range of 7 to 127 repeats, the most common being 8 repeats (84/332; 25.3%).
Discussion and conclusion: The frequency of SCA27B caused by *FGF14* GAA repeat expansion in our Peruvian cohort is zero, although a large proportion of patients carried a complex non-GAA-pure expansion. The allelic distribution of the GAA repeat among our cohort of undiagnosed Peruvian patients significantly diverges from the European distribution, with no expanded GAA alleles found, unlike other populations where the expansion frequency ranges from 9% to 19%. The absence of SCA27B cases in the Peruvian cohort might be due to the Amerindian ancestry of the patients.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5010T: Accelerating novel gene discovery in nonsyndromic cleft lip/palate using whole genome sequencing

Authors:

A. Letra[1], Y. Yu[2], C. Huff[3], J. Hecht[4]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]UTMDACC, Houston, TX, [3]U.T. MD Anderson Cancer Ctr, Pearland, TX, [4]Univ Texas McGovern Med Sch, Houston, TX

Abstract:

Nonsyndromic with or without cleft lip/palate (NSCLP) is a complex trait with approximately 40 genetic risk loci in multiple signaling pathways identified as likely etiologic. However, variation in these genes/loci only account for ~25% of the estimated heritability of NSCLP, indicating that additional genes remain to be identified. To uncover the missing NSCLP heritability, we leveraged whole genome sequencing (WGS) data from 172 multigenerational NSCLP families of non-Hispanic white (NHW) and Hispanic ethnicities, along with genomic data of 616 additional NSCLP trios from the Gabriella Miller Kids First Orofacial Cleft Projects to identify rare and common variants contributing to NSCLP. Controls for this study comprised 2,131 parents of offspring with phenotypes unrelated to orofacial clefting in other Kids First projects, including

congenital diaphragmatic hernia, congenital heart defects, neuroblastoma, and Ewing sarcoma. We conducted joint variant calling using DeepVariant and GLnexus, followed by sample- and variant-level quality control (QC) and cross-platform QC using XPAT2. Pedigree validation was performed using PRIMUS. We then conducted a whole-exome gene-based association study on unrelated case and control samples using the Covariate Adjusted PERmutation (CAPER) analysis package. All rare protein-coding variants passing QC and with minor allele frequency (MAF) less than 0.005 among non-Finnish European and Hispanic populations in gnomAD were evaluated. We analyzed NHW and Hispanic populations individually and conducted a meta-analysis across the two populations. Type I error was well controlled in the population specific analyses ($\lambda_{NHW.=}$ 0.99 and $\lambda_{Hispanic}$ = 1) as well as the meta-analysis (λ_{meta} = 1.07). We replicated previously reported NSCLP genes, including *FGF8* (p_{meta} = 0.00046 and p_{NHW} = 0.0003) and *ARHGAP29* (p_{meta} = 0.0015 and p_{NHW} = 0.0003). Additionally, we identified several novel candidate genes, including three genes on chromosomes 3q13, 7q34, and 10p12 with suggestive significance level at p less than 1 x 10^{-4}. Our results represent a first step in our systematic approach to evaluate the contributions of both common and rare variants to NSCLP, detect segregation of putative causal variants within and across families, and test complex inheritance models that are critical for unraveling the genetic etiology of NSCLP.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5011T: Admixture mapping revels novel population-specific loci associated with kidney disease in African American individuals

Authors:

A. Horimoto[1], N. Franceschini[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Chronic kidney disease (CKD) affects 13.9% of U.S. adults, with a disproportionate burden on women and African American individuals. Several studies have shown the impact of population-driven genetic variation on the CKD risk. Admixture mapping leverages local ancestry to understand its influences on the disease risk and is an approach used to elucidate the genetic architecture of complex diseases in admixed populations. Here, we used data from 8,192 African American women from the Women's Health Initiative (WHI) to investigate population-specific genetic determinants of CKD and CKD-related traits of estimated glomerular filtration rate (eGFR), and end-stage kidney disease (ESKD). Local ancestry was called using RFMix to identify chromosome segments similar to population reference panels. The genome-wide

admixture mapping was performed using a mixed model adjusted for random (kinship coefficients) and fixed (age, recruitment region, and principal components) effects to test for association between the local ancestry at each locus and the outcomes. We identified seven significant loci on chromosome 9 for eGFR ($P < 7x10^{-6}$), in which each copy of the African allele conferred a risk effect for low eGFR (beta coefficient = -0.09, SE = 0.02, $P = 2.1x10^{-6}$). The most significant locus at 9q31.1 spans the genes *BAAT* and *ALDOB*, and comprises of 22 variants, including a missense variant and four variants with high deleteriousness score. Previous studies have implicated these protein-coding genes with CKD-related traits. *BAAT* has been previously associated with plasma succinoyltaurine levels in CKD and C-reactive protein levels, while *ALDOB* was associated with cystatin C levels and eGFR. We also identified suggestive signals for CKD at 22q13.31 ($P = 2.6x10^{-4}$), an intergenic locus including regulatory region variants, and for ESKD at 7q31.1 ($P = 1.9x10^{-4}$), an intronic region enriched for long non-coding RNA and enhancers. Our findings provide new information into population-specific genetic contributions to kidney disease risk in African American individuals and contributes to fill the diversity gap in the genetic studies.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5012T: Alcohol use disorder polygenic scores with effect sizes comparable to the family history and *ADH1B* in European ancestry populations ⭐

Authors:

D. Lai, M. Zhang, M. Abreu, T-H. Schwantes-An, Y. Liu, T. Foroud; Indiana Univ. Sch. of Med., Indianapolis, IN

Abstract:

Alcohol use disorder (AUD) is a chronic and relapsing disease with devastating effects on individuals, families, and society. Identifying high-risk individuals then applying intervention and prevention programs can significantly reduce the incidence of AUD. Previously, we used SNPs having the same directions of effects in different study cohorts (i.e., cohort-concordant SNPs) to calculate AUD polygenic scores (PGS) and substantially improved predictability and generalizability of AUD PGS. The PGS were also associated with AUD severity and remission from AUD. In this study, we used cohort-concordant SNPs to further increase PGS predictability and generalizability by including more discovery datasets and using target datasets ascertained for different purposes. The discovery datasets were genome-wide association studies of AUD related phenotypes from the Million Veteran Program (N=202,004), the UK Biobank

(N=121,604), and the FinnGen consortium (N=376,737). We used a series of P value and linkage disequilibrium (LD) r^2 thresholds to select independent SNPs for calculating PGS. Three cohorts ascertained to study AUD (COGA, OZALC, and SAGE, total N_{case}=6,323, total $N_{control}$=2,476) were used to screen for the best PGS. Then we used the All of Us research program (AOU, N_{case}=5,660, $N_{control}$=110,404), which consists of participants with diverse backgrounds and conditions, and the Indiana Biobank (IB, N_{case}=936, $N_{control}$=5,437), which consists of patients from Indiana University health care systems, to validate the PGS. All discovery and target datasets were European ancestry samples. The best PGS were calculated by using 9,154 SNPs having P value<0.01 in all three discovery cohorts and LD r^2<0.1. Compared to the rest of samples, those 5% having the highest PGS had odds ratio (OR)=1.96 (95%confidence interval (CI) 1.78-2.16, P=4.10E-43) and 2.07 (95%CI 1.59-2.71, P=9.15E-08) to develop AUD in AOU and IB, respectively. This is comparable to the family history of AUD (ORs=1.91-2.38) that is clinically used to screen high-risk individuals. Compared the rest of samples, those 5% having the lowest PGS have OR=0.53 (95%CI 0.45-0.62, P=6.98E-15) and 0.57 (95%CI 0.39-0.84, P=4.88E-03) to develop AUD in AOU and IB, respectively, which is comparable to the of protection effect of *ADH1B* (ORs=0.41-0.57). We also found that PGS had similar effects between males and females. Our findings demonstrated that AUD PGS can potentially be used to predict AUD risk and thus personalized prevention and treatment strategies could be developed.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5013T: ALS Identified: Enhancing ALS Genetic Diagnosis Through a Sponsored Genetic Testing Program

Authors:

C. Paisan-Ruiz[1], S. A. Goutman[2], N. A. Goyal[3], K. Payne[4], M. Kang[1], A. Mitchell[1], T. E. Fecteau[1]; [1]Biogen, Cambridge, MA, [2]Univ. of Michigan, Ann Arbor, MI, [3]Univ. of California, Irvine, CA, [4]Indiana Univ. Hlth., Indianapolis, IN

Abstract:

Background: Amyotrophic lateral sclerosis (ALS) is a rare, fatal neuromuscular disease characterized by signs and symptoms of upper and lower motor neuron degeneration. This degeneration results in progressive weakness of the bulbar, limb, thoracic, and abdominal muscles, ultimately leading to death in approximately 2 to 5 years after diagnosis, typically due to respiratory failure. Approximately 60%-70% of familial ALS (fALS) and 10% of sporadic ALS (sALS) cases are caused by mutations in six genes, including C9orf72, SOD1, OPTN, TARDBP, FUS, and VCP.[1-3]**Methods:** ALS Identified, a genetic testing program, launched on June 30, 2021, in

the U.S., uses the Invitae Amyotrophic Lateral Sclerosis panel to test 22 ALS-associated genes at no cost for individuals ≥18 years in the U.S. and Puerto Rico with an ALS diagnosis or with a known family history of ALS. Nine additional genes that currently lack a definitive clinical association with ALS but may prove to be clinically significant in the future, are available for add-on screening. Free post-test genetic counseling through Invitae is available to participants upon request. The ordering physician has the option to provide patient demographics and clinical characteristics.**Results:** In the first two years, 998 healthcare professionals from 681 care sites ordered the panel. By October 26, 2023, 8,054 individuals were examined, with 7,483 (92.9%) diagnosed with ALS, with a mean age of 62 (SD ±13) years. Of these, 57.7% were male, and 42.3% were female. Common symptoms included progressive muscle weakness (76.5%) and muscle fasciculations (61.3%). Smaller groups reported psychiatric symptoms (2.7%), behavioral changes (5.5%), and cognitive decline (8.4%). Among those diagnosed with ALS, 1,810 (24.2%) showed genetic variations in ALS-associated genes, with 865 individuals (47.8%) carrying pathogenic (P) variants, and 44 (2.4%) likely pathogenic (LP) variants. The most frequently mutated genes in our ALS population were C9orf72 and SOD1.**Conclusion:** The ALS Identified program provides standardized genetic testing across the U.S., creating one of the largest genetic ALS patient datasets. This initiative enhances the understanding of the genetic landscape of ALS, underscores the importance of genetic testing and counseling for molecular diagnoses, and may aid in developing targeted therapies by identifying key genes in the ALS population. **References**:1. PMID: 21914052; 2. PMID: 35334234; 3. PMID: 37620070

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5014T: Altered bile acid metabolism in Down syndrome ⭐

Authors:

L. Dunn[1,2], K. Schade[1,2], C. Brown[1,2], N. Eduthan[1,2], K. Waugh[1,2], P. Araya[1,2], M. Galbraith[1,2], J. Espinosa[1,2], K. Sullivan[2,1]; [1]Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, [2]Linda Crnic Inst. for Down Syndrome, Aurora, CO

Abstract:

Trisomy of chromosome 21 (T21) gives rise to Down syndrome (DS) and results in a unique disease spectrum. Individuals with DS have increased rates of numerous conditions including autoimmune disorders, heart defects, and Alzheimer's Disease, while being protected from other conditions such as many solid tumors. Although modest increases in biliary atresia and autoimmune hepatitis have been reported in DS, these conditions remain extremely rare. Liver function in DS has been almost exclusively studied through the lens of transient abnormal

myelopoiesis, a perturbation in fetal hematopoiesis, which takes place in the liver. Several recent reports have identified widespread liver dysfunction via imaging-based methods. Few studies, however, have examined the impact of T21 on liver function throughout development and into adulthood and none have aimed to elucidate the mechanisms by which this phenotype occurs. Here we report the results from proteomic and metabolomic analyses of human plasma from around 400 individuals from the Human Trisome Project Biobank. Using the Dp16 mouse model of DS, we also report histopathology and transcriptome analysis of the liver, along with plasma proteomics and metabolomics. Proteomic and metabolomic analyses from individuals with DS reveal the dysregulation of several pathways consistent with liver dysfunction, broadly involving the immune system, protein synthesis, and metabolism. Consistent with previous work from our group, we report that the Dp16 mouse model of DS exhibits substantial liver pathology without elevation of ALT or AST. Transcriptome analysis of Dp16 livers revealed potential mechanisms of pathogenesis and plasma samples show similar alterations of key biosignatures to the human population with DS. Individuals with DS may experience widespread liver dysfunction, however the prevalence and clinical relevance remains unclear. Further research is also needed to define the mechanisms driving liver dysfunction in DS, understand the role of trisomy of human chromosome 21 in this process, and elucidate potential therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5015T: Alzheimer's disease pathway-specific polygenic risk score training and performance evaluation in the UK Biobank cohort

Authors:

K. Bazemore[1], J. Sha[1], W-P. Lee[1], J. Haut[1], Y. Zhao[2], T. Iqbal[1], O. Valladares[1], L-S. Wang[1], G. Schellenberg[3], A. Naj[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [3]Univ Pennsylvania Sch Med, Philadelphia, PA

Abstract:

Recent developments in polygenic risk score (PRS) methods include pathway-specific PRS (ps-PRS) that aggregate trait-risk variants with putative effects along one or a cluster of biological pathways. Tools and approaches for constructing ps-PRS vary across applications; PRSet is widely used to implement a clumping and thresholding approach. To identify pathways with genetic association to Alzheimer's Disease (AD), we constructed and evaluated 24 ps-PRS for AD in the UK Biobank (UKB). We used MAGMA to perform competitive pathway enrichment analysis on Kunkle et al (*Nat Genet* 2019) summary statistics and clustered pathways with overlapping genes meeting an adjusted P-value threshold. We built PRS for each cluster,

including those composed of a single pathway, based on the Kunkle summary statistics. For each ps-PRS, we considered only SNPs mapped to genes in the included pathways. Ps-PRS were constructed using clumping and thresholding on a UKB training set (true cases = 1,186; proxy cases = 29,298; controls = 218,667). We selected the best performing PRS for each pathway based on variance explained (Nagelkerke's R^2). The best performing PRS were validated in a UKB testing set (true cases = 433, proxy cases = 9,800, controls = 72,772). We identified 38 pathways from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) meeting an adjusted P<0.25. 23 pathways clustered with at least one other identified pathway, based on a threshold of 0.4 for gene overlap. 15 pathways were in a cluster of a single pathway, for a total of 24 clusters. Cluster size varied substantially, with 2,050-57,837 SNPs mapped to genes in each cluster, prior to clumping. P-value thresholds for best-fit ps-PRS varied across pathways, though for all pathways were P<0.3. Prior to *APOE* haplotype adjustment, the highest AUC model (AUC=0.52) included Regulation of Amyloid-Beta Clearance (GO: 1900221) (OR=1.05, P=1.95×10-5), along with the top 5 principal components (PCs). Following *APOE* adjustment, the best performing model AUC (AUC=0.53) included Regulation of Humoral Immune Response Mediated by Circulating Immunoglobulin (GO:0002923) and Regulation of Complement Activation (GO:0030449) (OR=1.05, P=2.81×10-6). Our analyses implicated pathways for amyloid-beta clearance and immune response regulation. We found that the use of default thresholds might result in ps-PRS with poor predictive performance, though individual AUCs for best-performing ps-PRS remain moderate. Future directions include exploring novel pathway annotations and comparing the performance of Bayesian and machine learning methods for constructing ps-PRS to traditional pruning and thresholding.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5016T: Analysis of 36,361 whole genome sequenced samples from the Alzheimer's Disease Sequencing Project (ADSP) ⭐

Authors:

S. Choi[1], B. Akgun[2], J. J. Farrell[3], D. Wang[1], J. Yang[1], C. Zhu[1], N. Kushch[2], K. Hamilton-Nelson[2], A. Naj[4], G. M. Peloso[1], B. W. Kunkle[5], A. L. Destefano[6], E. R. Martin[2], Alzheimer's Disease Sequencing Project (ADSP); [1]Boston Univ., Boston, MA, [2]Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Boston Univ Med, Laconia, NH, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Univ. of Miami, Miami, FL, [6]Boston Univ Sch Med., Boston, MA

Abstract:

Background The Alzheimer's Disease Sequencing Project (ADSP) aims to uncover genetic variations influencing the development or prevention of Alzheimer's disease (AD) across diverse populations. The recent ADSP whole genome sequencing (WGS) data release includes data from over 36,000 individuals across 37 study cohorts. **Methods** Extensive quality control (QC) steps were executed at the genotype, variant, and sample levels. Current analyses focus on AD and related dementias (ADRD) as the outcome with all cases and controls restricted to age ≥ 55 years. Two main genetic association approaches were implemented; 1) Population-specific analysis: a Gaussian mixture model of 10 principal components was used to identify genetic similarity clusters for defining populations; analyses included logistic regression and mixed model single-variant analyses for common variants, and coding-based and noncoding-based set analyses for rare variants using SAIGE and 2) Pooled population analysis: Common and rare gene-based association tests were conducted using a customized version GENESIS/R, accounting for sex, technical covariates, population structure, and empirical relationships. Predicted loss-of-function and predicted deleterious missense variants were aggregated in the rare-variant analyses, with Bonferroni correction for statistical significance of multiple rare-variant tests. **Results** After QC, more than 293 million single nucleotide variants, 15 million deletions, and 6 million insertions were retained. Using high-quality variants, nine clusters were defined based on genetic similarity for population-specific analyses. These clusters (Clusters 1-9) comprised approximately 5,500, 1,500, 1,000, 2,000, 12,000, 5,700, 5,300, 2,600, and 320 participants, respectively. Their genetic affinities were compared with HGDP and 1000G references. For the pooled analysis, 8,697 ADRD cases and 14,758 controls were included. Among the ADRD cases, 3,271 (37.6%) were female, with a mean age-at-onset of 75.6 years. Common variant associations confirmed known ADRD signals in the *APOE* and *CR1* regions, and the rare variant gene-based analysis identified *TNS1* (Tensin 1) as suggestively associated with ADRD (OR=6.67, P=3.05x10^{-6}). **Conclusion** The analysis of WGS data from the ADSP has identified a novel suggestive ADRD association at *TNS1* at which a common variant was previously associated with cognitive decline in non-demented older adults. Additional work in the diverse populations of ADSP is expected to reveal further genetic associations in and across genetic similarity groups, providing new insights into ADRD pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5017T: Analyzing specific candidate genes implicated in the development of Class III skeletal malocclusion

Authors:

S. Srikanth; Mira Loma High Sch., El Dorado Hills, CA

Abstract:

Jaw malocclusion is characterized by a misalignment between the upper and lower jaws, and is a situation that poses a range of challenges, including difficulty in chewing, speech issues, and aesthetic concerns. This study researched the roots of the Class III skeletal malocclusion specifically; By researching and gathering information from genetic databases, candidate genes associated with this mandibular prognathism were identified and analyzed. The findings of this research were deemed statistically significant, and showed a very strong correlation between the mandibular prognathism and genes in the Fibroblast Growth Factor (FGF) family, suggesting that these genes play a role in the craniofacial development seen in an underbite. After further investigation into the molecular pathways controlled by the FGF family genes, it was determined that these genes may have involvement in processes like facial morphogenesis, bone growth, and tooth development. I hypothesized that disturbances in these pathways, possibly due to genetic mutations, could lead to malocclusion. This study is useful because it allows for potential therapeutic solutions to underbites, and early detection of the disorder from genes. This allows orthodontists to develop more personalized treatment strategies, based on specific individual genetics profiles. This will offer better treatment outcomes and improved quality of life for patients with Class III skeletal malocclusion. However, I recognize that malocclusions are influenced by various factors, including genetics, environment, and development. Further research is needed to understand how these factors work together to shape malocclusion phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5018T: *APOBR* is downregulated in EBV+ tonsils of children with obstructive sleep-disordered breathing

Authors:

H. Gomez[1], C. L. Elling[1], L. Mayher[1], O. L. Diala[2,3], C. Gardner[1], K. Willford[4], V. Z. Zamora[5], A. Agyepong[5], N. K. Lee[1], K. K. Green[1], O. A. Darr[1,2], C. R. Francom[1,2], E. D. Larson[1,6], S. A. Gitomer[1,2], A. E. Schell[1], D. N. Frank[7], N. R. Friedman[1,2], B. W. Herrmann[1,2], R. L. Santos-Cortez[1]; [1]Dept. of Otolaryngology-Head and Neck Surgery, Sch. of Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Dept. of Pediatric Otolaryngology, Children's Hosp. Colorado, Aurora, CO, [3]Dept. of Family Med. and Population Hlth., Virginia Commonwealth Univ., Richmond, VA, [4]Pathways in Genomic Res. for Undergraduate Experiences (PATH-GREU) Program, Univ. of Colorado, Denver, CO, [5]Colorado Res. Experiences (CORE) Program, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [6]Dept. of Basic and Translational Sci., Sch. of

Dental Med., Univ. of Pennsylvania, Philadelphia, PA, [7]Div. of Infectious Diseases, Dept. of Med., Sch. of Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Obstructive sleep-disordered breathing (oSDB) is a heterogeneous phenotype that is increasing in prevalence worldwide and has many potential co-morbidities that could severely affect quality of life. There is a need to identify biomarkers for oSDB and its co-morbidities to improve clinical management, particularly in children. We performed bulk mRNA-sequencing, differential expression analysis, and qPCR replication of selected differentially expressed genes (DEGs) using RNA samples extracted from tonsils of children with oSDB. Two variables were used as classifier, namely detection of Epstein-Barr virus (EBV) in tonsils, and need for continuous positive airway pressure (CPAP) treatment. Standard statistical tests were used to determine associations across clinical, EBV and DEG variables. Nineteen genes were dysregulated in tonsils that are EBV+ or from children needing CPAP. Of these genes, *APOBR* was downregulated in both EBV+ and CPAP+ tonsils, and this downregulation was replicated by qPCR in an independent set of pediatric samples. In the tonsils of adult patients with oSDB, *APOBR* was positively correlated with age, and potentially with diastolic blood pressure. Taken together, *APOBR* and DEGs in tonsillar tissues may be useful as potential biomarkers of oSDB severity and co-morbidity across the lifespan, with *APOBR* levels being dependent on latent EBV infection.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5020T: Are germline and somatic mutation rates correlated in fertile and subfertile men?

Authors:

J. Kunisaki, M. Goldberg, T. Sasani, L. Hiatt, Y. Guo, K. Aston, J. Hotaling, A. Quinlan; Univ. of Utah, Salt Lake City, UT

Abstract:

Male infertility is associated with elevated rates of aneuploidy and DNA breaks in spermatozoa and germline precursors. This condition is not well understood and is associated with poor individual and familial somatic health relative to fertile men. These observations mirror prior work from our group using multi-generational pedigrees, which showed that men with increased germline mutation rates have reduced lifespans. Together, these results indicate possible pleiotropy of mutation rate modifiers and that germline hypermutation may predict increased somatic mutation, impaired spermatogenesis, and poor health in subfertile men.

To explore these associations, we study single-nucleotide *de novo* mutations in germline and somatic cells from fertile and subfertile men using deep duplex sequencing. This technology achieves an error rate of ~1 in 100M bp by distinguishing true mutations on complementary DNA strands from errors that arise on single strands.
We first explored links between subfertility and increased mutation by duplex sequencing bulk sperm from subfertile and fertile men. We observed a significant 1.4 to 2.2-fold increase in age-adjusted germline mutation frequencies in subfertile men, indicating that their germlines are predisposed to hypermutation by the time they begin puberty. However, we failed to identify a significant difference in age-adjusted blood mutation frequencies. These contrasting findings in subfertile men highlight the pleiotropic nature of mutagenesis where across unrelated individuals, germline hypermutation may correlate with increased somatic mutation burdens and poor health in distinct tissue types.
To test this possibility, we are commencing duplex sequencing of multiple cell types from various developmental lineages to examine patterns of mutagenesis within and across male organ donors. For our proposed analyses, we will study mutations in epididymal sperm, testis, prostate, colon, blood, pancreas, and skin from twelve donors between 23 and 71 years old, including two subfertile men. Though tissues are collected from otherwise healthy organ donors, we hypothesize that age-adjusted germline mutation rates will correlate with reduced epididymal sperm count and age-adjusted somatic mutation burdens in a subset of tissues per individual. We will also present progress toward understanding how mutation rates in specific tissue types vary across unrelated individuals and how somatic mutation rates vary across organs in a single individual. Our efforts to examine links between germline and somatic mutagenesis may illuminate precise mechanisms driving poor reproductive and somatic health in subfertile men.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5021T: Assessing the Performance of a Multi-Ancestry Polygenic Risk Score for LDL with Gene-Environment Interactions of Smoking and PM2.5 Exposure in Racially and Ethnically Diverse Groups: Results from the PAGE Study

Authors:

J. Sharma[1], Y. Cai[2], Z. Wang[3], M. Lin[4], C. Gignoux[4], M. Graff[5], C. Haiman[6], C. Kooperberg[2], T. Matise[7], K. North[5], U. Peters[2], E. Kenny[3], M. Jones[1], G. Wojcik[1]; [1]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]Fred Hutchinson Cancer Ctr., Seattle, WA, [3]Icahn Sch. of Med. at

Mount Sinai, New York, NY, [4]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Univ North Carolina, Chapel Hill, NC, [6]Keck Sch. of Med., Los Angeles, CA, [7]Rutgers Univ, Piscataway, NJ

Abstract:

Introduction: The extent to which environmental contexts, such as smoking behavior and air pollution exposure, may influence the predictive performance of polygenic risk scores (PRSs) is not well studied, limiting our understanding across settings. LDL is one such biomarker affected by genetics, smoking, and air pollution. In the Population Architecture Using Genomics and Epidemiology (PAGE) Study (n=31,729), we estimated the predictive performance of a multi-ancestry PRS on low-density lipoprotein cholesterol (LDL), incorporating smoking as a gene-environment interaction. In a subset of 11,398 participants enrolled in the NYC BioMe biobank with available genetic and zip code data, we assessed the modifying roles of both smoking and area-level PM2.5 exposure in LDL prediction.

Methods: We modeled the principal component-adjusted PRS-LDL relationship via linear regression including age, sex, BMI, and HDL cholesterol. Smoking status was assessed as a modifier to the PRS-LDL relationship in the full sample and across racial and ethnic groups in PAGE. All analyses were stratified by racial or ethnic identity to investigate the influence of intersectional contexts as part of a broader goal of investigating cardiovascular health disparities. In BioMe participants, annual PM2.5 exposure at the 3-digit zip code level was assessed using 2016 Air Pollution measurements from NASA's Socioeconomic Data and Applications Center (SEDAC), split into tertiles, and assessed as an additional modifier to the above PRS-LDL relationship.

Results: Addition of both smoking and PM2.5 exposure to the base model with PRS improved model performance differentially by race and ethnicity (incremental R2 range: 0.004 to 0.053). The interaction of PRS-smoking was statistically significant for the PAGE sample for former vs. never smokers ($\beta_{former}=-1.24$, $p=0.025$), with a smaller mg/dL change in LDL per unit increase in PRS for former smokers. The PRS-PM2.5 interaction in BioMe participants was statistically significant for those in the highest tertile of PM2.5 exposure ($\beta_{tertile3}=13.20$, $p=0.047$), but not when stratified by self-identified race and ethnicity.

Conclusion: The PRS-LDL relationship and PRS-smoking interaction varied across racial and ethnic groups, with stronger influence of PRS in non-smokers. The PRS-PM2.5 interaction depicted a synergistically higher effect of the PRS on LDL in individuals at higher levels of PM2.5 exposure. The intersectional incorporation of specific environmental variables in genetic risk prediction models deepens our understanding of health disparities in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5022T: Assessment of polygenic risk score performance in East Asian populations for ten common diseases: A Korean cohort study

Authors:

H. Jung[1], H. Jung[1], E. Baek[1], J-O. Kang[2], S. Kwon[1], J. You[3], J. Lim[2], B. Oh[4,2,3]; [1]Dept. of BioMed. Sicence, Graduate Sch., Kyung Hee Univ., Seoul, Korea, Republic of, [2]Dept. of Biochemistry and Molecular Biology, Sch. of Med., Kyung Hee Univ., Seoul, Korea, Republic of, [3]Fortuna Helix Inc, Seoul, Korea, Republic of, [4]Dept. of BioMed. Sci., Graduate Sch., Kyung Hee Univ., Seoul, Korea, Republic of

Abstract:

Polygenic risk score (PRS) uses genetic variants to assess an individual's susceptibility to diseases. Although the prediction performance of PRS has been well studied in Europeans, its accuracy in East Asians has been relatively unexplored. In this study, we assessed and compared the performance of East Asian PRS-continuous shrinkage (PRS-CSs), derived from single-population genome-wide association studies (GWAS), with transferability PRSs (PRS-CSx) that integrated large-scale European and East Asian GWAS. This assessment was conducted for ten common diseases within the Health Examinees (HEXA) cohort (n = 55,870) in Korea. The predictive performance in the HEXA cohort was assessed using four metrics: the likelihood ratio test (LRT), odds ratio per 1 standard deviation (SD) PRS (perSD OR), Net Reclassification Improvement (NRI), and R2. The East Asian PRS-CSs exhibited statistical significance for LRT and perSD OR across all ten diseases and for NRI in seven diseases. PRS-CSx consistently demonstrated significant transferability compared with PRS-CS in ten diseases with an enhanced on-average predictive metric in LRT [1.31-fold], perSD OR [1.04-fold], and NRI [1.24-fold]. The difference in correlation (R2) values between PRS-CS and PRS-CSx analyzed using the r2redux method was statistically significant across all diseases, demonstrating an average increase of 0.35 % in the R2 value in PRS-CSx. Furthermore, we assessed the relative performance of these East Asian PRSs compared with their respective European PRSs for seven diseases, resulting in an average performance of 85.69 %. Our results indicate that while transferability of PRSs helps improve the performance of East Asian PRSs, it is insufficient to match the performance of European PRS. This emphasizes the significance of large-scale East Asian GWAS data in bridging the performance gap between the European and East Asian PRSs for effective disease risk prediction in East Asian populations.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5023T: Association analysis of HLA-DQB1 with narcolepsy type 2 and idiopathic hypersomnia

Authors:

T. Miyagawa[1], M. Shimada[2], T. Kodama[1], Y. Honda[1], K. Tokunaga[2], M. Honda[1]; [1]Tokyo Metropolitan Inst. of Med. Sci., Setagaya-ku, Japan, [2]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Abstract:

Narcolepsy type 2 (NT2) is a rare sleep disorder characterized by excessive daytime sleepiness (EDS) and rapid eye movement (REM) sleep abnormalities, but no cataplexy. Narcolepsy type 1 (NT1), characterized by cataplexy, EDS and REM sleep abnormalities, is strongly associated with human leukocyte antigen (HLA)-*DQB1*06:02*. Most NT1 patients consistently have *DQB1*06:02* in many populations. In this study, *DQB1* alleles were tested in 209 Japanese NT2 patients and 1,418 control subjects. The frequency of *DQB1*06:02* in the NT2 group was significantly higher than that in the control group (allele frequency: 17.2 vs. 7.8%, $P=2.6x10^{-10}$, odds ratio (OR)=2.47; carrier frequency: 32.1 vs. 14.7%, $P=3.8x10^{-10}$, OR=2.74). Clinical differences based on polysomnography (PSG), Epworth Sleepiness Scale (ESS) and multiple sleep latency test (MSLT) data between *DQB1*06:02*-positive and -negative NT2 patients were evaluated. More sleep onset REM periods (SOREMP) and shorter REM latency on the MSLT were observed in the *DQB1*06:02*-positive group than in the -negative group (SOREMP: $P=9.2x10^{-4}$, REM latency: P=0.01), suggesting that *DQB1*06:02*-positive NT2 patients are more associated with REM sleep abnormalities. We tested whether the *HLA-DQB1* allele distribution (excluding DQB1*06:02) in *DQB1*06:02*-positive (n=67) and -negative NT2 (n=142) patients was similar to that of NT1 patients (n=580). The allele distribution of *DQB1*06:02*-negative NT2 was significantly different from that of NT1 ($P=6.5x10^{-5}$), while no significant difference was observed in the distribution between *DQB1*06:02*-positive NT2 and NT1 (P=0.36), suggesting that at least *DQB1*06:02*-negative NT2 has a different etiological pathway from NT1. In addition, we examined *DQB1* alleles in 566 Japanese patients with idiopathic hypersomnia (IH) which is a rare sleep disorder characterized by prolonged and disabling EDS. No significant associations were observed between IH and *DQB1* alleles. We also analyzed *HLA* imputation data obtained from SNP arrays (116 NT2 patients and 280 IH patients) to find other associated *HLA* genes besides *DQB1*. However, no significant *HLA* associations were seen with the patient groups, except for *DQB1*06:02* with NT2.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5025T: Associations Between Gut Microbiota and Alzheimer's Disease Biomarkers in a Korean Population

Authors:

Y. Jung[1], S. Lee[1], S. Lee[1], S. Seo[2,3], J. Kim[2,3], H-N. Kim[1,4]; [1]Dept. of Clinical Res. Design and Evaluation, Samsung Advanced Inst. for Hlth.Sci. and Technology, Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Dept.s of Neurology, Samsung Med. Ctr., Sungkyunkwan Univ. Sch. of Med., Seoul, Korea, Republic of, [3]Samsung Alzheimers Convergence Res. Ctr., Samsung Med. Ctr., Seoul, Korea, Republic of, [4]BioMed. Statistics Ctr., Res. Inst. for Future Med., Samsung Med. Ctr., Seoul, Korea, Republic of

Abstract:

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by cognitive decline and memory loss. Despite extensive research, its pathogenesis remains unclear. Recent studies suggest the gut microbiome may influence AD development. This study analyzed gut microbiota associated with AD and explored potential AD biomarkers. We recruited 176 participants including AD dementia (DAT), mild cognitive impairment (MCI), and cognitive unimpaired (CU) from Samsung Medical Center. Using 16S rRNA gene sequencing, we investigated bacterial diversity and differentially abundant bacteria in fecal samples. We evaluated association between the gut microbiome and potential AD biomarkers, including *APOE4* status, plasma biomarkers Aβ42/40 and phosphorylated tau and MRI visual scores. The MCI group showed significantly lower α-diversity in Observed, Faith's, and Shannon indices ($p < 0.05$) and significant dissimilarity in the ß-diversity index ($p < 0.05$) compared to CU group. However, no significant differences were found between the DAT and CU groups in any of the diversity indices. Significant differences in AD biomarkers were found in Faith's index for *APOE4* status, in Observed and Shannon indices for Centiloid RCL global, and in evenness index for GFAP. Differential abundance analysis showed *Butyribacter* (coef: -0.591, $q < 0.05$) was significantly depleted in MCI group compared to CU group, but lost significance after adjustment. *Butyribacter intestini*, *Faecalibacterium longum*, *Clostridium* unclassified species was marginally depleted in MCI group compared to CU group, even after adjustment covariates ($q < 0.1$). No significant association between AD biomarkers and gut microbiome differential abundance was identified ($q > 0.1$). In conclusion, our results indicate a potential influence of the gut microbiome on AD, particularly in early cognitive impairment. These findings suggest alterations in gut microbiota composition and diversity may be linked to AD pathogenesis and progression. Further studies are needed to clarify these associations and evaluate the potential of gut microbiota as AD biomarkers. This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2023R1A2C2006416).

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5026T: Associations between HLA polymorphism and COVID-19 severity, susceptibility, and mortality outcomes

Authors:

P. McClelland, S. Zhou, C. Bhérer, D. Taliun; McGill Univ., Montreal, QC, Canada

Abstract:

The HLA region is associated with many complex disease phenotypes, including COVID-19. However, many associations between HLA and COVID-19 were reported by small studies and remain to be replicated. We investigated HLA allele associations with COVID severity, susceptibility, and mortality in the Quebec COVID-19 Biobank (BQC19).
We identified 75,100 high-resolution HLA alleles in 11 class I and II HLA loci using high-quality whole-genome sequences from 3,610 participants (2,270 of inferred European genetic ancestry (EUR), 244 Middle Eastern, 191 African (AFR), 141 East Asian, 111 South Asian, 100 admixed American (AMR), and 553 undetermined/admixed). Allele frequencies correlated well with the Hema-Quebec donor registry (r=0.992, n=11,827) and a global HLA panel (r=0.958, n=21,546). We performed ancestry-stratified logistic regressions followed by multi-ancestry meta-analyses for COVID-19 outcomes compared to COVID-positive controls in 2,833 participants. We tested asymptomatic COVID infection (119/2,673 cases/controls (CC)), COVID infection with respiratory support (1,739/1,053 CC), very severe COVID infection (511/2,270 CC), and death due to COVID (97/2,684 CC). No previously reported associations were replicated after correction for multiple testing (p<2.35×10^{-4}). Few suggestive hits were with alleles previously implicated in COVID, including HLA-DQB1*03:01 (protective effect in EUR, AMR, & AFR against respiratory support; OR=0.698, p=9.34×10^{-4}) and HLA-DRB1*10:01 (increased risk of death in EUR; OR=5.37, p=3.68×10^{-3}). After adjusting for comorbidities and medication use, these associations retained nominal significance and relative effect size. Among the retained nominally significant associations were HLA-DPB1*17:01 (increased risk of death in EUR; OR=5.43, p=7.641^{-4}) and HLA-DQB1*05:04 (asymptomatic COVID in EUR; OR=10.62, p=1.93×10^{-3}), which were not linked previously to COVID-19. We compared association test statistics in the top 30 hits (5%) before and after adjusting for comorbidities and medications. For COVID deaths, adjusting for comorbidities gave a small but significant improvement in median test statistics. In COVID with respiratory support, adjusting for antibiotic use alone or with comorbidities significantly lowered median test statistics.
These results highlight the challenges in replicating reported COVID-19 associations, partly driven by small samples and heterogeneity in phenotypes, populations, and methods. Most

studies overlook the complex interaction between HLA, comorbidities, medications, and COVID-19 outcomes; larger studies and better methods are needed.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5027T: Bi-ancestral analysis of the utility of type 2 diabetes polygenic score for predicting incident diabetes

Authors:

R. Hanson, S. Kobes, L. Wedekind, E. Vazquez Arreola, M. Ramirez Luzuriaga, W-C. Hsueh; NIDDK, Phoenix, AZ

Abstract:

Genome-wide association analyses (GWASs) have identified numerous genetic variants associated with type 2 diabetes, but the utility of polygenic scores (PSs) derived from these associations for predicting future incident diabetes, beyond that provided by routinely measured clinical variables, remains uncertain. We analyzed the ability of PS to predict incident diabetes in longitudinal studies consisting of 3,886 African Americans (AfrAm) and 17,345 European Americans (EurAm) with GWAS data, who were initially without diabetes, from the publicly available dbGAP resource (from ARIC, CARDIA, MESA and Framingham cohorts). There were 688 incident diabetes cases in AfrAm and 2,304 in EurAm. Among six different PSs, we found that a score composed of 1.2M variants, derived from a multi-ancestry GWAS by PRS-CSx (Ge, *Genome Medicine,* 2022) provided the best prediction with the highest standardized hazard ratio (HR) and R^2 measures. With adjustment for age, sex, parental history of diabetes, body mass index, fasting plasma glucose and serum lipid levels, and accounting for whether the study was used to derive the PS, the HR per 5000 risk alleles was 1.48 overall ($p=6.8\times10^{-32}$), 1.24 in AfrAm ($p=4.1\times10^{-3}$), and 1.55 in EurAm ($p=1.4\times10^{-31}$). To assess improvement in model prediction, the difference in C-statistic (ΔC) between the model with clinical variables only and that which also included the PS was calculated: ΔC was 0.0093 overall ($p=4.9\times10^{-9}$), 0.0141 in AfrAm ($p=2.6\times10^{-3}$), and 0.0087 in EurAm ($p=2.9\times10^{-7}$). These analyses show that PS for type 2 diabetes strongly associate with incident diabetes in AfrAm and EurAm, but the overall improvement in prediction is modest. To further assess the ability of the PS to reclassify individuals correctly according to diabetes risk, the continuous net reclassification index (NRI) was calculated: NRI was 0.302 overall ($p=1.1\times10^{-12}$), 0.313 for AfrAm ($p=3.5\times10^{-3}$) and 0.300 for EurAm (8.6×10^{-11}). The NRIs for commonly measured clinical variables were also calculated: NRIs for body mass index were 0.481, 0.377, and 0.498 for overall, AfrAm and EurAm, respectively; NRIs for serum lipids were 0.251, 0.164, and 0.266; this suggests that PS provide

utility for classifying diabetes risk that is similar to that for widely used clinical variables. The HR was significantly different between AfrAm and EurAm ($p=7.8\times10^{-4}$), but ΔC and NRI were not (p=0.21 and 0.89, respectively). Although the association is stronger in EurAm than in AfrAm, the reclassification information is comparable; this suggests that differences in strength of association between groups may be of little clinical importance, but further study of this is required.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5028T: Can the polygenic component of idiopathic Parkinson's disease (PD) explain the penetrance of monogenic PD? Findings from polygenic scores.

Authors:

A. Caliebe[1], S. Sendel[1], Z. Landoulsi[2], K. Lohmann[3], B-H. Laabs[4], I. R. König[4], P. May[2], C. Klein[3]; [1]Inst. of Med. Statistics, Kiel Univ., Kiel, Germany, [2]Luxembourg Ctr. for Systems Biomedicine, Univ. of Luxembourg, Esch-sur-Alzette, Luxembourg, [3]Inst. of Neurogenetics, Univ. of Lübeck, Lübeck, Germany, [4]Inst. of Med. Biometry and Statistics, Univ. of Lübeck, Lübeck, Germany

Abstract:

Background: To date, the causes of Parkinson´s disease (PD) remain elusive, with genetic and environmental factors contributing. Less than 15% of PD cases are due to monogenic causes, often showing reduced penetrance. It remains an open question whether the polygenic causes of idiopathic PD also explain part of the penetrance of monogenic PD. This can be investigated by applying polygenic scores (PGS) for idiopathic PD on carriers of pathogenic variants.
Aim: Our objective was to develop and validate an advanced PGS for idiopathic PD to explain more thoroughly the polygenic background. We then assessed how this PGS performs in affected and unaffected carriers of pathogenic variants in known PD genes (monogenic PD) and whether it can help elucidate its penetrance.
Methods: 1,762 PD patients and 4,227 controls (European ancestry, age at onset (AAO) ≥ 40) from the ProtectMove cohort (www.protectmove.de) were used to generate a PGS utilizing advanced PGS construction methods. The best-performing PGS was validated in two independent cohorts, COURAGE-PD (10,854 patients, 8,626 controls) and LuxPARK (822 patients, 814 controls). The same PGS was then applied to 771 carriers of pathogenic variants (335 patients, 436 controls) and 66 early AAO PD patients without any known pathogenic variants (AAO < 40 years).

Results: The best-performing PGS (928,814 SNPs) achieved an area-under-curve (AUC) of 0.680 [0.665, 0.695] on our dataset and was validated with AUCs of 0.718 [0.711, 0.725] (COURAGE-PD) and 0.667 [0.639, 0.695] (LuxPARK), outperforming existing PD-PGS. The *SNCA* region made a remarkable contribution to the PGS's discriminative ability. In carriers of pathogenic variants, the PGS achieved an AUC of 0.646, indicating that the polygenic background of idiopathic PD also accounts for part of the penetrance of monogenic PD. The PGS performed differently across mutation types, having a better discrimination for *GBA1* than for *PRKN* (AUCs: 0.647 vs. 0.605). The PGS yielded an AUC of 0.636 when applied on the early AAO patients.
Conclusion: In light of an estimated heritability of only 27% for idiopathic PD, our PD-PGS exhibits good discrimination between patients and controls. This can allow a comprehensive risk analysis on an individual basis in the future, possibly in combination with other prodromal biomarkers and environmental risk factors. The fact that the PGS also achieves an AUC well above 0.5 in carriers of pathogenic mutations suggests that genetic risk factors in idiopathic PD also influence the penetrance of monogenic forms of PD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5029T: Cardiac arrhythmia risk scales with cumulative coding and noncoding genetic liability

Authors:

T. Monroe[1], L. Pesce[1], S. Kearns[1], P. Page[1], L. Castillo[1], G. Webster[2], M. Puckelwartz[1], E. McNally[3]; [1]Northwestern Univ., Chicago, IL, [2]Lurie Children's Hosp., Chicago, IL, [3]Northwestern Univ, Chicago, IL

Abstract:

Background Cardiac arrythmias are highly heritable disorders that can be life-threatening. Clinical genetic testing yields highly penetrant (likely) pathogenic monogenic variants in some patients, but this approach is agnostic to polygenic disease architectures. Identifying genetic risk is important for predicting who might benefit from clinical intervention. We hypothesized that a biologically informed analysis of genome sequencing combining common, regulatory, and coding variation would improve genetic risk prediction.
Approach: We established a diverse cohort of 1042 individuals at low or high arrhythmia risk including those with cardiomyopathies and classified participants as cases (n=505) or controls (n=537). Rare variant analysis of whole genome sequencing centered around cardiomyopathy/arrhythmia (CMAR), and epilepsy genes, as well as the regulatory noncoding elements that map to these genes. Common variant analysis relied on computing established

polygenic scores for 18 cardiac metrics or diseases. We then created a risk stratification algorithm that incorporated all 3 analyses.
Results: Polygenic scores for blood pressure, QRS interval, and LV ejection fraction were differentially distributed between the low and high-risk groups. A composite common variant score that included these polygenic scores held prognostic value, with an odds ratio of 5.0 [95%CI 3.8-6.7] when comparing the top quintile to the lowest quintile of scores. Coding variant analysis revealed an enrichment of ultrarare and pathogenic missense CMAR and epilepsy gene variants as well as protein disrupting variants in the arrhythmia cases. Patients with a high risk clinical phenotype had an accumulation of these rare variants. These affects were synergistic: individuals with a high CMAR variant score were most likely to have arrhythmia if they also carried rare variants for epilepsy or if they carried a high common variant load. We also found >1.5-fold enrichment of ultrarare regulatory variants in the high risk group. A risk predictor that combined the common, rare, regulatory variants outperformed any single class of variants or subset, with the top 20% of the cumulative risk score carrying an odds ratio of 18.4 [95%CI 12.7-26.9]. Forward and reverse validation replicated these results. The predictor also accurately identifies high-risk individuals who lack traditional Mendelian pathogenic variants ("genotype-negative" cases).
Conclusions: Incorporating noncoding and polygenic analyses improves genetic risk stratification for arrhythmias.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5030T: Cell type-specific gene-regulatory effects of structural variants on neuropsychiatric diseases

Authors:

X. Han[1], M. Kellis[2]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]MIT, Brookline, MA

Abstract:

Structural variants (SVs) significantly contribute to genetic diversity and are implicated in human complex diseases. However, most current genome-wide association studies (GWAS) and quantitative trait loci (QTL) research focuses nearly exclusively on single nucleotide polymorphisms (SNPs) and small insertions/deletions (indels). Despite their potential importance in both gene expression and disease phenotypes, the mechanisms and cell-type-specific gene-regulatory effects of SVs on human brain function and neuropsychiatric disorders remain poorly understood.
Here, we leverage short-read whole-genome sequences (WGS) and scRNA-seq data spanning

2.3 million nuclei across 427 postmortem human brains generated by our lab to infer structural variant QTLs (SV-eQTLs), and integrate WGS with scATAC-seq data spanning 437k individual nuclei from 92 individuals to infer SV-aQTLs, and similarly infer SV-spliceQTLs. We integrate GWAS summary statistics from neuropsychiatric diseases to assess the cell type-specific genetic effects of structural variations on a variety of disorders, including schizophrenia (SCZ), bipolar disorder (BD), autism spectrum disorder (ASD), and Alzheimer's disease (AD).
We characterised approximately 100k SVs with paired cell type-specific gene expression and chromatin accessibility spanning 25 cortical cell types. Our results reveal that multiple SVs contribute to at least one molecular phenotype within specific cell types. We observed reproducible SV-xQTLs across various cell types. Moreover, we identified SV associations that were closely linked with nearby SNPs, providing a potential causal mechanistic interpretation for strong-effect QTLs involved in neuropsychiatric disorders, including SCZ, BD, ASD, and AD. Overall, our study provides crucial insights into the cell type-specific gene-regulatory roles of SVs in regulating brain molecular function and modulating neuropsychiatric disease risk.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5031T: Characterization of copy number variation in people with Pacific ancestry using PacBio HiFi Sequencing

Authors:

M. Pezzolesi[1], C. Littlefield[1], J. M. Lazaro-Guevara[2], D. Stucki[1], M. Lansford[1], M. H. Pezzolesi[1], E. J. Taylor[1], J. Taloa[1], K. Lao[1], J. P. Tavana[3], W. L. Holland[1], K. L. Raphael[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]McGill Univ., Quebec, QC, Canada, [3]Brigham Young Univ., Provo, UT

Abstract:

Individuals with Pacific ancestry have been underrepresented in human genomic research, especially in structural genetic variation analyses. Although some whole-genome sequence data from the Human Genome Diversity Project indicate that people of Pacific ancestry are likely to carry more population-specific copy number variants (CNVs) compared to other populations, comprehensive studies are lacking. This gap is significant as understanding genetic variation in individuals with Pacific ancestry could shed light on health disparities observed in this population, such as high rates of obesity and Type 2 Diabetes. To begin to address this, we sequenced the genomes of 23 individuals from Utah with ancestry from Tonga (6), Samoa (6), Fiji (4), Philippines (2), Marshall Islands (2), Guam (1), Tahiti (1), and Pohnpei (1) using PacBio's HiFi long-read sequencing technology, achieving an average depth of 33X. We generated 46 high-quality partially-phased haploid assemblies using Hifiasm and identified whole-gene CNV

duplications using liftoff and Gencode Hg38 gene annotations. Additionally, we analyzed CNV gene duplications in 94 phased haploid assemblies from individuals without Pacific ancestry from the Human Pangenome Reference Consortium (HPRC). We limited our analysis to protein-coding genes and identified 50-100 duplicated genes per assembly from individuals with Pacific ancestry, totaling 704 unique whole-gene duplications, with 376 specific to those with Pacific ancesty; 39 of these were found in 2 or more assemblies. Notably, duplications in the *NPHP1* gene, which is associated with renal disease, were absent in all HPRC assemblies but were found in 3 assemblies from individuals with Pacific ancestry. We also observed differences in gene duplication frequencies between both sets of assemblies. There were 29 genes with more than a 10% frequency increase in assemblies from individuals with Pacific ancestry and 56 genes with more than a 10% frequency increase in HPRC assemblies. The *CLPS* gene, encoding pancreatic colipase and linked to Type 2 Diabetes, was duplicated in 57% of HPRC but only 32% of assemblies from individuals with Pacific ancestry. These findings provide the most accurate analysis of whole-gene CNV duplications in individuals with Pacific ancestry to date and identify potential CNVs for further functional studies.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5032T: Characterization of the Phenotypic Consequences of the Duffy-Null Genotype

Authors:

M. Hysong[1], M. Shuey[2], J. Huffman[3], P. Auer[4], A. Reiner[5], L. Raffield[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Palo Alto Veterans Inst. for Res., Boston, MA, [4]Med. Coll. of Wisconsin, Milwaukee, WI, [5]Univ. of Washington, Seattle, WA

Abstract:

Homozygosity at rs2814778-C ("Duffy-null") is frequent in populations with ancestry from regions where malaria has historically been endemic. Utilizing data from the All of Us (AoU) cohort, we show that this genotype is present in considerable frequencies in US populations that have African-like (68%) and Middle Eastern-like (14%) genetic similarity to 1000G/HGDP reference populations as well as in self-identified Mixed (12%), Pacific Islander (7%), and Hispanic (5%) populations. While this variant is known to lead to lower average white blood cell counts, neutrophils in particular, most evidence suggests that it does not directly impact health outcomes (i.e., benign neutropenia). To investigate the phenotypic consequences of the Duffy-null genotype, we conducted a Phenome-wide association study (PheWAS) across more than

1,400 clinical conditions. We find that Duffy null status is only reproducibly associated with changes in white blood cell counts and not with any disease outcomes in AoU, the Vanderbilt University Medical Center's Biobank, or Million Veterans Program (MVP). Moreover, Duffy-null associated neutropenia is on average less severe than neutropenia in individuals without homozygous rs2814778-C. However, prior studies have shown that the misinterpretation of Duffy-null associated neutropenia and leukopenia leads to inequities in medical care and health outcomes such as unnecessary bone marrow biopsies, disparities in cytotoxic and chemotherapeutic treatment courses, and under-enrollment in clinical trials. While the majority of previous investigations into homozygosity at rs2814778-C has focused on African Americans, we show that race is not an accurate predictor of Duffy-null status or associated benign neutropenia. Our research suggests that broad genetic screening of rs2814778 across all populations could provide a more robust and accurate understanding of white blood cell counts and mitigate resulting health disparities.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5033T: Characterizing Early Life Insulin Level-defined Asthma Endotypes with Population-level Single-Cell Multi-omics Profiling

Authors:

J. Ding[1,2], H-J. Kang[2], A. Spangenberg[2], F. Martinez[2], T. Carr[2], D. Cusanovich[1,2]; [1]The Dept. of Cellular and Molecular Med., The Univ. of Arizona, Tucson, AZ, [2]Asthma and Airway Disease Res. Ctr., The Univ. of Arizona, Tucson, AZ

Abstract:

Asthma is one of the most prevalent airway diseases that has its roots in early life. A recent study from the Tucson Children's Respiratory Study revealed a notable correlation between elevated serum insulin levels at age 6 and a significantly heightened prevalence of active asthma later in life. However, the underlying mechanisms of this correlation remain uncharacterized. In our investigation, we analyzed peripheral blood mononuclear cells (PBMCs) from 57 participants of the same study, collected at age 40 and stratified by their insulin levels measured at age 6, along with subsequent asthma diagnoses. Employing an enhanced approached, we generated single-cell data including both transcriptome and epigenome profiles of 281,489 PBMCs across the study cohort, observing 6 major immune cell types and 24 subtypes. Our analysis pointed to a distinct immune signature in female individuals with an asthma diagnosis and elevated early-life insulin levels. Notably, we identified over 900 differentially expressed genes, predominantly in natural killer (NK) cells, B cells, and mucosal-associated invariant T (MAIT) cells, comparing

asthmatic females with elevated early-life insulin levels to non-asthmatic females with normal early-life insulin levels. Moreover, we observed a significant disparity in the proportion of T cells between these two groups in females, suggesting a pronounced impact on T cell populations. Leveraging our single-cell data, we also conducted genotyping and cis-QTL mapping within this study population. Our results highlighted likely causal pathways linking single nucleotide polymorphisms to differential chromatin accessibility, subsequently influencing gene expression, and ultimately potentially contributing to disease onset. In conclusion, our investigation of childhood insulin-level-stratified asthma has demonstrated a persistent shift in the immune profile of subjects into adulthood. These findings suggest a female-specific subtype of asthma driven by the intricate relationship between early-life insulin regulation and later-life asthma susceptibility. Our results offer novel insights that may inform future preventative and therapeutic strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5034T: Claudin genetic variants in kidney stone pathogenesis

Authors:

I. Y. Liu[1,2], E. MacKenzie[1,2], J. Haverfield[2], L. Dufresne[2], J. C. Engert[1,2], A. K. Ryan[1,2], I. R. Gupta[1,2,3]; [1]McGill Univ., Montreal, QC, Canada, [2]Res. Inst. of the McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [3]Montreal Children's Hosp., Montreal, QC, Canada

Abstract:

Kidney stones are hard calcium-based deposits that affect 1 in 10 adults, causing severe acute pain. Half of first-time stone formers risk recurrence, which can lead to progressive loss of kidney function via protracted inflammation. There is strong genetic and biological evidence that claudin genes (*CLDNs*) are associated with kidney stone formation: 1) CLDNs are pivotal for regulating calcium and other ion transport in the renal epithelium, 2) rare recessive mutations in *CLDN16* and *CLDN19* cause a lethal form of kidney stone with early-onset kidney failure, 3) *CLDN14* variants were associated with kidney stones in a genome-wide association study (GWAS), 4) novel rare variants in 11 *CLDN* genes were previously discovered in our local kidney stone cohort.
We hypothesized that other coding variants in *CLDN* genes increase kidney stone risk.
Methods: Using whole-exome sequencing data from the UK Biobank, we conducted a targeted association study of loss-of-function (LoF) and missense variants across all 24 human *CLDN* genes. These variants were tested for association with kidney stone formation in the European-ancestry subset, the largest available genetic-ancestry (n > 400,000).

Results: After false discovery rate correction ($\alpha = 0.05$, $N = 460$ variants), we identified 44 missense or LoF variants (MAF < 0.02) across 23 *CLDN* genes significantly associated with kidney stones. Among these, 31 variants were predicted to be functionally deleterious (scaled-CADD > 15). The most significant associations were novel variants in *CLDN16*, which, unlike the previously identified recessive mode of action, conferred risk in the heterozygous state. We also confirmed the significant association of two common *CLDN14* synonymous variants, rs219780[C] (OR = 1.12, $P = 1.97 \times 10^{-9}$) and rs219779[C] (OR = 1.13, $P = 7.89 \times 10^{-12}$), reported in a kidney stone GWAS.
Conclusions: Our study uncovered 44 *CLDN* variants significantly associated with kidney stones, suggesting that coding variants in *CLDNs* contribute to kidney stone pathogenesis. The membrane localization of CLDNs makes them promising drug targets. Future research should validate these findings in additional cohorts containing other genetic ancestries and perform functional experiments in cell models to elucidate their biological impact and therapeutic potential.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5035T: Clonal hematopoiesis is associated with a seven-year delay in the onset of Alzheimer's disease in *APOE* ε4/ε4 cases

Authors:

Y. Le Guen[1], J. Park[1], L. Talozzi[1], M. Belloy[2], M. D. Greicius[1]; [1]Stanford Univ., Palo Alto, CA, [2]Washington Univ., Saint Louis, MO

Abstract:

Background: Recently, clonal hematopoiesis of indeterminate potential (CHIP) was found to be associated with a reduced risk of Alzheimer's disease (AD) and somatic mutations found in the blood of CHIP carriers were also in microglia-enriched brain samples (Bouzid et al., 2023, *Nature Medicine*). We aimed to validate this finding in a larger dataset, explore the effect across *APOE* genotypes, and examine the association of CHIP with age-at-onset. **Methods:** Bouzid et al. (2023) analyzed the Alzheimer's Disease Sequencing Project (ADSP) whole-exome sequencing (WES) 10k subjects release and limited their analysis of this cohort to ε3/ε3 individuals, while we analyzed the new ADSP WES 20k subjects release and considered all *APOE* genotypes. We identified CHIP variants from joint-called vcf files focusing on genes and variants provided in the previous study appendix. We further filtered variants requiring a median read depth above 30 among carriers and variant allele fraction below 40%. As in Bouzid et al., we restricted our analysis to European ancestry participants whose DNA was

extracted from blood (**Table 1**). Using the generalized linear model implemented in the statsmodel Python package, we tested the association with diagnosis (Binomial family) and age-at-onset (Gaussian family) across and by *APOE* genotype. Analyses were adjusted for sex, and the first 3 principal components accounting for genetic ancestry. Additionally, the analyses considering all *APOE* genotypes were adjusted for ε2 and ε4 dosages. **Results:** We confirmed CHIP association with reduced AD risk in ε3/ε3 individuals (OR = 0.68 [0.54; 0.86], p = 0.0016), and observed a similar direction of effect across *APOE* genotypes (OR = 0.77 [0.64; 0.93]; p = 0.0058), except in ε2/ε3 (**Table 2**). In addition, we found that CHIP was associated with 2.5 years delayed age-at-onset across *APOE* genotypes (β=2.53 [1.27; 3.79]; p = 0.0001, **Table 3**), with a marked 7.7 years delayed onset in ε4/ε4 cases (β=7.71 [2.66; 12.77]; p = 0.0028). **Conclusion:** Taken together, these results provide stronger human genetic evidence that CHIP is associated with reduced AD risk and delayed age-at-onset, particularly in *APOE*-ε4 carriers. These findings corroborate the notion that rescuing microglial phagocytic capacity during aging may be a therapeutic avenue for AD, most notably in *APOE*-ε4 carriers.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5036T: Cognitive Polygenic Index and occupational complexity

Authors:

A. Tsapanou; Columbia Univ., New York, NY

Abstract:

Although the impact of occupation on cognitive skills has been extensively studied, there is limited research examining if genetically predicted cognitive score may influence occupation. We examined the association between Cognitive Polygenic Index (PGI) and occupation, including the role of brain measures. Participants were recruited for the Reference Ability Neural Network and the Cognitive Reserve studies. Occupation complexity ratings for Data, People, or Things came from the Dictionary of Occupational Titles. A previously-created Cognitive PGI and linear regression models were used for the analyses. Age, sex, education, and the first 20 genetic Principal Components (PCs) of the sample were covariates. Total cortical thickness and total gray matter volume were further covariates. We included 168 white-ethnicity participants, 20-80 years old. After initial adjustment, higher Cognitive PGI was associated with higher Data complexity (B=-0.526.069, SE=0.227030, Beta= -0.526 p=0.022, R^2=0.259) (lower score implies higher complexity). Associations for People or Things were not significant. After adding brain measures, association for Data remained significant (B=-0.496059, SE: 0.245, Beta= -0.422, p=0.045, R^2=0.254). Similarly, for a further, fully-adjusted analysis

including all the three occupational complexity measures (B=-0.568073, SE=0.237030, Beta= -0.483, p=0.018, R^2=0.327). Cognitive genes were associated with occupational complexity over and above brain morphometry. Working with Data occupational complexity probably acquires higher cognitive status, which can be significantly genetically predetermined.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5037T: Common and rare variant analyses reveal novel genetic factors underlying Idiopathic Pulmonary Fibrosis and shared aetiology with COVID-19

Authors:

A. Kousathanas[1], C. Odhams[1], J. Cook[1], Y. Hu[2], S. Klee[2], M. Erzurumluoglu[2], F. Ramirez[2], C. Mayr[2], D. Schäfer[2], L. Beck[2], I. Christ[2], F. E. Herrmann[2], Boehringer Ingelheim – Global Computational Biology and Digital Sciences, J. Duboff[1], L. Moutsianas[1], M. Brown[1], J. Kriegl[2], J. N. Jensen[2], M. J. Thomas[2], Z. Ding[2]; [1]Genomics England, London, United Kingdom, [2]Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

Abstract:

Idiopathic pulmonary fibrosis (IPF) is a devastating lung disease characterized by irreversible loss of lung function. Genetic studies have provided clues to genes and mechanisms that may be causally associated with IPF, but the disease still has limited therapeutic interventions. Recent studies have also linked IPF with severe COVID-19 through shared genetic risk factors, which could unlock additional therapy angles. We carried out a genome-wide association study (GWAS), post-GWAS and rare variant analyses utilising the whole genome sequencing data (WGS) from the 100,000 Genomes Project (100K) of a cohort of IPF participants (n=586) to identify novel associations and potential drug targets. A meta-analysis combining 100K and published GWASs of IPF (total 11,746 cases and 1,416,493 controls) identified a novel GWAS association signal at the 1q21.2 locus (rs16837903, OR[95%CI]=0.88[0.85, 0.92], P=9.54×10^{-9}) which was also successfully replicated with independent data and linked to probable effector gene *MCL1*. Rare variant burden analysis identified *ANGPTL7* as a novel IPF candidate gene (OR[95%CI]=28.79 [8.51-97.43], P=6.73×10^{-8}). *ANGPTL7* encodes a secreted glycoprotein expressed in type-1 alveolar epithelial cells and fibroblasts, and may contribute to IPF by influencing the interstitial extracellular matrix formation or inhibiting alveolar angiogenesis. We also investigated shared genetic mechanisms between IPF with severe COVID-19 and expanded the list of shared genetic loci with three novel colocalised signals at 1q21.2, 6p24.3 and 16p13.3 with probable effector genes *MCL1*, *DSP* and *RHBDF1*, respectively. By leveraging the genetic

correlation between IPF and severe COVID-19 through multi-trait meta-analysis, we further identified and replicated a novel candidate IPF signal at 2p16.1 with probable effector gene BCL11A. Our study provides a novel robustly replicated genomic signal for IPF at 1q21.2 identifying *MCL1 as* a new potential effector gene for IPF. The expanded list of shared genetic signals between IPF and severe COVID-19 contributes to the growing understanding of the shared disease mechanisms between these diseases which could unlock novel potential therapeutic treatments for IPF.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5038T: Comparison of gestational and type 2 diabetes mellitus transcriptome-wide association studies in South Asian and multi-ancestries populations

Authors:

V. Trivedi[1], H. Wheeler[2]; [1]Loyola Univ. Chicago, Chicago, IL, [2]Loyola Univ Chicago, Chicago, IL

Abstract:

Type II diabetes mellitus (T2D) and gestational diabetes mellitus (GDM) are two forms of diabetes directly linked with lifestyle and genetic components. South Asian populations in particular have increased rates of both forms of diabetes. The age-standardized prevalence of GDM for South Asian women is 7.1%, compared to 2.9% in white women. In a prior study, South Asian women with the highest 25% of T2D-derived polygenic risk scores (PRS) were found to have a 3-fold increased risk of GDM than the lowest 25%. Furthermore, it has been shown that the cumulative incidence of T2D in South Asian women who had had GDM within the past 7.6 years was 31.8% versus 3.6% for South Asian women without GDM. In addition to GDM, South Asians in the United States have a significantly higher prevalence of T2D as compared to other ancestral populations at 23%. We aimed to identify and compare genes associated with T2D to those associated with GDM in both South Asian populations and larger, predominantly European, multi-ancestries cohorts. We performed transcriptome-wide association studies (TWAS) using S-PrediXcan with GWAS summary statistics for T2D and GDM from the Pan-UK Biobank (PUKB). We used transcriptome prediction models from the Genotype-Tissue Expression Project (GTEx) from whole blood, visceral adipose, and pancreatic tissues. We used the Bonferroni correction for the number of genes tested per tissue to determine our significance threshold of $p<3.8E-6$. Sample sizes for the GDM GWAS summary statistics were n=7,762 for the PUKB multi-ancestries population (META) and n=519 for the Central South Asian population (CSA). For GDM, *HKDC1* ($p=7.0E-6$, $\beta=-0.85$) was the most significant gene reported

from the META population with no significant hits in the CSA population. Decreased expression of *HKDC1* in the pancreas was associated with GDM risk. In addition, increased expression of *SH2B2* in the pancreas was associated with GDM in the META (p=8.3E-5, β=0.97) and *SH2B2* was found to contribute to GDM and changes in patient glucose metabolism through literature review. We performed TWAS with the same GTEx tissue models for T2D using the META (n=438,684) and CSA (n=8,845) PUKB GWAS summary statistics. The most significant gene in the META population was *CCND2* in visceral adipose tissue (p=3.7E-30, β=-0.85) with an overall of 152 genes revealed as statistically significant. The most significant gene in the CSA population was *ARAP1* (p=1.3E-5, β=3.32). Decreased expression of *HMGXB4* was associated with T2D (p=6.1E-4, β=-0.48) and GDM (p=9.0E-4, β=-2.6) in the META populations. Moving forward, we aim to analyze cohorts with larger sample sizes using other biobanks like All of Us.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5039T: Complex trait analysis of cardiovascular imaging phenotypes via manifold learning

Authors:

C. Beeche, M-J. Dib, W. Witschey, J. A. Chirinos; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: Common diseases and complex traits have a highly polygenic genetic architecture, though genome-wide association studies (GWAS) lack the power to detect many shared causal genetic loci with small effects. Manifold learning can reveal the underlying architecture of multidimensional data by projecting into a low dimensional domain. Here, we perform a GWAS using the first principal component from a collection of cardiac derived imaging phenotypes to uncover the genetic architecture of shared loci identified through linear manifold learning. **Methods:** We automatically extracted 35 unique cardiac imaging derived phenotypes (IDPs) using a previous segmentation approach in the UK Biobank(n=50,086). Firstly, we performed single trait GWAS on each IDP. Next, we performed dimensionality reduction on the same 35 cardiovascular IDPs using principal component analysis and performed GWAS on the first principal component. All GWAS were performed using PLINK and were adjusted for age (at the time of imaging), sex, and the first ten genetic principal components. LDSC was used to compute heritability of PCA-GWAS and its genetic correlation to each IDP-GWAS. We performed MAGMA gene expression analysis on 19,089 genes for each individual IDP GWAS as well as the PCA-GWAS to identify significant gene-level associations. **Results:** The first principal component of the cardiac IDPs explained 64.8% of IDP variance, whereas LDSC heritability analysis revealed

that it was moderately heritable (h^2=25.5%(±1.7%)). PCA-GWAS identified 88 genes that were significantly enriched ($p < 0.05/19089$), with 8 genes not being identified in any single trait GWAS. Furthermore, multiple significant gene-sets were identified related to cardiac atrium, chamber, and heart development. More specifically, PCA-GWAS identified one genomic locus (11p14.3-14.2) outside of 250KB LD blocks of any single trait GWAS loci that was upstream of the gene *ANOCTAMIN3* (*ANO3*). Furthermore, 11p14.3-14.2 displayed significant histone modification (H3K4Me3) related to transcription initiation, suggesting a significant role in the gene-expression of *ANO3*. Genetic correlation between PCA-GWAS and single IDP GWAS revealed an average correlation of 0.41(±0.36). Interestingly, PCA-GWAS was negatively correlated with left ventricle ejection fraction, right ventricle ejection fraction, left atrium ejection fraction, and right atrium ejection fraction. **Conclusion:** We demonstrate the feasibility of PCA to uncover variants not detected in single trait GWAS. Further studies should be performed to evaluate this approach in detailed cardiovascular genetics studies.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5040T: Comprehensive analysis of Human Leukocyte Antigen (HLA) diversity and association with infectious and immune-mediated diseases in the BioMe biobank

Authors:

A. Richardson, A. Abhijith, C. Caggiano, R. Shemirani, E. Kenny, S. Asgari; Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, 10027, New York, NY

Abstract:

Background: *HLA* refers to a cluster of polymorphic genes on chromosome 6 which encode for proteins responsible for antigen presentation. Due to this role, HLA genes are associated with numerous infectious, autoimmune, and allergic diseases. These associations result from studies which primarily include individuals of European ancestry. Therefore, information regarding imputation quality, HLA diversity, and disease associations in non-European populations remains limited. **Methods:** To address this gap, we utilized genotyping data from a diverse cohort through The Mount Sinai BioMe Biobank, which includes over 50,000 individuals. We used admixture (K=10 clusters) to infer genetic ancestry and shared identity-by-decent (IBD) tracks to define fine-scale population structure, with the 1000 Genomes Project and HGDP as reference. We then used PhecodeX to consolidate phenotype information from electronic health records. Finally, we performed HLA imputation on Michigan Imputation Server with a diverse reference panel of 20,349 samples. Currently, we are testing the association of HLA with infectious and

immune-mediated diseases at the SNP, amino acid, and haplotype-level. **Results:** 64% of BioMe participants are of non-European ancestry. This includes 37,209 African (AFR), 13,702 Hispanic (HIS), and 1,848 East Asian (EAS) individuals. IBD analysis identified 24 fine-scale population clusters with Afro Caribbean, Puerto Rican, and Ashkenazi Jewish being the largest clusters (N = 10,761,8,296, and 7,245, respectively). Following HLA imputation, we found 394 classical HLA alleles at the two-field resolution. Imputation quality was high (>0.97) for all genetic ancestry and fine-scale population groups. Although different numbers of unique two-field alleles were found per ancestry (AFR = 332, EAS = 242, EUR = 301, HIS = 326, SAS = 218), this likely reflects different population size. However, even in populations with a small number of individuals, we identified a relatively large number of high-quality unique HLA alleles. For example, we identified 132 HLA alleles among the 157 individuals assigned to the Garifuna IBD cluster (compared to 297 alleles in the Puerto Rican cluster). Moving forward, we will test the association of HLA alleles with >80 infectious and inflammatory diseases with at least 50 cases in BioMe. **Conclusions:** Our results highlight the utility of BioMe to study HLA diversity in global populations, particularly those rarely included in past studies. Given the HLA diversity, and its pivotal role in pathogenesis of immune-mediated diseases, we expect to identify several novel HLA-disease associations in this cohort.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5041T: Context Matters: Alzheimer Disease Heritability Differs by Population

Authors:

S. Liu[1], W. S. Bush[1], B. W. Kunkle[2], G. S. Byrd[3], C. Reitz[4], G. Tosto[4], J. M. Vance[2], M. Pericak-Vance[2], J. L. Haines[1], D. C. Crawford[1], S. M. Williams[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Miami, Miami, FL, [3]NC A&T State Univ, Greensboro, NC, [4]Columbia Univ, New York, NY

Abstract:

Late-onset Alzheimer disease (LOAD) is highly heritable. However, little is known about the range of narrow sense LOAD heritability estimates across populations. Here, we thoroughly examine the impact of context, defined as recorded race/ethnicity and cohort, on LOAD narrow sense heritability using Alzheimer's Disease Sequencing Project (ADSP) pedigrees and genome-wide data with overlapping participants. ADSP pedigrees included five groups identified by recorded race/ethnicity or nationality: US non-Hispanic white ($n_{pedigrees}$=404; $n_{participants}$=7,024), Dominican ($n_{pedigrees}$=100, $n_{participants}$=3,988), Dutch isolate ($n_{pedigrees}$=10; $n_{participants}$=655), African

American ($n_{pedigrees}$=13; $n_{participants}$=286), and Puerto Rican ($n_{pedigrees}$=3; $n_{participants}$=132). Family-based heritability was estimated adjusting for 1) age and sex, 2) age, sex, and APOE4 carrier status, 3) age, sex, and cohort, and 4) age, sex, cohort, and APOE4 carrier status using Statistical Analysis for Genetic Epidemiology (S.A.G.E.) and Sequential Oligogenic Linkage Analysis Routines (SOLAR). Regardless of method or adjustments, the resulting estimates were highly variable by race/ethnicity and cohort. Using S.A.G.E. adjusted for age and sex (model 1), estimates were highest for the Dutch isolate (78.3%) followed by African Americans (39.1%), Dominicans (31.7%), and non-Hispanic whites (29.1%). The inclusion of APOE4 allele status (model 2) reduced heritability estimates by an average of 4.9%, with the largest reduction in African Americans (6.7%) and non-Hispanic whites (6.2%). Adjustment for cohort (models 3 and 4) had the most impact in non-Hispanic whites (average 3.2% decrease) and African Americans (average 3.8% increase), reflecting the larger number of contributing cohorts for these populations compared with the other populations. The Puerto Rican dataset failed to converge in all models due to small sample size. In general, narrow sense heritability estimates were almost always higher (45.4% - 80.2%) using SOLAR compared with S.A.G.E (20.4% - 81.9%), reflecting modeling and missing value handling differences, with trends remaining consistent regardless of magnitude. While family-based estimates complement previous twin-based and genome-wide studies, the wide range of narrow sense heritability estimated across populations underscores the need for cautious interpretation both regarding the context of the study population and methodological assumptions. Work supported by AG074865 and AG072547 is ongoing to estimate chip-based heritability in overlapping participants for direct comparisons with family-based heritability.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5042T: Context-specific eQTLs reveal causal genes underlying shared genetic architecture of critically ill COVID-19 and idiopathic pulmonary fibrosis

Authors:

T. Dalapati[1], L. Wang[2], A. Jones[1], J. Cardwell[3], I. Yang[3], D. Ko[1]; [1]Duke Univ., Durham, NC, [2]Duke Univ, Durham, NC, [3]Univ Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Background: Identification of the causal genes and pathogenic mechanisms underlying disease susceptibility are critical challenges in human genetics. Most genetic variants identified through genome-wide association studies (GWAS) are suspected to be regulatory, but only a small

fraction colocalize with expression quantitative trait loci (eQTLs, variants that are associated with expression of a gene). Therefore, it is hypothesized but largely untested that integration of disease GWAS with context-specific eQTLs and other molecular QTLs will reveal the underlying genes driving disease associations.

Methods and Results: Using the interactive Cross-Phenotype Analysis of GWAS database (iCPAGdb), a novel tool developed by my lab to compare GWAS and discover pleiotropic variants, I identified five genome-wide significant variants associated with both critically ill COVID-19 and idiopathic pulmonary fibrosis (IPF). IPF is clinically similar to pulmonary fibrosis, or lung scarring, that may develop in patients recovering from COVID-19. While a previously reported variant in the *MUC5b* locus colocalized with GWAS and healthy tissue eQTL datasets, the other four variants did not colocalize. Instead, through harmonization and colocalization analyses of multiple context- and tissue-specific eQTL datasets, my results demonstrate that two of the four variants colocalize in only disease-specific and lung cell-specific eQTL datasets. These analyses point to low *DPP9* expression, from the G allele of rs12610495, acting in fibroblasts and in the context of IPF lungs to drive increased risk of IPF and critically ill COVID-19. Similarly, increased expression of *ATP11A*, from the C allele of rs12585036, in monocytes and in lungs from smokers, has opposite effects in increasing risk of IPF while decreasing risk of critically ill COVID-19. Consistent with my genetic associations, analysis of causal genes in RNA sequencing datasets of human tissues demonstrates significantly lower expression of the causal genes in bulk lung tissue from IPF and COVID-19 patients compared to non-diseased controls. *ATP11A* was significantly lower in epithelial cells, while *DPP9* was lower in epithelial, immune, and stromal cells relevant to COVID-19, implicating these cell types in disease risk and pathogenesis.

Conclusions: My findings highlight the power of integrating GWAS and context-specific eQTL datasets to harness human genetic variation to identify causal genes that may play a role in multiple human diseases. Further experimentation involving the identified causal genes may elucidate mechanistic roles during both pulmonary fibrosis and COVID-19.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5043T: Convergence of evidence between two methods for cell-type specific association studies robustly identifies differentially expressed genes in schizophrenia

Authors:

K. Aberg, L. Xie, T. Campbell, M. Zhao, E. Van den Oord; Virginia Commonwealth Univ., Richmond, VA

Abstract:

Schizophrenia (SZ) is a common devastating mental disorder, characterized by hallucinations, delusions and disorganized thinking and speech. The aim of this study was to identify differentially expressed genes for SZ in the prefrontal cortex (PFC) on a cell-type specific level. For this purpose, we sought a convergence of findings between two distinct methodologies to eliminate artifacts specific to any one method, thus enabling more robust conclusions. Analyses were performed on data from PsychENCODE, including single-nucleus (sn) (450 000 nuclei from 65/75 cases/controls) and bulk RNA-seq (315/300 cases/controls) data as well as a newly generated bulk RNA-seq dataset (304/291 cases/controls). Cell-types were identified using snRNA-seq data from 92 unaffected "control" subjects from 7 other snRNA-seq studies. Seventeen cell-types were robustly detected across all studies, subregions of the PFC, sex and age groups. Next, we trained a classifier in this reference data to map the nuclei for the SZ snRNA-seq data on these 17 cell-types with a mean/median assignment probability for the nuclei of 0.973/0.993. The reference data was also used to create a panel that allowed association tests for the same 17 cell-types in the bulk RNA-seq data through a statistical deconvolution approach. The cell-type specific transcriptome-wide association studies showed no signs of test-statistic inflation (lambda ranging from 1.05-1.25 for snRNA-seq and 0.94-1.09 for bulk RNA-seq). Analyses between suggestively significant findings from the snRNA-seq analysis and the cell-type specific bulk analysis showed significant enrichment for 7 of the 17 cell-types ($P<0.05$). This suggested robustness in findings. Using Fisher's method for combining P values, we identified the majority of overlapping findings in excitatory neurons (deep layer 2/3, 175 genes/185), interneurons (PV/SV2C, 132/105), and glial cells (microglia/oligodendrocytes/OPC, 70/55/97). Gene Ontology analysis of these overlapping genes revealed biological themes such as immune response, DNA conformation change, and DNA- and RNA binding.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5044T: Copy number variants in familial bipolar disorder ascertained from Anabaptist founder populations

Authors:

H. Sung[1], L. Kassem[1], E. Besancon[1], F. Lopes[1], S. Detera-wadleigh[1], N. Akula[1], A. Nardi[2], T. Schulze[3], A. Shuldiner[4], F. McMahon[1]; [1]HGB/NIMH/NIH, Bethesda, MD, [2]Inst. of Psychiatry, Federal Univ. of Rio de Janeiro, Rio De Janeiro, Brazil, [3]Inst. of Psychiatric Phenomics and Genomics, Univ. Hosp., LMU Munich, Munich, Germany, [4]Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

The Amish Mennonite Bipolar Genetics (AMBiGen) study aims to identify genetic variants that substantially increase the risk for bipolar disorder (BD) and related conditions within Anabaptist founder populations. Participants were ascertained through clinical settings or advertisements and assessed using semi-structured psychiatric interviews. In this study, the affected individuals diagnosed with BD, schizoaffective disorder, recurrent major depression, or schizophrenia were compared with their unaffected relatives. Copy number variants (CNVs) were detected based on whole exome sequencing with the Copy number estimation using the Lattice-Aligned Mixture Models (CLAMMS) algorithm. A total of 508 unique CNVs met the quality control criteria of a quality score &gt 1, copy number not equal to 2, CNV size between 10kb and 2 Mb, and carrier frequency &lt 0.1 in samples with a total CNV count &lt 40. CNV overlaps across samples were called based on shared genes. A total of 300 unique CNVs (88 deletions and 212 duplications) in 913 individuals were analyzed further. Of these individuals, 253 met narrow criteria for BD, 396 met broad criteria that included BD and related conditions, 367 were psychiatrically healthy and 150 were unphenotyped (considered unaffected under the narrow criteria, missing under the broad criteria). The effective number of independent samples was estimated from related individuals within each group using the genetic relationship matrix. Fisher's exact tests were conducted to test the association between BD affection status and 1) carrier status for each CNV in an individual, and 2) carrier status of the most loss-of-function intolerant gene within all CNVs detected in an individual. Some known neuropsychiatric CNVs were detected in different families, including duplications and deletions on 1q21.1, 15q11.2, 16p11.2, and 22q11.2. A duplication overlapping the *CES1* gene at 16q12.2 - not previously reported in neuropsychiatric disorders -- was significantly associated with BD (both narrow and broad criteria) after Bonferroni correction. Overall, duplications of intolerant genes (pLi&gt0.9 and LOEUF&lt0.03) were significantly enriched in narrow affected individuals. These results suggest a novel association with a duplication overlapping ***CES1*** and support a role for CNVs overlapping highly intolerant genes in familial BD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5045T: Cumulative incidence estimates across polygenic score strata in global populations

Authors:

B. Wolford[1], D. Dunca[2], V. Tozzo[3], K. Zguro[4], M. Kanai[5], Y. Wang[6], P. De Macedo Gomes[7], M. Kohailan[8], R. M. Badji[8], Y. Al-Sarraj[8], K. Lin[9], K. Läll[10], B. Wingfield[11], B. Pasaniuc[12], R. G. Walters[13], D. A. Bennett[14], D. van Heel[2], H. Heyne[15], A. Martin[6], N. Mars[16], A. Ganna[17], S.

Ripatti[17], S. Kanoni[2]; [1]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [2]Queen Mary Univ. of London, London, United Kingdom, [3]Dept. of Computational Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [4]Univ. of Siena, Siena, Italy, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Massachusetts Gen. Hosp., Boston, MA, [7]Hasso Plattner Inst. for Digital Engineering, Univ. of Potsdam, Potsdam, Germany, [8]Qatar Biobank, Qatar Precision Hlth.Inst., Doha, Qatar, [9]Nuffield Dept. of Population Hlth., Univ. of Oxford,, Oxford, United Kingdom, [10]Univ. of Tartu, Tartu, Estonia, [11]European Molecular Biology Lab., European Bioinformatics Inst., Wellcome Genome Campus, Cambridge, United Kingdom, [12]UCLA, Los Angeles, CA, [13]Univ. of Oxford, Oxford, United Kingdom, [14]Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [15]Hasso Plattner Inst., Potsdam, Germany, [16]Univ. of Helsinki, Helsinki, Finland, [17]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland

Abstract:

Background: Age- and sex-specific polygenic score (PGS) effects have been previously reported in European ancestry individuals, but it is unknown to what extent these occur in multi-ancestry groups. Within the INTERnational consortium of integratiVE geNomics prEdiction (INTERVENE), we investigated PGS effects on cumulative disease incidence in over 1.1M biobank participants (30% non-European).
Methods: We leveraged information from 10 biobanks with participants from groups with genetic similarity to reference populations for European (1000G-EUR-like), East Asian (1000G-EAS-like), South Asian (1000G-SAS-like), African, Middle Eastern and admixed American ancestries. We focused on 18 diseases which contribute substantially to the global burden of disease as measured by the Global Burden of Disease (GBD) study. We calculated MegaPRS-derived PGSs and used the PGS Catalog Calculator (pgsc_calc) to normalize PGS across genetic ancestry similarity groups. We employed our recently developed framework using GBD baseline statistics to estimate cumulative incidence of disease via Cox proportional hazards models.
Results: PGS had a larger effect for coronary heart disease (CHD) in 1000G-SAS-like males compared to females (HR male = 1.34 [95% CI: 1.29-1.39] versus HR female = 1.20 [95% CI: 1.14-1.27], PGS*SexInteraction P-value = 7.92x10-4). At age 80, the cumulative incidence for males in the top 5% of polygenic of risk was over double that of females in the same PGS strata with 42.9% (95% CI: 36.5-48.7%) for males and 20.3% (95% CI: 15.4-25.5%) for females. For Type 2 diabetes (T2D), the effect was larger in 1000G-EAS-like females compared to males (HR male = 1.53 [95% CI: 1.50-1.55] versus HR female = 1.62 [95% CI: 1.59-1.66], PGS*SexInteraction P-value = 1.76x10-5). However, females in the top 5% of polygenic risk still had a lower cumulative incidence at age 80 than males (28.7 % [95% CI: 25.9-31.1%] versus 37.5% [95% CI: 34.5-40.2%] respectively) due to the underlying risk estimated from GBD. Heterogeneous effect sizes by age were detected for CHD, Type 1 diabetes and T2D in 1000G-SAS-like, while T2D and prostate cancer exhibited age-specific effects in 1000G-EAS-like. Trends for age and sex-specific effects of

the PGS were often consistent with the 1000G-EUR-like, indicating a generalized genetic effect across populations.
Conclusion: Our results demonstrate variable sex- and age-specific cumulative indices by PGS strata across global ancestries. This emphasizes the importance of country- and ancestry-calibrated cumulative incidence estimates for global clinical utility of PGS. In the future, we aim to implement ancestry-aware PGS scores.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5046T: Deciphering the Genetic Connection in Juvenile Idiopathic Arthritis and Adult Rheumatic Diseases

Authors:

J. Li[1], H-Q. Qu[1], J. Glessner[2], H. Hakonarson[3]; [1]the Children s Hosp. of Philadelphia, Philadelphia, PA, [2]Childrens Hosp. of Philadelphia/PENN, Philadelphia, PA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Background: Juvenile idiopathic arthritis (JIA) shares clinical and molecular features with adult-onset arthritis, yet the full extent of its genetic relationship with adult rheumatic diseases remains to be elucidated. Understanding this relationship is crucial for improving treatment strategies and patient outcomes.
Methods: We conducted genetic correlation analyses to identify genomic regions presenting significant correlation with JIA and three adult rheumatic diseases (Rheumatoid Arthritis/RA, Systemic Lupus Erythematosus/SLE and Systemic sclerosis/SSc). Cross-trait meta-analysis was utilized to identify common loci, and candidate genes and potential drug targets were prioritized through various methods. Mendelian Randomization (MR) was employed to identify proteins with potential contribution to disease etiology and to explore causal relationship between diseases. Furthermore, scRNA-seq data and GWAS signals were integrated to determine key cell types implicated in rheumatic disorders.
Results: Our analysis revealed a significant global correlation between JIA and adult rheumatic diseases. Local genetic correlation analysis revealed 84 regions harboring common signals linked to multiple diseases, notably highlighting the JAK-STAT pathway. Cross-trait analyses uncovered 8 novel JIA loci and 4 novel SSc loci. Additionally, 20 loci demonstrated shared effects between JIA and adult rheumatic diseases. Proteomics data integration revealed 12 proteins with a significant causal effect on rheumatic disorders, with 9 as potential drug targets. MR analysis suggested a positive causal link from JIA to RA, indicating a potential increased risk of

developing adult rheumatic diseases in JIA patients. Cell-type enrichment analysis highlighted both shared and organ-specific/disease-specific critical cell types.
Conclusion: Our findings reveal significant genetic overlap and causal links between JIA and adult rheumatic diseases. The discovery of shared genetic loci and proteins supports a new JIA classification strategy focused on the risk of developing adult rheumatic diseases. This shared genetic foundation also suggests the potential for repurposing adult rheumatic disease drugs for pediatric patients with similar mechanisms.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5047T: DementiaSeq2, an extended public resource of whole genome sequencing data for study of Lewy Body Dementia and Frontotemporal Dementia / Amyotrophic Lateral Sclerosis genetics

Authors:

C. Alba[1], G. Sukumar[1], E. Rice[1], S. Christenson[1], M. Ognoon[1], P. Ruffo[2], A. Ray[3], M. Rasheed[2], C. Fevga[2], J. Ding[2], S. J. Lubbe[4], C. M. Morris[5], J. Attems[5], J. Agin-Liebes[6], M. D. Cykowski[7], J. Troncoso[8], C. Cruchaga[9], D. J. Irwin[10], G. Serrano[11], T. Beach[11], B. F. Boeve[12], D. Dickson[13], R. Chia[2], A. Chiò[14], J. Gibbs[2], O. Ross[13], J. Landers[15], B. Traynor[2], S. Scholz[16], C. Dalgard[17], International LBD Genomics Consortium; [1]Henry M Jackson Fndn. / The American Genome Ctr., Bethesda, MD, [2]Lab. of Neurogenetics, Natl. Inst. of Aging, NIH, Bethesda, MD, [3]Neurodegenerative Diseases Res. Section, Natl. Inst. of Neurological Disorders and Stroke, NIH, Bethesda, MD, [4]The Ken and Ruth Davee Dept. of Neurology, Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [5]Translational and Clinical Res. Inst., Campus for Ageing and Vitality, Newcastle Univ., Newcastle, United Kingdom, [6]Dept. of Neurology, Columbia Univ. Irving Med. Ctr., New York, NY, [7]Dept. of Pathology and Genomic Med., Houston Methodist Hosp., Houston, TX, [8]Johns Hopkins Sch. of Med. Faculty, Baltimore, MD, [9]Washington Univ., Saint Louis, MO, [10]Dept. of Neurology, Univ. of Pennsylvania Sch. of Med., Philadelphia, PA, [11]Banner Sun Hlth.Res. Inst., Sun City, AZ, [12]Dept. of Neurology, Mayo Clinic, Rochester, MN, [13]Mayo Clinic, Jacksonville, FL, [14]Rita Levi Montalcini Dept. of NeuroSci., Univ. of Turin, Turin, Italy, [15]UMass Mem. Med. Ctr., Dept. of Neurology, Worcester, MA, [16]NIH, Bethesda, MD, [17]The American Genome Ctr., Collaborative Hlth.Initiative Res. Program, Uniformed Services Univ., Bethesda, MD

Abstract:

Lewy body dementia (LBD) and frontotemporal dementia/amyotrophic lateral sclerosis (FTD/ALS) are common complex neurodegenerative diseases with poorly understood genetic

etiologies despite the presence of familial cases. The availability of high-throughput, cost-effective whole genome sequencing technologies has led collaborative efforts to initiate population-scale genomics profiling studies of patients diagnosed with LBD and FTD/ALS, previously collected biosamples from neuropathologically confirmed cases, as well as controls (matched age greater than 60 years, neurological assessments, and/or histopathological negative features after autopsy examination). Previously, we completed DementiaSeq1 by sequencing 9,500 individuals to advance the understanding of the genetic underpinnings of LBD and FTD, including data sharing with the public. Here, we describe the extension of this dataset during the DementiaSeq2 effort. To address an aim to evaluate the genetic determinants of LBD and FTD/ALS in diverse populations, the sample assembly includes additional whole genome sequencing data from approximately 12,000 individuals (3,000 LBD, 6,000 FTD/ALS, and 3,000 neurologically normal controls) of an enriched selection from diverse subpopulations outside of EUR. Whole genome sequencing data includes generation on Illumina NovaSeq 6000 and NovaSeq X Plus platforms. As part of this updated resource, we also harmonize data by re-computing alignment, variant calling and annotation of DementiaSeq1 and merging with current DementiaSeq2 data generation using the DRAGEN 4.3 pipeline by On Premise and Google Cloud environments. Furthermore, this pipeline leverages graph reference genome alignment and an assembly of additional calling pipelines, including HLA, repeat expansion, pharmacogenomics, etc. callers. The raw, sample-level and cohort-level data will be made publicly available on multiple data repositories matching previous DementiaSeq1 data-sharing accessions. We hope DementiaSeq2 offers potential to identify novel genetic markers associated with non-Alzheimer dementias and inform the development of targeted therapies.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5048T: Design of a pragmatic clinical trial to assess outcomes after incorporating genetic susceptibility into coronary heart disease risk estimates

Authors:

M. Hamed[1], M. Naderian[2], H. Bangash[1], S. Saadatagh[3], A. Sherafati[4], A. Miller[1], J. Petrzelka[1], A. Kamkazabek[1], M. Beasley[5], J. Cortopassi[6], B. Davis[7], A. Espinoza[8], Q. Feng[9], B. Goff[7], Y. Luo[8], S. Wood[10], J. Cimino[5], G. Jarvik[11], L. Kottyan[12], C. Prows[13], W-Q. Wei[14], N. Limdi[6], I. Kullo[1]; [1]Mayo Clinic, Rochester, MN, [2]Mayo Clinic, ROCHESTER, MN, [3]Baylor, Houston, TX, [4]Univ. of Toledo, Toledo, OH, [5]Univ. of Alabama, Birmingham, AL, [6]Univ. of Alabama at Birmingham, Birmingham, AL, [7]UAB, Birmingham, AL, [8]Northwestern Univ., Evanston, IL, [9]Vanderbilt Univ. Med. Ctr., Nashville, TN, [10]Cincinnati Children'S Hosp. Med.Ctr., Cincinnati, OH, [11]Univ Washington Med

Ctr., Seattle, WA, [12]Cincinnati Children s Hosp. Med. Ctr., Cincinnati, OH, [13]Children s Hosp Med Ctr, Cincinnati, OH, [14]Vanderbilt Univ, Nashville, TN

Abstract:

Introduction: We describe the design of the eMERGE IV clinical trial to assess outcomes after return of a 'genome-informed risk assessment' (GIRA) report for coronary heart disease (CHD) to 20,000 adult participants. **Methods:** A multi-ancestry polygenic risk score (PRS) (PGP000602) was calculated from high-density genotype data and pathogenic/likely pathogenic variants in *LDLR, APOB, PCSK9* and *LDLRAP1* ascertained from a targeted sequencing panel. Family history was obtained using MeTree (Duke Center for Precision Medicine) and/or a shorter survey. These data, along with relevant clinical variables, were exported to a REDCap database to generate the GIRA report. Participants with a PRS in the top 5th percentile, monogenic risk, or family history of CHD were considered to have high CHD genetic risk, and an integrated score that combined PRS and Pooled Cohort Risk Equations (PCE) to estimate 10-year CHD risk was displayed in such participants aged >40 years. High PRS/monogenic and positive family history results were returned by a study coordinator/genetic counselor/study physician and letter, respectively, and an in-basket message was sent to clinicians. The primary outcome, initiation or intensification of lipid lowering therapy (LLT), within 12 months after return of results will be compared between the high CHD genetic risk (high PRS, positive monogenic or positive family history) and non-high CHD genetic risk groups. Secondary outcomes include new CHD diagnosis, Cardiology referral, CHD screening tests, and participant-reported lifstyle changes. **Results**: As of May 16th, 2024, 9,788 adults (66.1% females, mean age 51.2 years, 55.0% self-reported White) had data for PRS, monogenic risk, and family history. 1,482 (14.4%) participants had a high risk GIRA for CHD; 419 (4.3%), 74 (0.7%), and 976 (10.0%) had high PRS, monogenic risk, and family history for CHD respectively. Compared to participants without high CHD genetic risk, those with high CHD genetic risk were older (mean, 54.0 vs 50.8, years) and more often had self-reported CHD (8.8% vs 3.7%), baseline PCE ≥ 7.5 (40.2% vs 31.0%), baseline integrated score (IS) ≥ 7.5 (43.6% vs 29.5%), and were more often on medications for cardiovascular risk factors (lipid lowering 37.5% vs 24.1%; anti-hypertensives 47.5% vs 33.6%; anti-diabetes 17.4% vs 11.9%). **Conclusion:** We describe the design of a pragmatic clinical trial to assess outcomes after return of a genome-informed risk assessment for CHD and report baseline characteristics of the participants. Integrating PRS in the PCE identified an additional 3.6% who qualify for statin initiation providing the opportunity for early intervention and prevention.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5049T: Development of a scoring system to guide action based on genetics, biomarkers, and lifestyle data: Health Composition

Authors:

J. Zhan, J. Ashenhurst, B. Koelsch; 23andMe, Sunnyvale, CA

Abstract:

Genetic research has significantly advanced our prediction for the most likely disease risks for individuals. Polygenic risk scores (PRSs) have been notably successful in this domain, with considerable efforts dedicated to enhancing their predictive accuracy. Despite the advancements, PRSs face inherent limitations: the immutable nature of genetic information, the modest proportion of heritability explained by genetic variation alone, and etc.

To address these challenges, our study integrates PRSs with comprehensive lifestyle data, family history, and biomarkers. Utilizing the rich datasets of the UK Biobank, we developed a heuristic model to categorize individuals into personalized and nuanced risk profiles for the purpose of highlighting areas of either clinical intervention or behavioral change. This approach allows for a more granular risk stratification than PRSs alone, as shown through comparative analysis using KM curves. Our findings demonstrate significantly enhanced stratification capabilities when integrating more comprehensive risk factors, indicating the potential for more targeted and effective health interventions prior to the onset of disease.

Incorporating lifestyle factors into the risk assessment systems offers actionable opportunities for individuals to modify their health risk proactively. By linking these modifiable factors with genetic predispositions, we not only predicts future risk but also empower individuals with personalized prevention strategies. This dual focus on prediction and prevention addresses a critical gap in current genetic risk assessment methodologies and provides opportunities for health care systems to both treat and prevent disease.

Moreover, our research extends to the predictive refinement within and between related disease categories, cardiovascular disease, cancer, metabolic diseases, and neurodegenerative diseases. By integrating disease-specific elements from each of these six categories (genetics, family history, biomarkers, and representations of diet, sleep, and exercise), our model achieved superior predictive performance across these major health challenges. Our approach underscores the importance of a comprehensive view of disease prediction that considers the interplay of both genetic and behavioral factors.

This study highlights the transformative potential of combining state of the art PRSs with broader health determinants to refine disease risk predictions. Our findings advocate for a shift towards more comprehensive, personalized health risk assessments that can adapt to the diverse genetic backgrounds and lifestyles of the global population.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5050T: Developmentally driven gene program in fat depot-specific adipogenesis.

Authors:

S. Kumar[1], M. Aceves[1], J. Granados[1], L. Guerra[1], A. Leandro[2], J. Peralta[2], S. Williams-Blangero[2], J. Curran[2], J. Blangero[2]; [1]UTRGV, Sch. of Med., McAllen, TX, [2]UTRGV, Sch. of Med., Brownsville, TX

Abstract:

The subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT) contribute differentially to metabolic disease risk. The latter has been associated with a higher risk of type 2 diabetes, cardiovascular disease, non-alcoholic fatty liver disease/steatohepatitis, and certain cancers. Recent studies have focused on the genetic and transcriptomic basis of SAT and VAT predisposition, however, underlying molecular mechanisms are yet to be fully understood. We used a novel induced pluripotent stem cell (iPSC) based in-vitro approach to identify molecular mechanisms involved in SAT and VAT development propensity. Adipogenic differentiation of validated iPSC lines generated from six participants of our Mexican American Family Study (MAFS) was performed using two independent differentiation protocols. The individuals belonged to two sibships, two had high MRI-measured liver fat (>15%) and body mass index (BMI) below 35, two had BMIs above 45, and two had low liver fat and BMI below 30. The generated adipocytes were analyzed by cellular phenotyping and genome-wide mRNA sequencing analyses. Adipocytes of two individuals who had high liver fat showed VAT propensity in both the differentiation experiments, whereas others were of SAT type. A total of 1189 and 854 genes were found significantly differentially expressed (DE) between adipocytes of VAT and SAT propensities generated in the two experiments, respectively. While all generated adipocytes showed significant upregulation of adipocytes associated genes with a highly significant enrichment of adipocytes upregulated transcriptome in the human gene atlas adipocyte gene set (enrichment p-value $\leq 1.48\times10^{-36}$) and a highly uniform transcriptomic profile across all samples ($r^2 \geq 0.90$), the expression of adipocyte-specific genes was comparatively lower in samples with VAT propensity as compared to SAT samples. The VAT and SAT samples also showed characteristic expression profiles of *HOX* gene clusters. The expression of genes of the *HOXB* cluster was upregulated in VAT samples and the expressions of genes of the *HOXA, HOXC,* and *HOXD* cluster were upregulated in SAT samples. The functional annotation and pathway enrichment analysis of genes that were DE between VAT and SAT samples showed significant enrichment and upregulation of TGF-β and Interleukin regulation of extracellular matrix and Oncostatin M pathways and downregulation of PPAR signaling and fatty acid

metabolism pathways in VAT samples. Overall, our results suggest that variation in TGF-β, Interleukin, Oncostatin M, and PPAR signaling pathways play a significant role in the predisposition to VAT adiposity and metabolic disease risk.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5051T: Differential Gene Expression Related to Opioid Use Behaviors in a Hybrid Rat Diversity Panel

Authors:

E. Aaron[1], C. Wood[2], M. Ehringer[3], R. Bachtell[3], L. Saba[2], K. Kechris[2]; [1]Univ. of Colorado, Denver, CO, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Univ Colorado, Boulder, CO

Abstract:

The opioid epidemic in the United States continues to grow in scale as more cases of opioid use disorder (OUD) are observed. Investigating genetic components associated with risk of OUD would provide key insights into how to treat and prevent OUD. Using the Hybrid Rat Diversity Panel (HRDP) provides a means to study genetic background, environmental factors, and related behaviors using systems genetics approaches. Our study explored the effects of oxycodone exposure on RNA expression levels from the nucleus accumbens (NAc), a brain region associated with addiction and reward behaviors and how genetics may influence this oxycodone-induced transcriptome response. Rats were exposed to oxycodone via self-administered doses through lever pressing and an infusion pump system. Bulk RNAseq expression was collected from the NAc of 26 oxycodone- and 26 saline-exposed rats from 7 HRDP strains. Following data processing, the genes were tested for differential expression based on the oxycodone exposure group using DESeq2, controlling for sex and strain in the model. Nine genes were differentially expressed (FDR < 0.10) out of the 17,657 genes tested. These genes included *Cldn11*, *Mbp*, *Alas2*, *Plp1*, *Cfap20dc*, *LOC100362738*, *LOC108352949*, *mt-Rnr2*, and *Mal* in which some are involved in the central nervous system as constituents of myelin sheath. This was further confirmed by gene set over representation analysis on the top 100 genes in the differential expression analysis where significant biological pathways were related myelination and glial cell differentiation. Our results implicate the role of myelination in the transcriptional response to oxycodone exposure. Supported by NIDA (U01 DA051937) and R25HG012919.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5052T: Differentially expressed transcripts in HIV-TB co-infected children from East and Southern Africa are enriched for genes impacting neutrophil function ⭐

Authors:

E. Katagirya[1], B. Mlotshwa[2], S. Kyobe[1], G. Mardon[3], N. Hanchard[4], Collaborative African Genomics Network (CAfGEN) ofthe H3Africa Consortium; [1]Makerere Univ., Kampala, Uganda, [2]Univ. of Botswana, Gaborone, Botswana, [3]Baylor Col Med, Houston, TX, [4]Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Abstract:

HIV-TB coinfection presents a significant global health challenge, with the two pathogens, *Mycobacterium tuberculosis* and HIV, potentiating one another and accelerating the deterioration of immunological functions. TB presentation in HIV-infected individuals maybe atypical complicating diagnosis. There is also incomplete understanding of the disease pathology. Both pathogens have been independently shown to cause changes in alternative splicing in different cells including macrophages and $CD4^+$ T-cells, resulting in differential transcript expression. This study aimed to characterize differential transcript expression (DTE) to better understand HIV-TB disease pathology in children. DTE can detect changes missed by traditional differential gene expression (DGE) analysis and provide potential diagnostic and therapeutic targets. We performed RNAseq analysis on samples from Uganda (East Africa, n=59) and Botswana and eSwatini (Southern Africa, n=38). Standard RNAseq quality control was performed before alignment using Salmon. Downstream analyses were performed using DESeq2 for DTE analysis (Significance threshold of log2fold change of 2 and adjusted p-value 0.05). STRING and Metascape were used for pathway and Gene Ontology analyses.
Raw sample counts ranged from 14 to 50 million reads, adequate for detection of low expression transcripts. East African samples had 1.8 times more differentially expressed transcripts than Southern African samples (572 vs 316). Of these, 387 and 197 transcripts were upregulated in children with HIV-TB coinfection in East Africa and Southern Africa, respectively with an intersection of 13 transcripts. Pathway analysis for genes corresponding to the upregulated transcripts for both regions showed enrichment in pathways and processes important in immunity, with neutrophil degranulation pathway genes being most enriched for East Africa (p.adj = 1.5e-10) and fourth highest for Southern Africa (p.adj = 3.2e-7) using Reactome. Pathway analysis using the common transcripts from genes with strong protein-protein interaction - *OLFM4, LTF, PRTN3, MMP8*, and *CD177* - showed significant enrichment in neutrophil extravasation (FDR = 0.02).
Our results revealed distinct transcriptomic signatures between Southern and Eastern African

children with TB-HIV coinfection with upregulated genes enriched for critical immune processes, including neutrophil degranulation, cytokine signaling, and innate immune response, with shared transcripts highlighting genes important in neutrophil extravasation. Adjunctive studies to deconvolute cell types from the RNAseq data also showed an increase in neutrophil count.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5053T: Discovery of Genetic Factors for Atypical Antipsychotic-Induced BMI Change: Obesity Related to Antipsychotic Liability & Exposure (ORAcLE) Consortium

Authors:

S. Gupta[1], G. Chittoor[1], N. Josyula[1], S. Lee[2], M. Preuss[2], A. M. Lake[3], M. Niarchou[4], ORAcLE Consortium, L. Davis[5], R. Birnbaum[6], A. E. Locke[7], G. Nicol[8], A. E. Justice[1]; [1]Geisinger, Danville, PA, [2]Mount Sinai Sch. of Med., New York, NY, [3]Vanderbilt Univ. Sch. of Med., Nashville, TN, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Vanderbilt Univ Med Ctr., Nashville, TN, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Regeneron Genetics Ctr., Tarrytown, NY, [8]Washington Univ. Sch. of Med., St. Louis, MO

Abstract:

Psychiatrically ill individuals are at increased risk for obesity and exhibit poor cardiometabolic health due to intake of antipsychotic medications. Uncovering the genetic loci associated with antipsychotic-induced weight gain (AIWG) may assist in identifying people who are predisposed to weight gain and cardiometabolic dysregulation. Hence, we conducted a genome wide association (GWAS) meta-analysis to identify loci associated with BMI change and BMI increase in individuals on atypical antipsychotic medications utilizing data from 12,306 adults (12.89% non-European genetic ancestry) from three clinical biobanks, MyCode, BioME and BioVU. Time on atypical antipsychotic medications was included in linear mixed effects model as both fixed and random effects to determine BMI change (BMIch) and BMI gain (BMIg), stratified by sex, and genetic similarity group. The resultant slopes were regressed on baseline BMI, baseline age, age^2, follow-up duration, and genetic principal components. The residuals were inverse normal transformed. Positive residual slopes signify the BMIg cohort. Population and sex specific GWAS were carried out using linear regression allowing for heterogeneity of effect by follow-up duration. We subsequently meta-analyzed GWAS results both stratified by sex and combined across sexes. Meta-analyses yielded five distinct significant (P<5e-08) loci: rs113470946 (chr5, β=0.76, *P*=4.7e-9) near *ISL1* and rs986413633 (chr13, β=-0.51, *P*=1.02e-8) near *KLF12* from BMIch strata, rs149462699 (chr17, β=-1.18, *P*=8.77e-9) in *TIMP2*, rs115901190 (chr1,

β=1.22, *P*=3.99e-10) near *MEGF6* and rs1878584767 (chr13, β=0.93, *P*=3.52e-9) in an intergenic region from BMIg strata. Of these, rs986413633, rs149462699 and rs115901190 are not within a previously reported cross-sectional BMI locus (index variant +/- 500 kb). Only one gene is nearby rs113470946, *ISL1*, which is known to bind to the enhancer region of the insulin gene and may in turn govern insulin gene expression. Further, rare variants in *ISL1* have been associated with severe obesity in a candidate gene study, and this gene is a known MODY (maturity-onset diabetes of the young), adult type 2 diabetes, and insulin associated gene. *MEGF6* as a gene has been previously associated with lithium clearance, a mood stabilizer with weight gain side effects that is commonly co-prescribed with antipsychotics. Additional analyses are in progress to confirm these results in independent studies. Our findings support the claim that the magnitude of AIWG is partly attributed to genetic factors and further provide novel insights and plausible mechanisms to comprehend body mass change induced by antipsychotics.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5054T: Dissecting the genetic basis of clinical phenotype heterogeneity within IBD in more than 55,000 IBD patients ⭐

Authors:

D. Li, IIBDGC; Cedars-Sinai Med. Ctr., Los Angeles, CA

Abstract:

Background: More than 200 genetic variants associated with Inflammatory Bowel Disease (IBD) susceptibility are known. However few large-scale studies have examined the genetics basis of the heterogeneity of clinical phenotypes in IBD, including disease location, behavior, age of diagnosis and the distinction between Crohn's Disease (CD) and Ulcerative Colitis (UC). Our aim was to examine the genetic association with IBD clinical phenotypes in the large International IBD Genetics Consortium(IIBDGC) GWAS cohort. **Methods:** Details of the sample recruitment, genotyping and quality control(QC) have been described previously. In brief, genotyping was performed using the Illumina GSA platform in more than 20 cohorts across the world and imputation was performed using the TopMed server. Clinical phenotype data, including CD/UC status, disease location, behavior as well as IBD-assocaited surgery, were collected following the IIBDGC core phenotype standard based on widely used Montreal classification. Association analysis was performed using Regenie. All subjects gave informed consent after IRB approval. **Results:** Currently 55,198 IBD cases (21,351 UC and 33,847 CD) were included in the analyses. In the CDvsUC comparison, 38 independent regions were identified and 9 of those

have not been associated with IBD before (e.g., *ZBTB10*, OR = 1.11, P=1.74E-9; *BRBKB2*, OR = 1.12, P=4.08E-8). We also identified multiple novel association with disease location and behavior (e.g., *COL24A1* with stricturing CD, OR=1.29, P=6.52E-8; *MYOCD* with CD disease location, OR=1.66, P = 7.09E-8; *MSH5-SAPCD1* with extensive disease in UC, OR=0.63, P=6.39E-11). Variants at *NOD2* and HLA are strongly associated with multiple clinical phenotypes including CDvsUC, CD disease location and behavior as well as age of onset. Further curation of the clinical phenotypes is on-going to further boost sample size and post-GWAS analyses including finemapping, pathway analyses and integration with functional data are underway. **Conclusion:** Using a large IBD cohort, we observed strong and novel association with IBD clinical phenotypes. This largest GWAS analysis on IBD clinical phenotypes will help to better understand the genetic basis of heterogeneity in IBD clinical phenotypes and provide insights for biomarkers and novel drug targets.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5055T: Diverse genomic alterations in single neurons after chronic brain trauma

Authors:

M. Miller[1], G. Dong[2], C. Ma[3], S. Naik[1], G. McDonough[1], S. Mao[4], A. McKee[5], Y. Huang[6], L. Eunjung Alice[1], C. Walsh[7]; [1]Brigham and Womens Hosp., Harvard Med. Sch., Boston, MA, [2]Harvard Univ., Boston, MA, [3]Harvard Med. Sch./Boston Children s Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]Boston Univ., Boston, MA, [6]Boston Children s Hosp., Boston, MA, [7]Children's Hosp Boston, Boston, MA

Abstract:

Chronic traumatic encephalopathy (CTE) is a neurodegenerative disease associated with repetitive head impact (RHI) although little is known about its molecular pathogenesis. Previous studies of single neurons showed that private somatic mutations increase both during normal aging and in neurodegenerative disorders, and show diverse mutational patterns.
We applied two orthogonal single-nucleus whole-genome sequencing (snWGS) methods to neurons isolated from the prefrontal cortex of 15 individuals with CTE, and 4 individuals with RHI but no CTE diagnosis, and compared mutational rates and spectra with neurons from neurotypical controls and Alzheimer’s disease (AD).
We found a modest but significant elevation of somatic double-stranded single-nucleotide variants (SNVs) that resembles a pattern previously reported in AD. In addition, we found a strikingly large burden of small insertions and deletions (indels) and used duplex sequencing to

show that these indels are mainly single-stranded, and again found a similar phenomenon in neurons from AD brain.
Our results suggest that neurons in CTE brain are exposed to stereotyped mutational processes, and that these processes are shared between AD and CTE suggesting potentially shared pathogenic mechanisms. Furthermore, the absence of similar changes in RHI neurons without CTE suggests that the development of CTE entails a mechanism beyond that caused by RHI alone.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5056T: Downregulated transcriptome in dopaminergic neurons from the ventral midbrain is enriched for schizophrenia risk loci.

Authors:

S. Singh[1], M. Iskhakova[1], T. Lambert[1], S. Marenco[2], P. Auluck[3], G. E. Hoffman[1], P. Roussos[1], K. Girdhar[1], S. Akbarian[1]; [1]Icahn Sch. of Med., Mount Sinai, New York City, NY, [2]Nathan Kline Inst. for Psychiatric Res., Orangeburg, NY, [3]Human Brain Collection Core, Natl. Inst. of Mental Hlth., Bethesda, Bethesda, MD

Abstract:

Dopamine neurons have been implicated in the dysregulation of neurotransmitter systems in individuals with schizophrenia (SCZ) and Bipolar Disorder (BD) via dopamine level mediation. While dopamine is primarily produced in the midbrain regions of the substantia nigra and ventral tegmental area (VTA), there is a dearth of research investigating changes in the genetic and epigenetic landscapes of these regions in SCZ and BD patients. To address this knowledge gap, we have developed a resource comprising cell-specific 111 RNA-seq and 34 HiC libraries sourced from midbrain tissue of SCZ, BD, and control patients. Libraries were prepared using dopaminergic and non-neuronal nuclei populations isolated through fluorescence-activated nuclei sorting (FANS) techniques using Nurr1+/NeuN+ immuno-tagging. Differential transcriptomic expression in dopaminergic neurons revealed 340 differentially regulated genes between SCZ and control donors, while no significant gene changes were observed in non-neuronal cells. Furthermore, downregulation in identified SCZ-specific genes showed enrichment for SCZ risk genes conferring heritable risk for midbrain dopaminergic neurons. Functional pathway analysis of these genes showed disruption in synaptic plasticity and neurotransmitter uptake regulation, highlighting the underlying molecular mechanisms in SCZ. Furthermore, by utilizing our cell-specific HiC data we investigated the spatial organization of the differentially regulated genes within the 3D chromatin architecture. We found that the

downregulated genes situated within dopaminergic neuronal TADs within compartment-A exhibit a 3-fold enrichment for SCZ risk genes compared to the entire downregulated genes. Overall, our research emphasizes the importance of midbrain dopaminergic transcriptome changes in SCZ, highlighting specific functional pathways. Supported by NIDA grant U01DA048279.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5057T: Dysregulated plasma miRNAs are associated with cerebrospinal fluid and neuroimaging biomarkers in Alzheimer's disease continuum

Authors:

S. Liu[1], T. Park[1], D. M. Krüger[2], T. Pena-Centeno[2], S. Burkhardt[2], A-L. Schutz[2], Y-N. Huang[1], T. Rosewood[1], S. Chaudhuri[1], M. cho[3], S. L. Risacher[1], Y. Wan[4], L. M. Shaw[4], F. Sananbebesi[2], A. Brodsky[5], H. Lin[6], A. Krunic[7], J. Blusztajn[7], A. Saykin[8], I. Delalle[7], A. Fischer[2], K. Nho[1]; [1]Indiana Univ., Indianapolis, IN, [2]German Ctr. for Neurodegenerative Diseases (DZNE), Göttingen, Germany, [3]Indiana Univ., Sungkyunkwan Univ., indianapolis, IN, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Brown Univ., Providence, RI, [6]Univ. of Massachusetts Chan Med. Sch., Worcester, MA, [7]Boston Univ., Boston, MA, [8]Indiana Univ. Sch. of Med., Indianapolis, IN

Abstract:

Background: Alzheimer's disease (AD) is the most common type of dementia. Effective diagnostic and treatment approaches for AD remain challenging despite decades of research. MicroRNAs (miRNAs) play important roles in gene expression regulation and AD pathogenesis. We aimed to identify miRNAs in plasma as a potential blood-based biomarker for AD related to central biomarkers.
Methods: We performed small RNA-Sequencing on plasma samples (N=803) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort and carried out thorough pre-processing steps including adapter trimming, read quality assessment, alignment, count generation, and batch effect evaluation. We investigated the association between plasma miRNAs and central AD biomarkers based on cerebrospinal fluid and neuroimaging endophenotypes from the Alzheimer's Disease Neuroimaging Initiative (ADNI; N=803): Amyloid, Tau, and Neurodegeneration (A/T/N). Differentially expressed miRNAs and their targets were identified, followed by pathway enrichment analysis. Machine learning approaches were applied to investigate the role of miRNAs as blood biomarkers.
Results: We identified nine, two, and eight miRNAs significantly associated with A/T/N positivity,

respectively. We identified 271 genes targeted by Amyloid-related miRNAs showing enriched pathways including ESR-mediated signaling. Additionally, 220 genes targeted by Neurodegeneration-related miRNAs showed enrichment in pathways including the IGF1 pathway. The classification performance of demographic information for A/T/N positivity was increased up to 9% with the inclusion of miRNAs.
Conclusion: Our results showed plasma miRNAs were associated with central A/T/N biomarkers for AD and improved their classification, highlighting their potential as AD blood biomarkers.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5058T: Effects of gene dosage on cognitive ability: A function-based association study across brain and non-brain processes

Authors:

G. Huguet[1], T. Renne[2], J. Mollon[3], L. Schultz[4], M. Jean-Louis[5], Z. Saci[5], Z. Pausova[5], T. Paus[5], L. Almasy[6], D. Glahn[7], S. Jacquemont[8]; [1]CHU Ste justine, Montréal, QC, Canada, [2]Université de Montréal, Montreal, QC, Canada, [3]Boston Children's Hosp., Boston, MA, [4]Childrens Hosp. of Philadelphia, Philadelphia, PA, [5]CHU Sainte-Justine, Montreal, QC, Canada, [6]Children`s Hosp. of Philadelphia, Philadelphia, PA, [7]Boston Children s Hosp. & Harvard, Boston, MA, [8]CHU Sainte justine Res. Ctr., Montreal, QC, Canada

Abstract:

Neurodevelopmental disorders (NDDs) are common (11-23%) in the general population and around 10-15% of NDD cases involve pathogenic Copy Number Variants (CNVs). Importantly, these same CNVs systematically affect cognitive ability in individuals from the general population without a diagnosis of NDD. Studies have focussed on a small set of recurrent CNVs. As a result, the full range of gene-dosage-sensitive biological processes linked to cognitive ability remains unknown, as the research has been limited to small parts of CNVs. This study evaluates how different gene sets, both related and unrelated to brain function, influence cognitive ability using the burden model, which includes all rare CNVs, not just the recurrent ones.We analyzed 258,292 individuals from 6 general populations, with genotyping and cognitive assessments. We studied all genetic CNVs >50Kb, called with PennCNV and QuantiSNP. Cognitive ability of each individual was measured using different assessments. The protein-coding genes were assigned to gene-sets (215 brain regions, 7 adult and 16 fetal brain cell types, and 36 whole-body tissues) based on overexpression. We performed linear regression to estimate the effect size of each gene-set on cognitive ability.
Among the 274 gene-sets associated with cognitive ability, only 11% showed significant effects

for both deletions and duplication. Accordingly, we systematically observed negative correlations between deletion and duplication effect sizes across all levels of biological observations. We quantified the preferential effects of deletions versus duplication using tagDS, a new normalized metric. Cognitive ability was preferentially affected by cortical, presynaptic, and negative-regulation gene-sets when duplicated. In contrast, preferential effects of deletions were observed for subcortical, post-synaptic, and positive-regulation gene-sets. A large proportion of gene-sets assigned to non-brain organs were associated with cognitive ability due to low tissue specificity genes, which were associated with higher sensitivity to haploinsufficiency. Overall, most biological functions associated with cognitive ability are divided into those sensitive to either deletion or duplications. In addition, genes most sensitive to haploinsufficiency are also those expressed across a broad spectrum of brain and non-brain tissues.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5059T: Elevated risk of schizophrenia in 22q11.2 Deletion Syndrome with genome-wide tandem repeat expansions

Authors:

M. Cheng[1], Y. Yin[2], W. Engchuan[2], T. Heung[3], International 22q11.2 Brain Behavior Consortium, A. Bassett[1], R. Yuen[2]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada

Abstract:

22q11.2 microdeletion is one of the most recurrent genetic risk factors for schizophrenia, but only ~25% of individuals with 22q11.2 deletion syndrome (22q11.2DS) develop schizophrenia. This may be attributed to the additional genetic factors that modify the outcomes. Known genetic modifiers, such as common risk SNPs (in aggregate as polygenic risk score; PRS) and rare copy number variants (CNVs), only account for part of the phenotypic heterogeneity, suggesting that other uncharacterized factors may be involved. We analyzed genome sequencing data of 88 de novo 22q11.2DS trio Canadian families (i.e., adult proband with typical 22q11.2 deletion and both unaffected parents), and 260 other unrelated individuals with typical 22q11.2 deletions from the International 22q11.2 Brain Behavior Consortium (IBBC) to assess the effect of genome-wide variants on developing schizophrenia. All individuals were of European background. Association analysis using a multiple regression model was performed on several types of genetic factors including PRS, rare (<0.1% population frequency) SNVs, rare CNVs, and tandem repeat expansions (TRE). We identified a significant association between genome-wide

rare TREs and schizophrenia risk in 22q11.2DS (Canadian cohort: OR = 2.47, p = 0.007; IBBC cohort: OR = 1.74, p = 0.04). The effect of rare TREs on schizophrenia risk, collectively, is on par with that conferred by PRS, and independent of the individuals' cognitive abilities. The identified schizophrenia-associated TREs are mostly located within intronic regions of genes that are FMRP targets (OR = 4.7, p = 0.001) and known schizophrenia risk genes identified by previous genome-wide association studies (OR = 9.3, p = 0.04). The TREs identified in 22q11.2DS with schizophrenia significantly overlap with those from our previous study of a community-based schizophrenia cohort without 22q11.2DS (p < 1.4 x 10E-6). Also, we found a similar significant effect of rare genic TRE burden on schizophrenia risk in an independent de novo 22q11.2 microdeletion trio Canadian family cohort (n=34) of non-European background (OR = 9.5, p = 0.02). Our study suggests that TRE is one of the major genetic modifiers contributing to schizophrenia risk in 22q11.2DS. The findings highlight the importance of expanding the analysis of genetic risk factors beyond the protein-coding sequences. Our approach provides a more comprehensive assessment of genetic modifying effects in genetic disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5060T: Enhancing cardiovascular disease risk prediction using polygenic risk scores and traditional risk algorithms in the UK Biobank

Authors:

M. Kals[1,2], L. Holst[2], E-M. Lotman[2,3], P. Palta[1,2]; [1]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Activate Hlth., Tallinn, Estonia, [3]Ctr. of Cardiology, North Estonia Med. Ctr., Tallinn, Estonia

Abstract:

Cardiovascular disease (CVD) is caused by a combination of environmental and genetic risk factors. Polygenic risk scores (PRS) have emerged as a powerful tool in predicting individual susceptibility to various diseases and there is a growing interest in the integration of PRS in guideline-recommended clinical risk prediction models. This study explores the potential of PRS alone and in combination with traditional risk algorithms, such as SCORE2, to enhance the prediction of CVD risk in the UK Biobank database.
Current CVD risk prediction models, like SCORE2, incorporate clinical and demographic factors, including age, sex, systolic blood pressure, total- and HDL-cholesterol, smoking status, and diabetes. While these models are effective, they do not account for the genetic predisposition that significantly impacts CVD risk. By integrating PRS with existing models, it is possible to provide a more comprehensive risk assessment.

In the study, we included UK Biobank participants of European origin aged 40-69 who had no history of CVD and for whom SCORE2 risk could be calculated and genotype information was available for polygenic risks (n=326,131). Participants were followed for more than one year until the occurrence of a CVD event, death, or the end of the study (whichever occurred first). During the follow-up period, 45,304 CVD incidence cases were recorded (13.9% of all participants). There were more CVD events in men (25,821 out of 145,935, 17.7%) than in women (19,483 CVD events out of 180,196, 10.8%).
For all included participants, SCORE2 risk estimates and PRS for seven traits related to the outcome or clinical risk factors of CVD were calculated - PRS for CVD, coronary artery disease, ischemic stroke, hypertension, type 2 diabetes, body mass index, and HDL cholesterol. PRS for CVD demonstrated the highest discrimination of CVD risk (C-index 0.547 in the Cox proportional-hazard model). We then assessed the predictive performance of PRS in combination with SCORE2. The results indicated that PRS significantly enhances the prediction accuracy of CVD risk (per SD hazard ratio was 1.20, 95% CI = 1.19-1.21). When PRS for CVD was combined with SCORE2, the model's discrimination C-index increased from 0.683 to 0.689.
In conclusion, integrating PRS with established risk algorithms like SCORE2 enhances the precision of CVD risk prediction. This combination leverages genetic information alongside traditional risk factors, offering a more robust tool for identifying high-risk individuals. Future work should focus on validating these findings in diverse populations and integrating PRS into clinical practice to optimize CVD prevention and management.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5061T: Enhancing Early Alzheimer's Disease Detection Using Machine Learning and SHAP-Enhanced Non-Invasive Biomarkers

Authors:

H. Javedani Sadaei; Scripps Res., La Jolla, CA

Abstract:

Predicting the onset of Alzheimer's Disease (AD) early is crucial for effective clinical management and improving patient outcomes. Traditional diagnostic methods are invasive and costly, emphasizing the urgent need for scalable, non-invasive diagnostic models. Our study leverages machine learning (ML) to develop a predictive model for AD, utilizing key non-invasive biomarkers collected from plasma and cerebrospinal fluid (CSF). These biomarkers include phosphorylated tau (p-tau) and amyloid-beta , which are instrumental in understanding AD pathology. Phosphorylated tau is associated with neurofibrillary tangles, while amyloid-beta is

involved in plaque formation, both of which are hallmark features of AD.Our model integrates ensemble learning techniques and utilizes SHAP (SHapley Additive exPlanations) values to interpret the impact of individual predictors, including the crucial roles of p-tau and Aβ levels. This approach enhances our understanding of the factors driving AD progression. By incorporating Polygenic Risk Scores (PRS) alongside these biomarkers, we capture both genetic and non-genetic risk factors, facilitating the early diagnosis and prediction of disease progression across various stages in older adults. The data for this study is sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which supports our approach by providing a rich dataset.Our advanced ML model demonstrates significant potential in improving the accuracy of AD predictions, emphasizing the transformative capability of non-invasive, PRS-enhanced methods in detecting and managing Alzheimer's Disease. This comprehensive approach offers a deeper insight into the biological processes underlying AD and provides a robust tool for early detection and ongoing monitoring of the disease.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5062T: Epistasis and Gene-Gene Interaction Network to Identify Core Genes of Complex Traits

Authors:

L. Evans; Univ. of Colorado, Boulder, CO

Abstract:

One largely unexplored reservoir of genetic effects on complex traits is interactions among genes (epistasis), the characterization of which could identify key and novel molecular pathways, networks, and cell types in which genetic risk of complex disease is manifest. We have developed a method to efficiently and exhaustively statistically test all pairwise gene-gene interaction associations with complex traits in large-scale biobank data using predicted gene expression and predicted gene loss of function variation, followed by gene set enrichment analyses to infer biological function and context in which such interactions influence trait variation. We have applied our published approach to multiple traits, but here present new, unpublished, and trans-ancestral analyses of aging phenotypes, including frailty and Alzheimer's Disease and related dementias. We find several core genes with many gene-gene interactions, and we highlight specific cell types in which those interactions appear most important.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5063T: ESCALATOR: An efficient bioinformatics pipeline to harmonize and apply external weights for PRS calculation in large-scale datasets ⭐

Authors:

M. Lin, M. Fisher, C. R. Gignoux; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Calculating polygenic risk scores (PRSs) using existing weights for testing, replicating, or evaluating independently developed PRS models is becoming a prevalent analysis step in biomedical research. While the process appears straightforward, it comes with challenges attributed to data heterogeneity between the test and the external training dataset, necessitating systematic harmonization of external weights before score calculation. These disparities may include unmatched strand information at the same variant, non-biallelic codes at overlapping genomic positions, and different genomic builds, all of which can arise from non-converged dataset generation methods and handling processes between datasets. To address these issues, we present ESCALATOR, a convenient pipeline that efficiently standardizes, harmonizes, and applies external PRS weight files to a test dataset of interest, and utilizes external PLINK software to complete the generation of score calculations of samples along with useful summaries. ESCALATOR was tested on the biobank of Colorado Center for Personalized Medicine for 73,346 samples with imputed dosages stored in the PLINK 2 binary format. It has proven cost-effective in terms of computational resources, requiring approximately 0.5 CPU hours and 826MB maximal memory consumption for a model of ~1M variants, typical of scores developed from popular Bayesian shrinkage methods such as LDpred2, and about 1 CPU hour and 905MB maximal memory consumption for a model of ~10M variants, among the largest published models we have encountered. By offering a containerized version of this pipeline compatible with widely used Docker and Singularity platforms, we hope ESCALATOR can simplify and expedite the risk score calculation analysis, especially for large test datasets residing in most computational cluster environments, including the ongoing implementation into PRIMED consortium's ANVIL platform for broader applications.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5064T: Evaluating Approaches for Multi-Ancestry Polygenic Risk Scores Development in Alzheimer's Among Hispanic Populations

Authors:

Y. Xu[1], D. Reyes-Dumeyer[1], T. GUNASEKARAN[1], M. Qiao[1], A. Piriz[1], R. Lantigua[1], M. Medrano[2], D. Rivera[3], D. Sanchez[1], A. Brickman[1], J. Manly[1], B. Vardarajan[1], R. Mayeux[1]; [1]Columbia Univ., New York, NY, [2]Pontificia Univ. Catolica Madre y Maestra, Santiago, Dominican Republic, [3]Univ. Pedro Henríquez Urena, Santo Domingo, Dominican Republic

Abstract:

Introduction: Polygenic risk scores (PRS) are powerful tools for summarizing genetic risk across the genome, with the potential to predict disease outcomes and guide treatment decisions. However, Alzheimer's disease (AD) PRS have been significantly underutilized and underdeveloped in Hispanic populations. A comprehensive evaluation of approaches for developing PRS among Hispanics is crucial for ensuring fair and effective implementation in clinical practice, ultimately benefiting this community. **Methods:** Leveraging genome-wide association studies of AD in European and African American populations, we have constructed ancestry-specific PRS and multi-ancestry PRS, independent of *APOE*, based on three different methods: PRSice-multi, PRScsx, and BridgePRS. We evaluated the PRS in 2,149 cases and 4,649 controls from two Hispanic cohorts: the Estudio Familiar de Influencia Genética en Alzheimer and the Washington Heights-Hamilton Heights-Inwood Columbia Aging Project. The performance metrics of each PRS were evaluated alone and with other predictors using the area under the curve (AUC) and logistic regression. Additional covariates included age, sex, *APOE* ε4, and the first 10 principal components. **Results:** Among Hispanics, the multi-ancestry PRS outperformed all ancestry-specific PRS in predicting clinical AD. The full model of PRS derived by PRSice-multi, PRScsx, and BridgePRS achieved an AUC of ~73%, which is similar to the AUC reported by a previous paper that constructed PRS based on a Hispanic GWAS. For PRS derived by all three methods, higher PRS was significantly associated with clinical AD ($OR_{PRSice\text{-}multi}$ = 1.19, 95% CI = 1.08-1.31; OR_{PRScsx} = 1.24, 95% CI = 1.14-1.34; $OR_{BridgePRS}$ = 1.22, 95% CI = 1.12-1.32). The association for PRS derived from all three methods was stronger among *APOE* ε4 carriers compared to non-carriers. **Conclusion:** Among Hispanics, multi-ancestry PRS based on available GWAS improve the ability to predict clinical AD more than European or African American derived ancestry-specific PRS. Our efforts represent the first step towards the implementation of multi-ancestry PRS into risk prediction for Hispanics.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5065T: Evaluating reported ethnicity, genetic ancestry and three-way local-ancestry admixture with preeclampsia in GenPE, a multi-center case-control study of Colombian women.

Authors:

N. Vaitinadin[1], K. Gray[2], C. Dupont[1], M. Maher[3], R. Saxena[4], J. Casas[5,6], C. Colmenares[7], N. SERRANO[8], A. Giri[9]; [1]Vanderbilt Univ Med. Ctr., Nashville, TN, [2]Univ. of Washington, Seattle, WA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Mass Gen. Hosp., Harvard Med. Sch., Boston, MA, [5]VA Hlth.care System, Boston, MA, [6]Brigham and Women's Hosp., Boston, MA, [7]Fundación Cardiovascular de Colombia, Bucaramanga, Colombia, [8]FUNDACIÓN CARDIOVASCULAR DE COLOMBIA FCV, Bucaramanga, Colombia, [9]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

The role of genetic ancestry in hypertensive pregnancy disorders in Latin-American women is poorly understood. Using data from a multi-center case-control study (GenPE) of preeclampsia in young Colombian women (median age 19) of predominantly low socioeconomic status (2364 controls and 1811 cases), who identify as Afro-Caribbean, White Hispanic, Amerindian, and Mixed ethnicity, we evaluated associations between 1) reported ethnicity, and 2) empirically estimated genetic ancestry, with preeclampsia. We first performed three-way admixture mapping using European (EUR), African (AFR) and Amerindian (AMR) ancestry references from the Human Genome Diversity Project using the FLARE software to estimate local and global ancestry in GenPE samples. Statistical significance threshold, for three-way local ancestry analyses, was empirically estimated using STEAM ($P= 3.45x10^{-6}$). In multivariable logistic regression models for reported ethnicity, Afro-Caribbeans were 33% more likely to have preeclampsia (OR = 1.33; P= 0.02) than white-Hispanic women. In models evaluating empirically estimated global genetic ancestry, AFR was positively associated (OR per 10% increase in ancestry = 1.05; P = 0.002), while AMR (OR = 0.91; P = 0.035) and EUR (OR = 0.95; P = 0.009) were inversely associated with preeclampsia. Additionally, adjusting for reported ethnicity in models evaluating global genetic ancestry and preeclampsia changed estimates only marginally for AFR (OR = 1.04; P = 0.025) and EUR (OR = 0.92; P = 0.009). Evaluation of genetic ancestry and preeclampsia in a subset of women who reported Afro Caribbean ethnicity, showed stronger estimates for all global ancestry estimates: AFR (OR = 1.11; P = 0.013, EUR (OR = 0.82; P = 0.026), and AMR (OR = 0.83; P = 0.01). Association analyses with AFR local ancestry identified three loci associated with preeclampsia. At the top locus at chromosome 11, rs2021740 (a smooth muscle enhancer in *OTOG1* and near *MYOD1*), each additional allele of African origin was associated with 27% increased odds of preeclampsia (OR = 1.27; $P = 1.13x10^{-7}$). The A-allele for this variant is found in greater frequency in African reference populations (22%) than in

Europeans (5%). Subgroup analyses with HELLP syndrome (279 cases and 2364 controls) shows intriguingly opposite findings with increased risk for global AMR and EUR ancestry and decreased risk for AFR ancestry. Using a genetically heterogenous Hispanic population, we show genetic ancestry is associated with preeclampsia independent of reported ethnicity and further demonstrate the power of admixture mapping to identify a candidate locus for preeclampsia.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5066T: Evidence for Y chromosome dosage effects on autism risk

Authors:

A. Berry[1], B. M. Finucane[2], S. M. Myers[1], L. K. Walsh[1], J. M. Seibert[1], C. L. Martin[1], D. H. Ledbetter[3], M. T. Oetjens[4]; [1]Geisinger, Danville, PA, [2]Geisinger Hlth.System, Lewisburg, PA, [3]U. Florida Coll. of Med.-Jacksonville, Fernandina Beach, FL, [4]Geisinger, Lewisburg, PA

Abstract:

A female protective effect has long been postulated as the primary explanation for the four-fold increase of autism spectrum disorder (ASD) diagnoses in males versus females. Investigations of the female protective effect in genetic and epidemiological studies have shown that this hypothesis is unable to explain the large difference in ASD prevalence between the sexes. Furthermore, it is not well understood if sex chromosome dosage contributes to the higher prevalence of ASD among males. To address this knowledge gap, we examined sex chromosome aneuploidy in a large ASD case-control cohort to evaluate the relationship between X and Y chromosome dosage and ASD risk. From these data, we modeled three relationships between sex chromosome dosage and ASD risk: the extra Y effect, the extra X effect, and sex chromosome haploinsufficiency. We found that an extra X chromosome was not associated with increased ASD risk relative to 46,XX or 46,XY (adjusted $P > 0.05$), however 47,XXY was significantly more likely to be associated with ASD compared to 46,XX (odds ratio (OR), 4.6; 95% confidence interval (CI), 3.1-6.6; adjusted $p < 0.001$), and 47,XYY was significantly more likely to be associated with ASD compared to 46,XY (OR, 2.4; 95% CI, 1.6-3.5; adjusted $p < 0.001$). Meta-analyses with data from the Danish iPSYCH cohort supported the finding that the extra Y effect increased ASD risk significantly more than the extra X effect. In meta-analyses, individuals with 47,XYY had increased ASD risk compared to individuals with 47,XXY (OR, 1.9; 95% CI, 1.2-2.9; adjusted $P < 0.01$), and individuals with 47,XXY had increased ASD risk compared to 47,XXX (OR, 2.4; 95% CI, 1.3-4.4; adjusted $P < 0.01$). Among females, we observed a large association between 45,X and ASD (OR, 6.4; 95% CI, 3.2-12.6; adjusted $P < 0.001$), confirming sex chromosome haploinsufficiency as a strong ASD risk factor. This study presents evidence that

one or more genes on the Y chromosome increase ASD risk among males and may contribute to the well-documented sex ratio difference in ASD prevalence.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5067T: Examining the link between 179 lipid species and 8 disease endpoints using genetic predictors ⭐

Authors:

L. Ottensmann[1], R. Tabassum[1], S. Ruotsalainen[1], M. Gerl[2], C. Klose[2], D. McCartney[3], E. Widen[1], FinnGen, K. Simons[2], S. Ripatti[1,4,5], V. Vitart[6], C. Hayward[6], M. Pirinen[1,4,7]; [1]Inst. for Molecular Med. Finland, University of Helsinki, Finland, [2]Lipotype GmbH, Dresden, Germany, [3]Ctr. for Genomic and Experimental Med., Inst. of Genetics and Cancer, University of Edinburgh, United Kingdom, [4]Dept. of Publ. Hlth., Clinicum, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [5]Broad Inst. of the Massachusetts Inst. of Technology and Harvard Univ., Cambridge, MA, [6]Med. Res. Council Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh, Western Gen. Hosp., Edinburgh, United Kingdom, [7]Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland

Abstract:

Background/Objectives: Genome-wide association studies (GWAS) of lipid species have identified many lipid loci associated with diseases. However, the influence of lipid species on disease risk remains poorly understood. We aim to build genetic predictors of 179 lipid species measured by mass spectrometry in 7,174 Finnish individuals from the GeneRISK cohort and assess the impact of these lipid species on 8 disease endpoints.**Methods:** We consider two approaches for building lipid species' genetic predictors: Utilizing loci with high impact on lipid species levels and utilizing genome-wide polygenic risk scores (PRS). We obtained regional heritability estimates in GeneRISK by fine-mapping loci associated with 179 lipid species. We derived PRS for 131 lipid species measured in GeneRISK with PRS-CS and PLINKv1.9 and validated the scores in an independent FinRISK cohort (N=1,032). To examine the link between lipid species and diseases, we performed association analyses for high-impact variants in FinnGen (N=500,348), UK Biobank White British participants (N=408,959), and Generation Scotland (N=20,032). We applied univariable Mendelian randomization (MR) analysis of lipid species and the disease endpoints and utilized multivariable MR (MVMR) analyses to assess if lipid species impact disease risk independently of standard lipids such as HDL-C and ApoB.**Results:** We identified 11 high-impact loci (*ABCG8*, *AL161670.1*, *APOE*, *CERS4*, *FADS2*, *GLTPD2*, *LINC01722*, *LIPC*, *NTAN1*, *TMC4*, *ZPR1*) with

regional heritability values >2%. PRS explained >4% of the variance for 25 lipid species. Excluding high-impact locus regions from the PRS calculation reduced the variance explained to 2.1%. Variants of high-impact loci are associated (P<6.25e-9) with Alzheimer's disease, Cholelithiasis, Type 2 diabetes, Ischemic heart disease (IHD), Statin medication, Pure hypercholesterolemia, and Age-related macular degeneration (AMD). Univariable MR analysis linked 4 species including 2 Phosphatidylcholines (PC) to Cholelithiasis, 3 Phosphatidylethanolamines (PE) to AMD, and 5 species including 3 Triacyglycerols to IHD. The effect of PEs on AMD remained significant after conditioning on HDL-C and was driven by *LIPC* variants. MVMR analysis showed a significant effect of ApoB on higher IHD risk independently of lipid species.**Conclusion:** We identified 11 loci with a high impact on the heritability of the plasma lipidome. Variants within these loci are associated with the 8 disease endpoints. Univariable MR suggested a causal effect of higher PC levels on lower Cholelithiasis risk. *LIPC* variants increase PE levels and lower AMD risk. MVMR supports a causal role of ApoB in IHD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5068T: Exploration of bulk TCR repertoire in Graves' disease patients to evaluate T cell receptor diversities and signature sequences associated with HLA alleles and disease phases

Authors:

Y-H. Huang[1], P-L. Chen[1,2,3], S-K. Lai[2,3], C-M. Hsieh[4]; [1]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ., Taipei, Taiwan, [2]Genome and Systems Biology Degree Program, Academia Sinica and Natl. Taiwan Univ., Taipei, Taiwan, [3]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [4]Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan

Abstract:

Graves' disease (GD) is the leading cause of hyperthyroidism resulting from the production of autoantibodies that stimulate the thyroid gland. The global incidence of Graves' disease is around 0.02-0.04% per year and is more common in females. Early twin and family studies have suggested that the combination of environmental and genetic factors may contribute to the development of Graves' disease. Human leukocyte antigen (HLA) genes play a crucial role in regulating immune responses by presenting antigens to T cells. Although several GD-associated HLA alleles have been identified (e.g. HLA-DRB1*03 in Caucasians, HLA-DPB1*05:01 and HLA-B*46:01 in Asian populations), further functional studies are needed to explore the mechanisms underlying Graves' disease. In this study, we aim to investigate the T cell receptor (TCR)

repertoire, the collective set of TCR in an individual, in GD patients to elucidate potential alterations in TCR associated with HLA risk alleles and different disease phases. Subjects with and without diagnosed Graves' disease at National Taiwan University Hospital (NTUH) were enrolled in this study. RNA extracted from PBMCs collected from participants were reverse-transcribed into cDNA library using TCR primers targeting the V, D, and J gene regions. High-throughput NGS was conducted to sequence the V(D)J junction of TCR. Subsequently, the computational algorithm TRUST4 was applied to reconstruct the rearranged complementarity-determining region 3 (CDR3) of TCR repertoires. The immune repertoire analysis R package (Immunarch, v0.9.1) was utilized to assess the repertoire composition and diversity. We observed differences in the TCR repertoire profiles between Graves' disease (GD) patients and healthy individuals. Additionally, the composition of TCR repertoire varied between GD patients carrying GD-related HLA alleles and those without these alleles. Furthermore, our results indicated changes in TCR repertoire diversity between the active and stable phases of Graves' disease patients. These findings suggested a potential association between TCR profiles and both HLA genotypes and disease phases in GD. In conclusion, this study shed light on the development of GD and the role of HLA alleles in disease susceptibility, providing valuable insights for future research with larger sample sizes to elucidate the underlying pathophysiological mechanisms.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5069T: Explore genetic modifiers of disease progression in amyotrophic lateral sclerosis

Authors:

Y. Huang[1], C-Y. Chen[1], R. Zwamborn[2], W. van Rheenen[2], J. Veldink[2], J. Hurt[1], M. Harms[3], K. G. Guruharsha[1], E. Tsai[1]; [1]Biogen, Cambridge, MA, [2]Univ. Med. Ctr. Utrecht, Utrecht Univ., Utrecht, Netherlands, [3]Columbia Univ., New York, NY

Abstract:

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease with high unmet need for disease-modifying therapies. Genetic factors contribute significantly to ALS susceptibility. Currently, familial ALS is estimated to encompass 5-10% of the total ALS population — of these, ~ 76% have a known genetic component. For sporadic ALS, which represents 90-95% of the total ALS population, ~25% have a known genetic component. (Akçimen 2023; Byrne 2011; Zou 2017) The most recent genome-wide association study (Rheenen 2021) identified fifteen loci associated with ALS risk, but many do not influence

disease progression. Our study aims to uncover genetic modifiers of disease progression in a population-based ALS cohort with whole-genome sequencing and longitudinal clinical data. By leveraging data from the Genomic Translation for ALS Care Study, we measured ALS disease progression by the slope of the Revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) score and forced vital capacity (FVC) across >1 timepoints in 1,302 participants. After genetic data quality control, we first explored genetic associations of known ALS mutations with ALSFRS-R and FVC slopes in 1,010 and 825 participants, respectively. We observed slower disease progression for SOD1 D91A carriers (N=4, p=0.05) while faster disease progression for SOD1 A4V variant carriers (N=5, p=0.02) and C9orf72 repeat expansion carriers (N=98, p=4e-8). In a genome-wide search for genetic modifiers of disease progression, we identified several emerging genetic loci (p<1e-5) to be associated with disease progression reflected in the slope measure of ALSFRS-R score as well as FVC. However, replication analysis in additional cohorts at smaller scale shows inconsistent results and strong heterogeneity. Intriguingly, variant-to-gene mapping identified genes in disease relevant pathways including sphingolipid metabolism and axonal guidance associated with progression in ALS. We plan to further increase our sample size to validate the genetic findings and better understand the genes and pathways that contribute to disease progression in ALS and their potential as disease-modifying therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5070T: Exploring Genetic Determinants of Type 2 Diabetes: The Causal Impact of *SREBF1* Revealed by Two-Sample Mendelian Randomization

Authors:

A. Jamshidi[1], Y. Zhang[2], K. Ashok[3], M. Sevilla Gonzalez[4], L. Ching-Ti[2,5], J. Mercader[3,6,7,8], J. Meigs[3,6], J. Dupuis[1,2], A. Manning[3,4,6]; [1]Dept. of Epidemiology, Biostatistics and Occupational Hlth., McGill Univ., Montreal, QC, Canada, [2]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [3]Programs in Metabolism and Med. & Population Genetic, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Clinical and Translational Epidemiology Unit, Massachusetts Gen. Hosp., Boston, MA, [5]Natl. Heart, Lung, and Blood Inst. and Boston Univ.'s Framingham Heart Study, Framingham, MA, [6]Dept. of Med., Harvard Med. Sch., Boston, MA, [7]Diabetes Unit, Massachusetts Gen. Hosp., Boston, MA, [8]Ctr. for Genomic Med. and Diabetes Unit, Endocrine Div., Dept. of Med., Massachusetts Gen. Hosp., Boston, MA

Abstract:

Introduction: Type 2 Diabetes (T2D) is a complex metabolic disorder influenced by genetic and environmental factors. Genetic determinants underlying T2D susceptibility can be used to understand disease mechanisms and develop targeted therapies. This study employs a Two-Sample Mendelian Randomization (MR) approach to elucidate the causal relationships between genetic expression in human tissues and T2D risk.
Methods: We utilized genetic association results from GTEx (Genotype-Tissue Expression), TIGER (Translational human pancreatic Islet Genotype tissue-Expression), and DIAMANTE-EUR (Diabetes Meta-Analysis of Trans-Ethnic association studies- European population). We used lassosum tool to select SNPs to use as instrumental variables (IVs) after filtering on Minor Allele Frequency and before linkage disequilibrium (LD) clumping. Three analyses were performed, varying the sources of exposure data, but all using the DIAMANTE-EUR genetic association results to link exposures to T2D. First, we used GTEx Whole Blood cis-eQTLs as IVs and selected 24 genes whose expression, measured by RNA sequencing in blood, is associated with T2D status in the Trans-Omics for Precision Medicine (TOPMED) program. Next, we included pancreatic islets (TIGER) eQTLs and GTEx Pancreas cis-eQTLs as IVs for the exposure and restricted our analysis to 20 genes whose expression in pancreatic islets is associated with T2D risk, identified from a study by the IMIDIA Consortium. We performed sensitivity analyses to detect heterogeneity or horizontal pleiotropy.
Results: Among the differentially expressed genes, our MR analysis identified a significant causal relationship between *SREBF1* (sterol regulatory element binding transcription factor 1) gene expression in whole blood and T2D susceptibility ($P = 1.1 \times 10^{-6}$, MR effect size=0.1) using 16 SNPs as IVs. The rs1889018, a known eQTL for *SREBF1* in whole blood showed an association with T2D, body mass index and HbA1c, a biomarker often used to diagnose T2D. Additionally, six other SNPs at the locus were associated with HbA1c. Finally, rs76504482 and rs8077176 showed association with fasting glucose, and insulin-like growth factor, respectively.
Conclusion: Our study highlights the importance of *SREBF1* as a genetic determinant in T2D through the application of a Two-Sample Mendelian Randomization approach. Sensitivity analyses did not detect any heterogeneity or horizontal pleiotropy, providing robust evidence for the role of SREBF1 in T2D pathogenesis. We have shown how genetic determinants underlying T2D susceptibility and gene expression levels can be used to identify potential therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5071T: Exploring the impact of rare variant burden on COVID-19 severity: A whole genome sequencing analysis

Authors:

C. Dowd[1], D. Howrigan[1], L. Nkambule[1], S. Chapman[1], F. Cerrato[1], S. Bryant[1], V. Kaur[2], E. Perez[3], L. Brenner[2], C. Pato[4], M. Pato[4], L. Jehi[5], S. Comhair[5], J. Barnard[5], F. Cheng[5], J. Florez[2], M. Talkowski[2], B. Neale[2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA, [4]Rutgers Univ., Newark, NJ, [5]Cleveland Clinic Fndn., Cleveland, OH

Abstract:

Patients with SARS-CoV-2 infections are known to exhibit widely varying clinical presentations and severity of outcomes. In addition to factors such as age, sex, ancestry, and comorbidities, an individual's genetic makeup can influence their susceptibility to SARS-CoV-2 infection and the severity of their illness. While several common variants have been linked to COVID-19 severity, less is known about the potential influence of rare variants. We use whole genome sequencing (WGS) data collected from patients with varying levels of COVID-19 severity to explore the impact of rare genetic variant burden on disease severity.

This project pulls together whole genome sequencing samples collected from three research institutions across the United States. Following quality control, our full cohort consists of 2830 SARS-CoV-2 cases and 4141 controls. We compared the burden of rare loss-of-function and damaging missense variants across groups of patients with differing degrees of disease severity. We additionally analyzed the effects of rare variant burden in loss-of-function constrained genes and immunologically relevant gene sets. Finally, we seek to validate the association identified by the COVID-19 Host Genetics Initiative (HGI) between severe COVID-19 and rare deleterious variants in the toll-like receptor 7 (TLR7) gene.

Our initial analyses considered rare variants as those with minor allele frequency less than 0.1%. Using this definition, we found a modest enrichment of rare protein-truncating variants (PTV) among severe COVID-19 cases of non-Finnish European ancestry (OR = 1.06, p = 0.025), but did not observe this effect in other ancestral groups. We also identify a handful of subjects (N = 6) in our cohort carrying rare deleterious mutations in the TLR7 gene, and plan to meta-analyze our association results for this gene alongside summary statistics provided by the COVID-19 HGI to better understand its role in disease outcomes. In all, this project leverages a significant cohort of COVID-19 cases and controls to advance our knowledge about the impact of rare variant burden on the clinical outcomes of infected patients.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5072T: Exploring the potential causal association of gut microbiota on panic & conduct disorder using a two-sample Mendelian randomization approach

Authors:

A. Fatoba[1], C. Simpson[2]; [1]Univ. of Tennessee Hlth.Sci. Ctr., Cordova, TN, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

Abstract:

The potential causal association of gut microbiota with panic disorder and conduct disorder remains unexplored. This study uses detailed Mendelian randomization (MR) analyses to unravel the causal association of specific bacterial taxa with these disorders. Genome-wide association studies summary-level dataset for gut microbiota (n = 18,340), panic disorder (n = 200,486), and conduct disorder (n = 216,179) were retrieved from the MiBioGen and FinnGen consortium, respectively, and subjected to two-sample Mendelian randomization analysis. The inverse-variance weighted method was used as the primary analysis to estimate the causal effect complemented by other MR methods. Sensitivity analyses were also carried out to assess the validity of our results. Based on IVW MR analysis, thirteen bacterial taxa were detected to be causally associated with panic and conduct disorders. Among these 13 bacterial taxa, the genera *Eubacterium hallii* (OR = 1.39, 95%CI = 1.09-1.78; *P* = 0.0077) and *Alistipes* (OR = 1.74, 95%CI = 1.22-2.47; *P* = 0.0018) increased the risk of panic disorder while genus *Coprococcus* (OR = 2.39, 95%CI = 1.16-4.92; *P* = 0.0179) and class *Coriobacteriia* (OR = 2.20, 95%CI = 1.01-4.77; *P* = 0.045) increased the risk of conduct disorder. The other 9 bacteria taxa function as a protective factor as they reduce the risk of the two psychiatric disorders. There was also the absence of horizontal pleiotropy and heterogeneity. Our study provides genetic evidence for the causal association of gut microbiota with panic and conduct disorder. This provides a bedrock for future prevention and treatment of these disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5073T: Fifty-one novel, replicated loci identified in genome-wide association study of polyunsaturated and monounsaturated fatty acids in 124,024 European individuals

Authors:

H. Xu[1], M. Francis[2], Y. Sun[1], J. Brenna[3,4,5], K. Ye[1,2]; [1]Dept. of Genetics, Univ. of Georgia, Athens, GA, [2]Inst. of Bioinformatics, Univ. of Georgia, Athens, GA, [3]Dell Pediatric Res. Inst. and Dept. of Pediatrics, The Univ. of Texas at Austin, Austin, TX, [4]Dell Pediatric Res. Inst. and Dept. of Chemistry, The Univ. of Texas at Austin, Austin, TX, [5]Dept. of Nutritional Sci., Coll. of Natural Sci., The Univ. of Texas at Austin, Austin, TX

Abstract:

Background: Circulating polyunsaturated and monounsaturated fatty acid (PUFA and MUFA) levels, whose imbalances co-occur with human metabolic diseases, have strong heritable components. Despite the identification of hundreds of genomic loci for circulating PUFA and MUFA levels, these loci account for only a tiny fraction of trait variance. **Methods:** We performed a genome-wide association study (GWAS) on fourteen PUFA and MUFA phenotypes, measured by nuclear magnetic resonance in plasma. To further increase the power of identifying additional association signals, we will employ a multi-trait analysis of GWAS (MTAG) by leveraging fourteen GWAS summary statistics. We will develop and validate polygenic scores for fourteen PUFA and MUFA phenotypes using single-trait and multi-trait GWAS summary statistics. **Results:** We identified 612 significant loci-phenotype associations (115 unique loci; $P<1.678*10^{-8}$) in a European cohort from UK Biobank Phase 1 (UKB-EUR; N=101,729). Replication of five phenotypes (omega-3, omega-6, DHA, LA, MUFAs) was conducted in two external European studies: FinMetSeq (N=8,751) and a meta-analysis by Kettunnen *et al*. (N=3,644-13,544). Meta-analysis of these three studies yielded 254 significant loci-phenotype associations (109 unique loci; $P<2.439*10^{-8}$); we identified 87 novel loci, 51 of which were replicated. A transcriptome-wide association study of the UKB-EUR cohort revealed an additional twelve novel loci. We are in the process of MTAG analysis and developing polygenic scores. **Conclusion:** Our study identifies 51 novel and replicated loci associated with fourteen PUFA and PMUFA phenotypes. This study improves our understanding of the genetic architecture of unsaturated fatty acids and can inform future gene-based dietary interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5074T: Functional screening of AD genes via CRISPR activation in microglia.

Authors:

R. Kosoy[1], J. Vicari[1], J. Fullard[1], C. Coleman[1], J. Bendl[2], S. Kleopoulos[1], Z. Shao[1], S. Argyriou[1], Y. Ma[1], J. Humphrey[1], E. Brophy[1], T. Raj[1], P. Katsel[1], G. Voloudakis[1], D. Lee[1], D. Bennett[3], V.

Haroutunian[1], G. Hoffman[1], P. Roussos[1]; [1]Mount Sinai Sch. of Med., New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [3]Rush Univ., Chicago, IL

Abstract:

Alzheimer's disease (AD) imposes a massive public health burden with no effective treatment or preventative measures available. Microglia are implicated as the primary cell type in AD etiology. Previously, we identified 21 AD genes via microglia-specific expression and chromatin accessibility quantitative trait loci (QTLs). Here, we tested the transcriptome for association with AD measures in primary ex-vivo human microglia from 63 cases and 68 controls. Combined with co-expression network analyses, we identified key driver genes upregulated (APOE, APOC1, CD81, IL15), or downregulated (CECR2, SELENBP1, MNDA), in AD. Elucidating whether these differentially expressed genes (DEGs) are causal, consequential, or compensatory to AD-associated changes requires functional studies.

We utilized a multiplex CRISPR platform to over-express 14 AD genes (7 AD GWAS genes and 7 DEGs) in the HMC3 microglia cell line. ECCITE-seq (Expanded CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes) enables screening of multiple genes at single-cell resolution. Gene activation was performed via dCASP9-VPR, with target specific guide RNAs delivered via lentiviral transduction. Since microglia are myeloid cells with a role in environmental surveillance, exposure to lentivirus triggers a profound transcriptional response. Importantly, these changes include many of the AD implicated genes. Using supervised and unsupervised approaches, we sought to define, and correct for, lentivirus-associated transcriptional changes and, in so doing, identify target gene specific changes.

After removing the lentiviral effect, a consistent signal was obtained for nine of the 14 AD genes tested. Over-expression of APOC1, up-regulated in AD, leads to a reduction in cell-cycle activity and higher expression of genes involved in oxidative phosphorylation, and PTPRG, a gene strongly associated with AD in microglia. Activation of APOE increases epithelial-mesenchymal transition, hypoxia response, and TNF-signaling via NF-Kb. Conversely, IL15 activation leads to downregulation of autoimmune responses, consistent with the observation of a reduced immune functional state in AD-associated microglia. Altogether, these functional explorations of AD associated genes enable us to distinguish between disease state establishing roles versus compensatory roles for those genes, revealing potential targets and new directions for therapeutic intervention. We also demonstrate that it is critical to address the inadvertent transcriptional changes induced by lentiviral transduction, the approach most commonly used in immune cell CRISPR screens.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5075T: Gene-set based analysis of whole genome sequencing data of families with schizophrenia and bipolar disorder from Portuguese Islands Collection

Authors:

H. Kazemi[1], J. Drake[1], T. Bigdeli[2], S. Bacanu[3], T-H. Nguyen[3], K. Benke[4], B. Maher[4], J. Knowles[5], S. McCarroll[6], C. Carvalho[7], H. Medeiros[8], R. Ferreira[9], M. Pato[5], C. Pato[5], V. Vladimirov[1], A. Fanous[1]; [1]Univ. of Arizona, Phoenix, AZ, [2]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [3]Virginia Commonwealth Univ., Richmond, VA, [4]Johns Hopkins Univ., Baltimore, MD, [5]Rutgers Univ., Rutgers–New Brunswick, NJ, [6]Harvard Univ., Boston, MA, [7]Univ. of the Azores, Ponta Delgada, Portugal, [8]Univ. of Southern California, Los Angeles, CA, [9]Inst. S. Joao de Deus, Lisbon, Portugal

Abstract:

Schizophrenia (SCZ) and bipolar disorder (BPD) are complex psychiatric phenotypes with substantial genetic components, characterized by polygenic inheritance involving numerous genetic variants. Here, we analyzed whole genome sequencing (WGS) data from the Portuguese Islands Collection (PIC), which included families with SCZ (n=44) and BPD (n=64) cases, and controls (n=62). Specifically, we performed genetic associations of 141 gene-sets (associated with psychiatric disorders) with SCZ and BPD using rare variants (minor allele frequency ≤ 0.01). Variants were annotated using ANNOVAR with GRCh38 and refGene. Combined Annotation Dependent Depletion (CADD) with hg38 was used to assign phred-like scores, focusing on variants in the top 10% and 20% most deleterious categories (CADD-PHRED scores ≥ 10 and ≥ 20, respectively). By filtering variants from the PIC dataset based on CADD-PHRED, we conducted gene-set based analysis using Linear Mixed Model (LMM) and Scalable and Accurate Implementation of Generalized mixed model (SAIGE). Our findings revealed significant associations between certain gene-sets, their constituent genes, and rare/deleterious variants with SCZ and BPD. The False Discovery Rate (FDR) method was used for multiple test comparisons. LMM identified significant relationships (p-value ≤ 0.05; q-value ≤ 0.1) for 6 gene-sets including microRNA-137, *RBFOX2*/*RBFOX13*, loss-of-function intolerant genes, abnormal long-term potentiation, abnormal neuron morphology, and abnormal somatic nervous system morphology with SCZ. Additionally, suggestive associations (p-value ≤ 0.05) were observed for 9 gene-sets encompassing *RBFOX2*/*RBFOX13*, synaptic genes, *CELF4*, abnormal choroid plexus morphology, *FMRP*, abnormal brain ventricle morphology, abnormal neurotransmitter levels, abnormal dopaminergic neuron morphology, and abnormal neuron proliferation/differentiation with BPD. SAIGE analysis corroborated these findings, revealing significant associations of 102 and 114 gene-sets with SCZ and BPD, respectively (p-value ≤ 0.05; q-value ≤ 0.1), including gene-sets related to nervous, neuronal, and synaptic systems. Furthermore, shared significant gene-

sets were identified between SCZ and BPD, as well as between LMM and SAIGE analyses for both disorders, with nonsynonymous variants being the most significant (Fisher's Exact Test p-value (SCZ) = 0.0104; p-value (BPD) = 0.0189). These findings suggest common genetic underpinnings and provide valuable insights into the genetic architecture of SCZ and BPD, highlighting potential shared molecular mechanisms that could inform future research and therapeutic strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5076T: Genetic Analysis in Over 51,000 Mother-Child Pairs Suggests Higher Maternal Testosterone Protects Daughters from Anxiety

Authors:

J. Leinonen[1], FinnGen, T. Tukiainen[2]; [1]FIMM, Univ. of Helsinki, Helsinki, Finland, [2]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland

Abstract:

Hormones such as testosterone may play a role in psychiatric health. While genetic studies have found limited evidence linking adult testosterone levels to behavioral outcomes, fetal hormonal exposure is speculated to influence such predispositions.
To investigate the role of maternal testosterone in psychiatric disease risk, we conducted Mendelian Randomization (MR) analyses for child anxiety (N=4824) and depression (N=7490) in over 51,000 genotyped mother-child pairs (child mean age 46.2) from the FinnGen study. We selected 99 SNPs associated with significant effects on female testosterone levels as the exposure and obtained conditional effect estimates for these SNPs on child anxiety and depression. Importantly, we could separate between the genetic effects of the mothers and their children in these analyses. Polygenic score (PGS) analyses and data from father-child and sibling pairs served as further sensitivity analyses.
While the child's own genome showed no statistically significant effects on anxiety or depression risk, higher maternal testosterone levels were associated with protection from child anxiety, particularly in daughters (IVW β=-0.193, P=0.022). This finding was supported by PGS analyses (for daughter anxiety, HR=0.956, P=0.037 per 1SD in mother testosterone PGS). No effect was observed in mother-son pairs or for depression (P>0.05). As a negative control, no significant associations (P>0.05) were found between paternal testosterone levels and child anxiety risk (N~44,000 father-child pairs), suggesting a mother-specific, possibly prenatal, mechanism. Intriguingly, alone mothers homozygous for a testosterone-lowering G-allele of rs45446698 in the CYP3A7 locus, affecting testosterone metabolism and associated with the

largest effect on female testosterone of all common SNPs, showed a substantial effect on child anxiety (β=0.586, P=0.024).
Our findings, based on a unique resource of mother-child pairs, link the maternal genome to effects on child psychiatric health. These results not only support the role of prenatal hormone exposure in modifying the predisposition to anxiety-related behavior in humans but also highlight the potential of large-scale familial data to elucidate the complex biology underlying diverse traits.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5077T: Genetic analysis of X-chromosome in myasthenia gravis

Authors:

M. Rasheed[1], R. Chia[2], A. Ray[2], E. Bayram[3], C. Provenzano[4], B. Traynor[5]; [1]Natl. Inst. on Aging, Bethesda, MD, [2]NIH, Bethesda, MD, [3]Univ. of California San Diego, San Diego, CA, [4]Fondazione Policlinico Univ.rio "A. Gemelli" Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Rome, Italy, [5]Natl. Inst Aging, Bethesda, MD

Abstract:

Myasthenia gravis is an autoimmune condition that affects the neuromuscular junction, causing muscle weakness. The clinical course is affected by age and sex, with females likely to have earlier disease onset. This suggests that genetic risks on chromosome X may contribute to the observed dimorphism in clinical outcomes. With more recent developments in consensus quality control metrics specific to chromosome X, the role of the X chromosome in disease susceptibility can now be investigated. Following previous work on identifying autosomal genetic risks in myasthenia gravis, we performed a chromosome X-wide association study (XWAS) on genotype data from the same cohort of 1871 acetylcholine receptor-positive myasthenia gravis cases and 36,106 neurologically healthy European-ancestry individuals. The discovery cohort constituted samples from two cohorts from the US and Italy. Genotype imputation was performed using the TOPMed Imputation Server. Genetic principal components were derived from both autosomes and X chromosomes. Covariates were selected to correct for confounding subpopulation stratification using the step function in R. Association analysis was performed in a sex-stratified manner and overall analysis of all samples. This was done separately for the US and Italian cohorts, followed by a fixed-effect meta-analysis. The number of haploblocks on chromosome X, specific to the cohorts, was determined and used to set the significance threshold for the XWAS (0.05/6570 haploblocks = 7.61 x 10-6). Our analysis identified three novel risk loci in the overall analysis. The index variants are located within

Xp22.2 (upstream to *GRPR*) (odds ratio (OR)=1.69, 95% CI=1.36-2.09, P=2 x 10-6), Xq23 (intron 2 of *TMEM164*; OR =1.49, 95% CI=1.26-1.78, P=5 x 10-6), and Xq27.3 (intron 3 of *ENSG00000227303*; OR=1.21, 95% CI=1.11-1.31, P=3x 10-6). The sex-stratified analysis of the male and female cohorts revealed no significant association. Our preliminary findings suggest a role for chromosome X in the pathogenesis of myasthenia gravis. Replication in an independent cohort and further investigations would help give new insights into molecular pathways that may be amenable to therapeutic interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5078T: Genetic architecture and comprehensive phenotypic assessment of alcohol use disorder and related outcomes in All of Us.

Authors:

B. Webb[1], A. Moore[1], M. F. Hassan[2], R. E. Peterson[3], M. Singh[2], B. P. Riley[2], S. A. Bacanu[2], J. D. White[1], D. B. Hancock[1], E. O. Johnson[1], A. E. Gentry[2]; [1]RTI Intl., Research Triangle Park, NC, [2]Virginia Commonwealth Univ., Richmond, VA, [3]SUNY Downstate Hlth.Sci. Univ.State Univ. of New York Downstate Hlth.Sci. Univ, Brooklyn, NY

Abstract:

Alcohol use disorder (AUD) is moderately heritable with significant social and economic impacts. The reported prevalence of AUD in the All of Us (AoU) is 1.9% when using EHR-based ICD ascertainment. In contrast, studies using structured interviews show much larger lifetime AUD prevalences (~29%). This discrepancy may impact genetics studies since misclassification reduces statistical power. AoU has information beyond AUD ICD codes including procedures, prescriptions, laboratory values, surveys, and self-report. A comprehensive assessment of these domains may identify additional cases, participants at increased risk, and screened controls. Cases were defined using self-report variables and EHR elements specific to AUD including prescriptions, health consequences, and treatments. Participants that were heavy drinkers, regular binge drinkers, or had a family history of AUD were defined as high-risk (HR). Screened controls excluded AUD cases, HR, non-drinkers, and unclassified participants. Unscreened controls included HR participants. Genetic analyses were performed within AoU provided groupings based on genetic similarity to reference populations. SNP-based heritabilities (h^2) were estimated using REML in BOLT-LMM for groups with at least 1000 genotyped AUD cases. Age, sex, and 16 genetic ancestry principal components (PCs) were included as covariates. Four case/control definitions were examined including a) AUD vs. unscreened controls and b) AUD, c) HR, and d) AUD+HR vs screened controls. Using release V7, we identified 25,988 (6.3%) AUD

cases, 19,288 (4.7%) HR, 307,139 (74.3%) screened controls, 39,955 (9.7%) non-drinkers, and 21,087 (5.1%) participants with insufficient information for classification. Sufficient unrelated AUD cases for h^2 analyses were present in the pre-assigned AFR (n= 4,872), AMR (n= 2,410), and EUR (n= 9,179) groups. AUD was more heritable when using screened (h^2_{AFR}=0.055, h^2_{AMR}=0.154, h^2_{EUR} =0.060) versus unscreened controls (h^2_{AFR}=0.037, h^2_{AMR}=0.133, h^2_{EUR} =0.054). Combining AUD and HR increased the case counts but (h^2_{AFR}=0.052, h^2_{AMR}=0.142, h^2_{EUR} =0.064) did not meaningfully change h^2. Comprehensive assessment across domains in AoU identifies additional cases, HR subjects, and screened controls and produces equal or higher h^2 from previous molecular studies of AUD. However, AUD+HR is still less prevalent (11%) than lifetime AUD from structured interviews, implying there are additional unidentified lifetime AUD cases. Further analyses, such as using genomic structural equation modeling, are needed to determine which AUD classifications will maximize power in future GWAS studies of AUD in AoU.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5079T: Genetic architecture of the limbic white matter in aging

Authors:

A. Lorenz[1,2], A. Sathe[1], D. Zaras[1], Y. Yang[1], A. Durant[1], N. Shashikumar[1], K. Pechman[1], M. E. Kim[3], C. Gao[4], N. Khairi[4], Z. Li[4], T. Yao[3], Y. Huo[4], L. Dumitrescu[1], K. A. Gifford[1], S. L. Risacher[5], L. L. Beason-Held[6], Y. An[6], K. Arfanakis[7], G. Erus[8], C. Davatzikos[8], D. Tosun[9], A. W. Toga[10], P. M. Thompson[11], E. C. Mormino[12], P. Zhang[1], K. Schilling[13], Alzheimer's Disease Neuroimaging Initiative (ADNI), The BIOCARD Study Team, The Alzheimer's Disease Sequencing Project (ADSP), M. Albert[14], W. Kukull[15], S. A. Biber[15], B. A. Landman[4], S. C. Johnson[16], B. Bendlin[16], J. Schneider[17], L. L. Barnes[17], D. A. Bennett[17], A. L. Jefferson[1], S. M. Resnick[6], A. J. Saykin[5], T. J. Hohman[1], D. B. Archer[1]; [1]Vanderbilt Memory and Alzheimer's Ctr., Vanderbilt Univ. Sch. of Med., Nashville, TN, [2]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Dept. of Computer Sci., Vanderbilt Univ., Nashville, TN, [4]Dept. of Electrical and Computer Engineering, Vanderbilt Univ., Nashville, TN, [5]Dept. of Radiology and Imaging Sci., Indiana Univ. Sch. of Med., Indianapolis, IN, [6]Lab. for Behavioral NeuroSci., Natl. Inst. on Aging, NIH, Baltimore, MD, [7]Dept. of BioMed. Engineering, Illinois Inst. of Technology, Chicago, IL, [8]Dept. of Radiology, Univ. of Pennsylvania, Philadelphia, PA, [9]Dept. of Radiology and BioMed. Imaging, Univ. of California San Francisco, San Francisco, CA, [10]Lab. of Neuroimaging, USC Stevens Inst. of Neuroimaging and Informatics, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [11]Imaging Genetics Ctr., Mark and Mary Stevens Inst. for Neuroimaging and Informatics, Keck Sch. of Med., Univ. of Southern California, Marina del Rey, CA, [12]Dept. of Neurology and Neurological Sci., Stanford

Univ. Sch. of Med., Stanford, CA, [13]Dept. of Radiology & Radiological Sci., Vanderbilt Univ. Med. Ctr., Nashville, TN, [14]Dept. of Neurology, Johns Hopkins Sch. of Med., Baltimore, MD, [15]Natl. Alzheimer's Coordinating Ctr., Univ. of Washington, Seattle, WA, [16]Wisconsin Alzheimer's Disease Res. Ctr., Sch. of Med. and Publ. Hlth., Univ. of Wisconsin, Madison, WI, [17]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL

Abstract:

Alterations in white matter (WM) microstructure contribute to cognitive decline in aging, but its genetic basis is unexplored. This study assessed the role of genetic factors in microstructure integrity across seven WM tracts of the limbic system, which are crucial for memory function. WM microstructure was measured in seven harmonized cohorts (N=2,614) using free water-corrected diffusion MRI metrics, including axial diffusivity (AxD), mean diffusivity (MD), radial diffusivity (RD), and fractional anisotropy (FA). Participants included in the analysis were non-Hispanic White individuals aged 50 to 100 years (mean=73.66, SD=9.76), 42.65% were male. All had genome-wide genotype data imputed using the TOPMed reference panel. Using GCTA, we found that 16 of 35 metrics across all limbic tracts were heritable with estimates ranging from 0.26 to 0.60 ($p_{FDR\text{-}corrected}<0.05$). Next, a cross-sectional GWAS for each metric of each tract was performed using linear regression models covarying for sex, age, and the first three genetic ancestry principal components. Following a meta-analysis across cohorts, we identified genome-wide significant signals ($p<5x10^{-8}$) for fornix AxD (rs78407651, EAF=0.03, z=-5.46, $p=4.77x10^{-8}$, near *SENP5*), inferior longitudinal fasciculus (ILF) FA (rs8026709, EAF=0.36, z=5.75, $p=8.94x10^{-9}$, near *RORA*) and AxD (rs8026709, EAF=0.36, z=5.84, $p=5.84x10^{-9}$, near *RORA*), in addition to superior temporal gyrus transcallosal tract (STG) RD (rs11542181, EAF=0.03, z=-5.57, $p=2.64x10^{-8}$, near *MIR1265*) and cingulum RD (rs56017587, EAF=0.13, z=5.55, $p=2.98x10^{-8}$, near *MIR548A1*). A locus with 36 genome-wide significant SNPs (rs12959877, EAF=0.44, z=5.83, $p=5.78x10^{-9}$) was identified for the middle temporal gyrus (MTG) RD, with additional nominal associations for STG and inferior temporal gyrus transcallosal tract (ITG) RD ($p<1x10^{-5}$). These SNPs are eQTLs for *CDH19* which is highly expressed in WM and functions as a calcium-dependent cell adhesion glycoprotein. Gene-level analyses using MAGMA highlighted an association between cingulum AxD and SERPINA12 (z=4.60, $p_{FDR\text{-}corrected}=0.03$) which has been associated with diabetes and atherosclerosis by modulating glycolipid metabolism. Furthermore, gene set analysis indicated insulin, immune response, and neurotrophic signaling pathways and microstructural changes were genetically correlated with multiple sclerosis, lipid measures, and cardio traits ($p_{FDR\text{-}corrected}<0.05$).These findings highlight a genetic contribution to changes in WM during aging and suggest the involvement of neural development, lipid metabolism, vascular health, and inflammation mechanisms.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5080T: Genetic correlates of treatment-resistant depression: Insights from polygenic scores across cognitive, temperamental, and sleep traits in the all of us cohort

Authors:

B. Xu[1,2], K. L. Forthman[2], R. Kuplicki[2], J. Ahern[1,3], R. Loughnan[1,3], F. Naber[1,2], W. K. Thompson[1,2,4], C. B. Nemeroff[5], M. P. Paulus[2,6], C. Fan[1,2,7]; [1]Population NeuroSci. and Genetics Ctr., Laureate Inst. for Brain Res., Tulsa, OK, [2]Laureate Inst. for Brain Res., Tulsa, OK, [3]Ctr. for Human Dev., Univ. of California, San Diego, La Jolla, CA, [4]Div. of Biostatistics and Bioinformatics, the Herbert Wertheim Sch. of Publ. Hlth.and Human Longevity Sci., Univ. of California, San Diego, La Jolla, CA, [5]Dept. of Psychiatry and Behavioral Sci., Dell Med. Sch., The Univ. of Texas at Austin, Austin, TX, [6]Dept. of Psychiatry, Univ. of California, San Diego, La Jolla, CA, [7]Dept. of Radiology, Univ. of California, San Diego, La Jolla, CA

Abstract:

Background:
Treatment-resistant depression (TRD) is a major challenge in mental health, affecting a significant number of patients and leading to considerable economic and social burdens. The etiological factors contributing to TRD are complex and not fully understood.
Objective:
To investigate the potential etiological factors of TRD by using polygenic scores (PGS) across various traits that were reported to be co-occur with TRD status.
Method:
Data from 292663 participants in the All of US Research Platform (AoU) were analyzed using a case-cohort design. TRD, treatment responsive Major Depressive Disorder (trMDD), and all others who have no formal diagnosis of Major Depressive Disorder (non-MDD) were identified through diagnostic codes and prescription patterns. PGS for 61 unique traits from seven domains were used and logistic regressions were conducted to assess associations between PGS and TRD. Finally, a Cox proportional hazard models were used to explore the predictive value of PGS for progression rate from the diagnostic event of Major Depressive Disorder (MDD) to TRD.
Results:
In the discovery set (104128 non-MDD, 16640 trMDD, and 4177 TRD), 44 of 61 selected PGS were found to be significantly associated with MDD. Eleven of them were found to have stronger associations with TRD than with trMDD, encompassing PGS from domains in education, cognition, personality, sleep, and temperament. Genetic predisposition for insomnia and specific neuroticism traits were associated with increased TRD risk (OR range from 1.05 to

1.15), while higher education and intelligence scores were protective (ORs 0.88 and 0.91). These associations are consistent across two other independent sets within AoU (n = 104388 and 63330). Among 28964 individuals tracked over time, 3854 developed TRD within an average of 944 days (95% CI: 883 ~ 992 days) after MDD diagnosis. All eleven previously identified and replicated PGS were found to be modulating the conversion rate from MDD to TRD. Those having higher education PGS would experiencing slower conversion rates than those who have lower education PGS with hazard ratios in 0.79 (95% CI: 0.74 ~ 0.85). Those who had higher insomnia PGS experience faster conversion rates than those who had lower insomnia PGS, with hazard ratios in 1.21 (95% CI: 1.13 ~ 1.30).
Conclusion:
Our results indicate that genetic predisposition related to neuroticism, cognitive function, and sleep patterns play a significant role in the development of TRD. These findings underscore the importance of considering psychosocial factors in managing and treating TRD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5081T: Genetic correlations inform machine learning based identification of heavy alcohol consumption using objective biomarkers

Authors:

M. Hassan[1], A. Gentry[1], R. Peterson[2,1], B. Webb[3,1]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]SUNY Downstate Hlth.Sci. Univ.State Univ. of New York Downstate Hlth.Sci. Univ, Brooklyn, NY, [3]RTI Intl., Research Triangle Park, NC

Abstract:

Alcohol use disorder (AUD) is a moderately heritable complex disorder with significant social and economic impacts with even moderate alcohol consumption (AC) increasing some health risks. While self-report via survey or structured interview is generally reliable, being able to predict AC using objective biomarkers could allow for unbiased and earlier detection of potential problems. Machine learning (ML) analyses of biomarkers can be used to construct robust predictive models of complex outcomes such as AC. To predict AC as measured by number of alcoholic drinks per week (DPW), we evaluated five ML models, including LASSO, ridge regression, Gradient Boosting Machines (GBM), Model Boosting (MBoost), and Extreme Gradient Boosting (XGBoost). This study was conducted using the UK Biobank Resource application number 30782. For DPW prediction, 338 initial predictors were considered including blood count (n=31), blood biochemistry (n=31), infectious diseases (n=25), urine assays (n=3), and NMR metabolomics (n=249). Discovery analyses were conducted on unrelated participants

of European ancestry with complete biomarker data and DPW (n=63,058) and were randomly allocated to the training/testing sets with a percentage of 80/20. In addition to evaluating performance using phenotype correlations, genetic correlation (r_G) analyses were performed using LDSC. Using linear regression, sex, age, statin use, and liver enzymes accounted for 0.105 of the variation in DPW. All five ML methods achieved significantly better but moderate prediction (R^2=0.304-0.356). XGBoost performed the best (R^2=0.356, MAE=5.214), but required a more complex model structure and training resources. While phenotype correlations were similar across methods, XGBoost produced similar heritability (h^2=0.0776) to observed DPW (h^2=0.0643). In contrast, the other ML methods produced higher h^2 but lower r_Gs. The best estimated r_G between observed and predicted DPW is 0.8774. All ML models were able to accurately predict if subjects were heavy drinkers using CDC based sex-specific cutoffs, with accuracies, sensitivities, and specificities of 87.8%, 72.2%, and 93.0% for women (DPW>8) and 89.0%, 73.4%, and 94.3% for men (DPW>15), respectively. Using ML methods in a large sample, we have demonstrated that a panel of objective biomarkers can sufficiently predict AC to accurately screen and identify heavy drinkers. Additionally, as demonstrated by the h^2 and r_G estimates of the predicted outcomes, training ML models on large biobank data can predict missing and unmeasured outcomes, thereby boosting sample sizes and power for genetic loci discovery.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5082T: Genetic covariance analysis identifies *TMEM106B* and *ACE* as genetic loci shared between Alzheimer's disease and primary psychiatric disorders

Authors:

A. Kumar[1], N. Ray[1], T. Hohman[2], A. De Vito[3], C. Stein[3], P. Del Rosario[1], M. Cuccaro[4], G. Beecham[4], E. Huey[3], C. Reitz[1]; [1]Taub Inst. for Res. on the Aging Brain, Columbia Univ. Med. Ctr., New York, NY, [2]Vanderbilt Memory and Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Dept. of Psychiatry and Human Behavior, Alpert Med. Sch. of Brown Univ., Providence, RI, [4]John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL

Abstract:

Background: Neuropsychiatric Symptoms (NPS) (e.g., aggression, psychosis, anxiety, apathy, depression, agitation, sleep disturbances, repetitive behaviors) occur in 85% of Alzheimer's Disease (AD) patients, and are associated with accelerated decline and greatly increased suffering of patients and their caregivers. Our understanding of the etiology of NPS in AD is

inadequate, with treatments for NPS often being ineffective and associated with serious adverse effects. Pharmacological treatments have generally been borrowed from their indications for psychiatric illness, despite limited understanding to what extent the molecular and genetic architectures underlying AD-associated NPS overlap with that of primary psychiatric disorders.
Method: To better characterize the genetic overlap between AD and major psychiatric disorders and identify potentially shared biological processes we conducted local genetic covariance analyses between AD and depression, and AD and schizophrenia using LAVA, capitalizing on the largest most recent GWAS summary statistics for these traits (AD: n=487,511; bipolar disorder: n=413,466; depression: n=1,154,267; schizophrenia: n=130,644). Identified overlapping genomic loci were followed up with finemapping and in silico functional analyses.
Results: Local genetic covariance analyses identified four loci shared between AD and depression and five loci shared between AD and schizophrenia. Finemapping and in silico functional analyses at these loci pinpoint a missense variant in *TMEM106B* as a causative variant shared with depression, and a regulatory region variant in *ACE* as a causative variant shared with schizophrenia. *TMEM106B* is a lysosomal protein critical for clearance of misfolded proteins and is associated with decreased cortical volume, degree of TDP-43 pathology, and decreased neuronal proportion. *ACE*, the main component of the renin-angiotensin system (RAS) critical for blood pressure regulation and homeostasis, is located on the cellular membrane of neurons, modulates innate and adaptive immunity, and cleaves several neuropeptides involved in neurologic and psychiatric disorders including glutamate, dopamine, and GABA.
Conclusion: These findings support the notion of pleiotropic overlap between AD and depression and schizophrenia, respectively, and pinpoint *TMEM106B* and *ACE* as shared genetic drivers. Further functional characterization of these loci and molecular pathways will be critical to characterize the specific underlying biological processes.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5083T: Genetic differences between male & female pattern hair loss in a Korean population

Authors:

K-W. Hong; Theragen Hlth.Co. Ltd., Seongnam, Korea, Republic of

Abstract:

Studies on androgenetic alopecia (AGA or patterned hair loss (PHL)) have suggested differences in its underlying pathological mechanisms when affecting males or females. To date, more than

a dozen genome-wide association studies have been conducted and numerous genetic factors have been identified for male hair loss; however, the genetic determinants of female hair loss remain unclear. In this study, we identified sex-specific genetic factors in approximately 1,000 individuals, including 436 males and 568 females. We conducted three independent genome-wide association studies for the total, male-only, and female-only groups and identified three novel loci (rs7814359, rs2163085, and rs4793158 of the TSNARE1, FZD1, and GJC1 genes, respectively). The single nucleotide polymorphism (SNP) rs7814359 showed a significant genome-wide association with AGA in the total sex group and a weak association in both the male-only and female-only groups. The rs2163085 SNP showed a significant genome-wide association with AGA in the total sex group, which was significant only in females. The remaining SNP, rs4793158, showed a suggestive association with AGA in both the total and female-only groups. The SNP rs7814359 is located in TSNARE1, which is involved in intra and extracellular vesicle transport. FZD1 encodes a frizzled protein that is a key regulator of the Wnt signaling pathway. GJC1 is a gap junction protein. The FZD1 and GJC1 loci were associated with female-specific AGA, implying that sex hormones (such as estrogen) influenced female pattern hair loss through novel candidate genes. These findings will contribute to our understanding of the sex-specific pathophysiology of AGA.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5084T: Genetic mechanisms define novel drug targets for IgA vasculitis ⭐

Authors:

L. Liu[1], L. Zhu[2], S. Monteiro-Martins[3], A. Griffin[1], L. Vlahos[1], F. Zanoni[1], F. Lugani[4], K. Mucha[5], M. Mizerska-Wasiak[6], L. Berthelot[7], J. Novak[8], W. Smoyer[9], R. Wyatt[10], R. Wyatt[10], R. Nelson[11], J. Martin[12], M. González-Gay[13], A. Köttgen[14], A. Califano[15], A. Gharavi[15], H. Zhang[2], K. Kiryluk[16]; [1]Columbia Univ., New York City, NY, [2]Peking Univ. Inst. of Nephrology, Beijing, China, [3]Univ. of Freiburg, Freiburg, Germany, [4]The Giannina Gaslini Children's Hosp. -- Genova, Italy (IGG), Genova, Italy, [5]Warsaw Univ. Med. Ctr., Warsaw, Poland, [6]Warsaw Pediatric Hosp., Warsaw, Poland, [7]French Inst. of Hlth.and Med. Res., Strasbourg, France, [8]Univ. of Alabama at Birmingham, Alabama, AL, [9]Nationwide Children's Hosp., Washington, WA, [10]Univ. of Tennesee, Memphis, TN, [11]Univ. of Utah, Salt Lake City, UT, [12]Inst. of Parasitology and Biomedicine Lopez-Neyra, Spanish Natl. Res. Council, Granada, Spain, [13]Hosp. Univ.rio Marqués de Valdecilla, Santander, Spain, [14]Inst. of Genetic Epidemiology Med. Ctr., Univ. of Freiburg, Freiburg, Germany, [15]Columbia Univ, New York, NY, [16]COLUMBIA Univ. HEALTH Sci.S, New York, NY

Abstract:

Background: IgA vasculitis (IgAV) is the most common form of systemic vasculitis in children. The pathogenic mechanisms remain unclear and thus there are currently no targeted treatments for IgAV.
Methods: We conducted genome, transcriptome, and proteome-wide association studies (GWAS, TWAS, and PWAS) for IgAV in 2,170 cases and 5,928 controls of European and East Asian ancestries. To study the consequences of the new risk alleles, we generated IgAV-specific maps of gene expression and splicing based on blood RNA-seq of 255 pediatric IgAV cases. We used systems biology approaches to build myeloid-specific regulatory networks and define a set of master regulators of IgAV co-modulated by the genes supported by GWAS.
Results: We observed strong association at the *HLA-DRB1* locus (OR=1.55, $P=1.1x10^{-25}$) and defined specific amino acid substitutions in DRβ1 driving this association. We discovered two novel non-HLA loci: *FCAR* (OR=1.51, $P=1.0x10^{-20}$) encoding a myeloid IgA receptor FcαR, and *INPP5D* (OR=1.34, $P=2.2x10^{-09}$) encoding SHIP-1, a known inhibitor of FcαR signaling. The *FCAR* risk locus co-localized with a cis-eQTL associated with increased *FCAR* expression; the risk alleles disrupted a binding motif of *PRDM1* within a myeloid enhancer of *FCAR*. Another locus encoding *IL6R* was suggestive in GWAS and TWAS, but proteome-wide significant in PWAS with risk alleles associated with elevated levels of soluble IL6R. The IL6R risk haplotype carried a missense variant causing accelerated cleavage of IL6R into a soluble form. Using systems biology approaches, we prioritized 28 master regulators of IgAV that were co-modulated by *FCAR*, *INPP5D* and *IL6R*. These regulators alter the activity of chemokine and interleukin signaling pathways in myeloid cells of patients with IgAV.
Conclusions: Our integrative genetic analyses provided novel insights into the genetic susceptibility to IgA vasculitis, and implicated *HLA-DRB1*, *FCAR*, *INPP5D* and *IL6R* in the disease pathogenesis, nominating these genes and related molecular pathways for therapeutic targeting.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5085T: Genetic risk estimation of pregnancy complications through polygenic risk scores from low-coverage NIPS data

Authors:

S. Becelaere, Y. Abakkouy, D. Jans, J. Vermeesch, T. Kivisild, I. Cleynen; KU Leuven, Leuven, Belgium

Abstract:

Non-invasive prenatal screening (NIPS) of cell free DNA (cfDNA) present in maternal blood samples is routinely used to detect fetal aneuploidy early in the pregnancy. Besides fetal cfDNA, maternal cfDNA is also abundantly present in the samples, representing an opportunity for genetic analyses. However, these cfDNA samples are usually sequenced at low-coverage, and appropriate imputation tools need to be applied before downstream analyses. We investigated the efficiency of using imputed low-coverage sequencing data from NIPS samples for polygenic risk score (PRS) calculations. We use height as a complex trait, and preeclampsia (hypertension after 20 weeks gestation with proteinuria) and gestational diabetes mellitus (GDM, hyperglycemia with onset during pregnancy), as complex disease examples.
NIPS samples were collected at UZ Leuven and imputed with QUILT. For each complex trait, PRS were calculated with three different tools (PRSice-2, lassosum and PRS-CS) and for various parameter settings. Only the results from the best PRS tool (with the best settings) for each complex trait are reported here. PRS for height were calculated for 17,768 NIPS samples based on summary statistics of the GWAS by Yengo et al. (2022). For preeclampsia, PRS were calculated for 177 cases and 1,908 controls (Honigberg et al. 2023). Finally, for GDM, GDM PRS and type-2 diabetes (T2D) PRS were calculated for 555 cases and 1,945 controls (Pervjakova et al. 2022, and Mahajan et al. 2022, respectively).
The phenotypic variance in height explained by the PRS developed with PRS-CS was 0.33, compared to 0.44 in literature, indicating that PRS from imputed NIPS data can be used for PRS calculation. The variance in preeclampsia explained by the PRS computed with lassosum was 0.082 (OR = 1.73 (1.49-2.02); p-value = 3.30e-12), with an area under the receiver operator curve (AUC) of 0.66 (0.62-0.70), which is rather low but not unexpected given the low number of cases and the limited GWAS size. The phenotypic variance in GDM explained by the GDM PRS computed with PRSice-2 was 0.023 (OR = 1.35 (1.23-1.49); p-value = 1.18e-09; AUC = 0.58 (0.55-0.61)) and by the T2D PRS computed with PRS-CS was 0.050 (OR = 1.55 (1.41-1.71); p-value = 7.05e-19; AUC=0.62 (0.57-0.63)). This demonstrates the value of using proxy phenotypes (here T2D instead of GDM) when the GWAS size is limited. Together, these results show that imputed NIPS data can be used for estimating PRS and that PRS effectiveness in stratifying pregnancy-related complex diseases relies on the underlying genetic architecture and GWAS cohort sizes, which are expected to improve in the future.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5086T: Genetic Risk, Adherence to Brisk Walking Pace, and Venous Thromboembolism

Authors:

W. Xian[1], X. Chen[2], J. M. Ranson[3], V. Napolioni[4], J. Huang[1]; [1]Southern Univ. of Sci. and Technology, Shenzhen, China, [2]The George Washington Univ., Washington, DC, [3]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [4]Stanford Univ. Sch. of Med., Palo Alto, CA

Abstract:

Background Mendelian mutations in *F2* and *F5* genes are established risk factors for venous thromboembolism (VTE). Walking pace is associated with the risk of coronary artery diseases, but no study has investigated its association with VTE. **Methods** We aimed to investigate the association and causality between walking pace and VTE, and compare its population risk with established Mendelian mutations. We followed up 445,261 UK Biobank participants free of VTE at baseline. Self-reported walking pace was collected via touchscreen questionnaire at baseline. The carrier status of two Mendelian mutations in the *F2* and *F5* genes was determined by the genotypes of rs1799963 (G20210A, c.*97 G > A) and rs6025 (p.R534Q), respectively. **Results** Over a median follow-up period of 12.8 years, 11,155 incident VTE cases were identified. The 10-year incidence rate for brisk and slow walking pace was 1.32% (CI: 1.27% - 1.37%) and 3.90% (CI: 3.71% - 4.09%), respectively. For non-carrier and *F5* carrier, the rates were 1.70% (CI: 1.66% - 1.73%) and 3.62% (CI: 3.39% - 3.84%), respectively. The overall risk of VTE for *F5* mutation carriers with brisk walking pace (2.65%) was smaller than that for non-carriers with slow walking pace (3.66%). For *F5* mutation carriers, brisk pace (but not steady pace) reduces the risk of VTE (*P* interaction <0.05). Mendelian Randomization analyses displayed a causal relationship (IVW $P=3.21\times10^{-5}$) from walking pace to VTE incidence, but no significant causal effect was found for VTE on walking pace. **Conclusion** Slow walking pace is causally associated with a higher risk of VTE. On the population scale, the protective effect of brisk walking pace offsets the risk of VTE caused by Mendelian mutations. ***Acknowledgement*** This research was conducted using the UK Biobank resources under application 66137. We thank the participants of UK Biobank for making this work possible.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5087T: Genetic variants affect diurnal glucose levels across the day

Authors:

H. Ollila[1], S. Strausz[2], L. Urpa[3], E. Abner[4], R. Saxena[5], S. E. Jones[1], N. Sinnott-Armstrong[6]; [1]Univ. of Helsinki, Helsinki, Finland, [2]Inst. for Molecular Med. Finland, Helsinki, Finland, [3]Univ.

of Helsinki, Helsinki, Finland, [4]Univ. of Tartu, Tartu, Estonia, [5]Mass Gen. Hosp., Harvard Med. Sch., Boston, MA, [6]Fred Hutch Cancer Ctr., Seattle, WA

Abstract:

Circadian rhythms not only coordinate the timing of wake and sleep but also regulate homeostasis within the body, including glucose metabolism. While a large number of genetic variants have been found for glucose levels, the genetic variants that contribute to temporal control of glucose levels have not been previously examined. Using data from 420,000 individuals from the UK Biobank and replicating our findings in 100,000 individuals from the Estonian Biobank, we show that diurnal and temporal regulation of glucose levels is under genetic control. We discover a robust association with temporal regulation of glucose levels across the day at the Melatonin receptor 1B (MTNR1B) (rs10830963, P = 1e-22) and a canonical circadian pacemaker gene Cryptochrome 2 (CRY2) loci (rs12419690, P = 1e-16). Furthermore, we show that sleep traits modulate glucose levels and the genetic variants have an additional, independent role in regulating diurnal glucose levels. Our findings together with earlier genetic and epidemiological evidence show a clear connection between sleep and metabolism and highlight variation at MTNR1B and CRY2 as temporal regulators for glucose levels.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5088T: Genetic variants associated with TST and LPS in insomnia patients: whole genome sequencing study

Authors:

N. Synnott[1], S. Smieszek[2], H. Bai[2], C. Xiao[2], C. Polymeropoulos[2], G. Birznieks[3], M. Polymeropoulos[2]; [1]Vanda Pharmaceutical Ltd, London, United Kingdom, [2]Vanda Pharmaceuticals Inc., Washington, DC, [3]Vanda Pharma, Washington, DC

Abstract:

Insomnia disorder is characterized as the difficulty in initiating or maintaining sleep and corresponding daytime dysfunction, and occurs in 10-20% of the population. Although the molecular underpinnings of sleep-onset insomnia remain unclear, many insomnia patients with sleep-onset insomnia have a misalignment of their circadian rhythms with respect to their desired bedtime. Thus far, little is known about the underlying genetic risk factors for sleep-onset insomnia patients. We explored the association of genetic variants with TST and LPS in patients diagnosed with insomnia with SL of >45 mins and TST of <6.5 hr. Samples were obtained from adults (ages 18-70) with primary insomnia participating in a randomized, double-

blind, placebo-controlled, multi-center study investigating 20 mg or 50 mg tasimelteon vs placebo in 322 patients over 5-weeks of treatment, using polysomnography measures of sleep. WGS was conducted with 30x read depth using whole blood samples. TST and LPS measures were obtained from daily sleep diary and PSG. GWAS using linear model was conducted on variants with MAF ≥0.01 using Plink with adjustment of the subject's age, sex, and the first 2 PCs. Top variants associated with LPS at baseline included loci within GRIK2 (rs11756558 beta=85.2 p-value=2.84e-8, MAF 0.02), in vicinity of GABRG2 and within TAFA1, PTPRD and RAPGEF5. GRIK2 - glutamate receptor, ionotropic, kinase 2 is known to be involved in a series of neurodevelopmental disorders as well as implicated in chronic fatigue syndrome. Top variants associated with TST at baseline were from loci within the DAB1 gene, as well as within BABAM2 and NCKAP5. The top variant was rs12070267 with a MAF of 0.02 (gnomAD all), where minor allele was associated with shorter TST (beta -84.5= 0.68, p=3.36e-7). We expanded the analysis with single gene tests which shows that DAB1 is significant, especially for MAF>0.05 variants (SKAT-O p=0.001). Interestingly, DAB1 variants were reported in the largest GWAS of insomnia using the UK biobank dataset. In this work, we identified potential insomnia loci associated with TST and LPS that further shed light upon understanding of the pathophysiology of insomnia disorder.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5089T: Genetics of postpartum depression in African American women

Authors:

C. DaCosta, C. M. Rose, S. Liu, D. C. Crawford; Case Western Reserve Univ., Cleveland, OH

Abstract:

Approximately 1 in 8 US women experience postpartum depression (PPD) after childbirth. PPD prevalence is higher among mothers <35 years and African Americans (AA) compared with mothers ≥35 years and other racial groups. The causes of PPD are unknown and few effective treatments are available. PPD risk factors include family history of psychiatric disorders, diagnosis of a psychiatric disorder pre-pregnancy, and social determinants of health. Twin studies suggest PPD is a complex disorder with a heritable component. Recent studies demonstrate an association between genetic risk scores (GRS) for major depressive disorder (MDD) and PPD. Most PPD genetic studies conducted to date include only European descent populations, and it is unclear if these findings are generalizable to AA. To address this gap, we accessed the Epidemiologic Architecture for Genes Linked to Environment dataset, a subset of

Vanderbilt University Medical Center's biorepository linked to de-identified electronic health records genotyped using the Illumina Metabochip. From the 15,863 adult patients with clinical data collected through 2011, 73% were AA. PPD cases were identified among females of reproductive age (12-45 years) using the first mention of a qualifying patient's International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) codes associated with pregnancy (V22.0, V22.1, V23.81, V23.82, V23.86, V23.89, V24.2, V28.0, V28.2, V28.9, V28.3, V82.9) and PPD (300.4, 309, 311). Identified cases (n=297) were majority AA (82%) with a mean age of 24.66 years. Identified non-cases (n=3,014) included women with pregnancy-associated ICD-9-CM codes who did not have PPD ICD-9-CM codes. Non-cases were majority AA (64%) with a mean age of 27.18 years. We queried the NHGRI-EBI GWAS Catalog and identified 415 MDD-associated (p<5x10-8) single nucleotide variants (SNVs). Of these, 17 were assayed on the Metabochip; 16 remained after quality control using PLINK. Individual-level unweighted (risk allele counting and summation) GRS were calculated. Tests of association for single SNV and unweighted GRS were limited to PPD AA cases and non-cases using logistic regression adjusted for age and PC1. In single SNV tests, intronic rs1363103 was associated at p=0.012 but was not significant after adjustment for multiple testing. AA PDD cases had lower average unweighted MDD GRS compared with non-cases, but this was not significant in a sex and PC1 adjusted logistic regression (OR=0.96, 95% CI: 0.90-1.01; p=0.123). Weighted GRS calculations, imputation, and access to larger datasets supported by 5R25GM075207 are ongoing to address present study limitations.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5090T: Genome sequencing identifies novel risk loci for multiple system atrophy

Authors:

A. Ray[1], R. Chia[2], E. Bayram[3,1], Z. Shah[1], J. Ding[2], P. Ruffo[2], M. Fujita[4], V. Menon[5], S. Saez-Atienzar[2], P. Reho[1], K. Kaivola[1], R. Walton[6], R. Reynolds[7], R. Karra[2], S. Sait[1], F. Akcimen[2], A. Fanciulli[8], Z. Alexander[9], W. Pirker[10], W. Meissner[11], F. Tison[11], M. Pellecchia[12], P. Mir[13], C. Proukakis[14], L. Parkkinen[15], M. Hu[15], C. Kobylecki[16], E. Garland[17], I. Litvan[3], R. Alcalay[4], S. Lubbe[18], V. Khurana[19], H. Kaufmann[20], C. Kim[21], M. Iba[21], E. Masliah[21], J. Troncoso[22], P. Pastor[23], J. Clarimon[24], A. Chiò[25], P. De Jager[26], Z. Wszolek[6], D. Dickson[6], B. Traynor[27], J. Gibbs[2], C. Dalgard[28], O. Ross[6], H. Houlden[29], S. Scholz[21], International MSA Genomics Consortium; [1]Natl. Inst. of Neurological Disorders and Stroke, Bethesda, MD, [2]Natl. Inst. on Aging, Bethesda, MD, [3]Univ. of California San Diego, San Diego, CA, [4]Columbia Univ. Irving Med. Ctr., New York, NY, [5]Columbia Univ. Med. Ctr., New York, NY, [6]Mayo Clinic, Jacksonville, FL, [7]Univ. Coll. London,

London, United Kingdom, [8]Med. Univ. of Innsbruck, Innsbruck, Austria, [9]Med. Univ. of Vienna, Vienna, Austria, [10]Klinik Ottakring - Wilhelminenspital, Vienna, Austria, [11]CHU Bordeaux, Bordeaux, France, [12]Univ. of Salerno, Salerno, Italy, [13]Univ. de Sevilla, Universidad de Sevilla, Spain, [14]Univ. Coll. London, london, United Kingdom, [15]Univ. of Oxford, Oxford, United Kingdom, [16]The Univ. of Manchester, Manchester, United Kingdom, [17]Vanderbilt Univ. Med. Ctr., Nashville, TN, [18]Northwestern Univ., Chicago, IL, [19]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [20]New York Univ. Sch. of Med., New York, NY, [21]NIH, Bethesda, MD, [22]Johns Hopkins Univ. Med. Ctr., Baltimore, MD, [23]The Germans Trias i Pujol Res. Inst. (IGTP), Badalona, Barcelona, Spain, [24]Univ.t Autònoma de Barcelona, Barcelona, Spain, [25]Univ. of Turin, Turin, Italy, [26]Columbia Univ Med Ctr, New York, NY, [27]Natl. Inst Aging, Bethesda, MD, [28]Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [29]Inst Neurology, London, United Kingdom

Abstract:

Multiple system atrophy (MSA) is a fatal neurodegenerative disease characterized by abnormal α-synuclein accumulation in oligodendroglial cells. MSA typically manifests with parkinsonism, cerebellar ataxia, and dysautonomia with sex differences for clinical profile, and results in death within 6-10 years after disease onset. Due to its rarity and clinical heterogeneity, the genetic landscape of MSA remains elusive, with limited focus on sex-specific risk factors. To address this critical knowledge gap, we performed whole-genome sequencing in 888 European-ancestry MSA cases and 7,128 neurologically healthy controls to define the genetic architecture of this understudied disease.

After stringent quality control steps, we included approximately 9.2 million autosomal variants in the genome-wide association analysis, under additive and recessive logistic regression models. Under the additive model, we discovered a genome-wide significant signal - in *GAB1* (odds ratio (OR)=1.37, 95% CI=1.23-1.52, p-value=1.43×10-8) and near *lnc-LRRC49-3* (OR=2.30, 95% CI=1.71-3.08, p-value=3.40 × 10-8). In the recessive model, we identified two additional risk loci - in *TENM2* (OR=7.40, 95% CI=3.74-14.67, p-value=9.73×10-9), and in *RABGEF1* (OR=1.95, 95% CI=1.55-2.46, p-value=1.65×10-8).

To examine the effect of the local genetic architecture on gene expression, we performed transcriptome-wide association and colocalization analyses using the Genotype-Tissue Expression eQTL dataset. *USP38-DT*, an *lnc*-RNA upstream of *GAB1*, was predicted to mediate MSA risk by decreased gene expression in the hippocampus (p-value=4.38×10-6, Z=-4.59). Reduced expression of *KCTD7* or *lnc-KCTD7-2* in twelve brain regions was predicted to increase MSA risk, with the lowest p-value for *lnc-KCTD7* was identified in the putamen (p-value=1.11×10-6, Z=-4.87).

We also tested overall and sex-stratified X-chromosome-wide associations in the 888 MSA cases and jointly called variants from 3,018 controls. We identified a significant 290 kb risk region only

in female MSA cases. Our genetic findings may provide deeper insights into the genetic underpinnings of sex-specific autonomic manifestations in MSA.
In conclusion, our genomic analyses revealed four novel autosomal risk loci for MSA susceptibility and an X-chromosome-specific risk region in females. This study advances our understanding of the genetic risk in this fatal neurodegenerative disease and provides a key genomic resource for developing targeted treatments emphasizing sex-specific approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5091T: Genome-based Analysis of Patients with IEI Increases Diagnostic Yield by 53%

Authors:

N. Di Donato[1,2], C. Du[1], T. Niehues[3], B. Grimbacher[4,5,6], S. Goldacker[4], K. Warnatz[4], G. Sogkas[7], U. Baumann[8], B. Auber[1], S. von Hardenberg[1]; [1]Dept. of Human Genetics, Hannover Med. Sch., Hannover, Germany, [2]Cluster of Excellence 2155 to Hanover Med. Sch., Hannover, Germany, [3]Ctr. for Child and adolescent Hlth., Helios Hosp. Krefeld, Avcademic Hosp. of the RWTH Aaachen, Krefeld, Germany, [4]Divison of Immunodeficiency, Dept. of Rheumatology and Clinical Immunology, Med. Ctr. - Univ. of Freiburg, Faculty of Med., Univ. of Freiburg, Freiburg, Germany, [5]German Ctr. for Infection Res., Satellite Ctr. Freiburg, Freiburg, Germany, [6]Cluster of Excellence 2155 to Hanover Med. Sch., Satellite Ctr. Freiburg, Hannover, Germany, [7]Inst. of Immunology, Hannover Med. Sch., Hannover, Germany, [8]Dept. of Pediatric Pulmonology and Neonatology, Hannover Med. Sch., Hannover, Germany

Abstract:

Background/Objectives: Whole exome sequencing (WES) and whole genome sequencing (WGS) have the significant advantage over panel analysis of enabling the detection of non-coding and structural variants. Additionally, WGS and WES allow for expanded phenotype- and genotype-driven evaluation, respectively. Therefore, we implemented short-read WGS for the genetic analysis of inborn errors of immunity (IEI).
Methods: 160 bp paired-end WGS was performed with an insert size of 350-400 bp on an Illumina NovaSeq 6000 or X Plus instrument. The analysis of WGS data was initially panel-based, followed by both a phenotype- and genotype-driven approach.
Results: This study explores the genetic underpinnings of IEI by analyzing 491 independent consecutive index patients (48% pediatric and 52% adult). The overall diagnostic yield was 10% (49 patients) with substantial variation across age groups (15% for pediatric and 5% for adult patients). Notably, 17 of these 49 positive cases harbored pathogenic variants outside of the

established IEI panel. In at least two of these patients, the diagnosis directly led to a curative hematopoietic stem cell transplantation. Furthermore, the diagnostic rate varied significantly between male and female patients (male 15%, female 5%) due to X-linked diseases. The difference become even more obvious in adult patients, where the definitive diagnosis was made in 10% of male patients versus only 2,4% of female patients.
Conclusion: These findings underscore the genetic heterogeneity of IEI and highlight the importance of comprehensive genomic analyses beyond traditional panel analysis. Our results suggest that expanding the scope of genetic testing could improve the diagnostic yield and understanding of IEI, ultimately enhancing patient care and targeted therapies.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5092T: Genome-wide association studies and integrative analysis with multi-omics data reveal novel loci and their molecular mechanisms for polyunsaturated, monounsaturated, and saturated fatty acids

Authors:

Y. Sun, H. Xu, K. Ye; Univ. of Georgia, Athens, GA

Abstract:

Background: Polyunsaturated fatty acids (PUFAs) are associated with decreased risks and mortality of various diseases. The circulating levels of fatty acids (FAs) exhibit heterogeneity across individuals due to genetic and environmental influences. Although previous genome-wide association studies (GWAS) have identified loci associated with FAs, the biological implications of these genetic associations remain largely unexplored. **Methods:** Here, we conducted GWAS to identify additional loci for 19 fatty acid (FA) traits, quantified by nuclear magnetic resonance spectroscopy, in UK Biobank participants of European ancestry (N = 239,268) and five other ancestries (N = 508 - 4,663). We leveraged the GWAS findings to characterize genetic correlations and colocalized regions among FAs, explore sex differences, examine FA loci influenced by lipoprotein metabolism, and apply statistical fine-mapping to pinpoint putative causal variants. The integrative analysis of GWAS and multi-omics quantitative trait loci (QTL) provides a better understanding of the genetic impacts on intermediate molecular phenotypes. **Results:** Altogether, we identified 215 significant loci for PUFA traits in European participants, 163 loci for monounsaturated fatty acids (MUFAs), and 119 loci for saturated fatty acids (SFAs), including 70 novel loci for PUFAs, 61 for MUFAs, and 54 for SFAs. Of these significant loci, one locus (around genes *GSTT1/2/2B*) overlapped with QTL signals for all six molecular phenotypes, including gene expression, protein abundance, DNA methylation,

splicing, histone modification, and chromatin accessibility. Across FA traits, we observed that 65% of GWAS loci overlapped with QTL signals for at least one molecular phenotype. **Conclusions:** Our investigation identifies additional novel genetic loci and provides the first evidence uncovering molecular mechanisms underlying FA-relevant variants. These findings offer insights that could facilitate patient stratification for precision nutrition and the identification of novel therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5093T: Genome-wide Association Study of Long COVID by Predictive Phenotyping of EHR Data using Deep Learning: Transferring Trained Model from National COVID Cohort Collaborative to All of Us Research Platform Enables Genetic Discovery

Authors:

A. Naseri, H. Yan, L. Rasmy, D. Zhi; Univ. of Texas Hlth.Sci. Ctr., Houston, TX

Abstract:

As the COVID-19 pandemic subsides, many patients continue to experience the long-term effects of their initial infection, known as “Long COVID” or post-acute sequelae of COVID-19 (PASC). Understanding and diagnosing complex and novel diseases such as PASC continue to be challenging making it difficult to identify genetic risk factors that contribute to the susceptibility or severity of the disease. Although there is an ICD code for PASC conditions (U09.9), it may not be comprehensive enough to record all complications and may not be accurately recorded for every patient. To address the missing or unrecorded PASC diagnosis codes, we trained a Recurrent Neural Network (RNN) model to predict patient risk for PASC using their longitudinal health record considering the time difference between events. We defined the PASC cases as COVID patients who developed any PASC symptoms or conditions after 28 days of their first positive test result or diagnosis. We trained our model using 24,774 identified PASC patients available in N3C. Our model showed a discriminative accuracy of 81.8% measured by the area under the receiver operating characteristic curve (AUROC). Our model achieved an AUROC of 86.8% in All of Us data using a test cohort of 2752 patients (after fine-tuning on 9634 patients) demonstrating the transferability and generalizability of our model. Utilizing both genetic and EHR data, alongside the transferred RNN model predictions, we conducted GWAS to uncover potential genetic links to the identified patient characteristics. We have identified new genetic regions associated with PASC using whole-genome sequencing of 9475 individuals in All of Us

and AI-generated scores. These regions have very low minor allele frequencies in individuals of European ancestry, but significantly higher allele frequencies in African and American ancestries. 10 genetic loci passed the Bonferroni threshold in our analysis. One of the loci (chr11:76613341:C:CAAAAAAA, p=8.15E-09) is located near a regulatory T-cells (Tregs) marker (LRRC32) that has already been suggested to play a role in the severity of COVID and PASC. Our AI-generated scores identify genome-wide significant loci missed by using only diagnostic labels, enriching genetic discovery for this new and complex disease.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5094T: Genome-wide meta-analysis of Alzheimer's disease reveals genes causally involved in disease risk by linking to gene expression loci in whole blood and brain.

Authors:

D. Baird, D. Chasioti, C-Y. Chen; Biogen, Cambridge, MA

Abstract:

Alzheimer's disease (AD) has a complex genetic etiology, pinpointing the causal genes is a key yet challenging step in identifying therapeutic targets. Here, we combine data on gene expression quantitative trait loci (eQTLs) with AD meta-analysis using a Mendelian Randomization (MR) and colocalization framework to identify genes underlying disease susceptibility.

To genetically proxy for gene expression levels, we obtained genome-wide significant cis-eQTLs from whole blood in eQTLGen and brain cortex tissue in MetaBrain (MB). For the AD outcome, we prepared a meta-analysis by pooling together the GWAS summary statistics available from Wightman et al. and FinnGen to obtain 624,220 European individuals (n=49,608 cases and 574,612 controls). Palindromic eQTLs were removed before LD clumping was conducted using a UKBB European reference panel. Single SNP Wald ratio test was then performed on the eQTL-disease pairs available for the MR. We selected the MR findings that reached significance threshold of $P<1x10^{-6}$ and then undertook both coloc-abf and coloc-Susie analysis on the 1Mb region around the primary eQTL to verify if gene expression and AD share the same causal association in the genomic region (PP4 > 0.7 for either coloc-abf or coloc-susie).

A total of 41,336 eQTL-AD pairs (15,469 genes) were obtained from eQTLGen and 14,866 pairs (11,060 genes) from MB, resulting in a total of 18,374 genes being analyzed where 8,155 genes were available for both datasets. A total of 62 genes reached statistical significance in the MR, of which 21 genes were colocalized in eQTLGen and 9 in MB (none colocalized in both). Higher

AD risk was associated with increased expression for 11 genes (downregulation targets) and decreased expression for 19 genes (upregulation targets). Magnitudes of effects for the downregulation targets found in the eQTLGen ranged between 1% to 11% for a 1 SD change in gene expression in whole blood, and for MB ranged between 3% to 35% for 1 SD change in expression in brain cortex. For example, the MB MR found evidence for upregulation of *GRN* estimating that a 1 SD change in the gene's expression in brain cortex would reduce AD odds by ~4%. Mutations in *GRN* have previously been linked to frontotemporal dementia risk (OMIM 607485).
Our MR and coloc study detected a total of 30 genes where changes in the expression levels modulate AD risk, which provide a set of potential candidates for therapeutic intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5095T: Genomic analysis identifies risk factors in restless legs syndrome

Authors:

F. Akcimen; Natl. Inst. on Aging, Bethesda, MD

Abstract:

Restless legs syndrome (RLS) is a neurological condition that causes uncomfortable sensations in the legs and an irresistible urge to move them, typically during periods of rest. The genetic basis and pathophysiology of RLS are incompletely understood. Here, we present a whole-genome sequencing and genome-wide association meta-analysis of RLS cases (n = 9,851) and controls (n = 38,957) in three population-based biobanks (All of Us, Canadian Longitudinal Study on Aging, and CARTaGENE). Genome-wide association analysis identified nine independent risk loci, of which eight had been previously reported, and one was a novel risk locus (*LMX1B*, rs35196838, OR = 1.14, 95% CI = 1.09-1.19, *p*-value = 2.2×10^{-9}). A genome-wide, gene-based common variant analysis identified *GLO1* as an additional risk gene (p-value = 8.45×10^{-7}). Furthermore, a transcriptome-wide association study also identified *GLO1* and a previously unreported gene, *ELFN1*. A genetic correlation analysis revealed significant common variant overlaps between RLS and neuroticism (rg = 0.40, se = 0.08, p-value = 5.4×10^{-7}), depression (rg = 0.35, se = 0.06, p-value = 2.17×10^{-8}), and intelligence (rg = -0.20, se = 0.06, p-value = 4.0×10^{-4}). Our study expands the understanding of the genetic architecture of RLS and highlights the contributions of common variants to this prevalent neurological disorder. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5096T: Genomic Insights into Autism Spectrum Disorder: Comprehensive Analysis of Genetic Variants in Indian Trios

Authors:

D. Manjegowda[1], S. Shankaran Chettiar[2]; [1]Dayananda Sagar Univ., Bangalore, India, [2]Genexplore Diagnostics & Res. Ctr., Ahmedabad, India

Abstract:

Autism Spectrum Disorder (ASD) is a multifaceted neurodevelopmental disorder marked by difficulties in social interaction, communication challenges, and restricted repetitive behaviors. The complexity and heterogeneity of ASD suggest that multiple genetic and environmental factors contribute to its onset. However, the specific genetic variants and the overall genomic landscape associated with ASD remain largely uncharted.To identify potential genetic contributors to ASD, we conducted an extensive genomic analysis. We recruited 50 Indian trios, each consisting of parents and their clinically diagnosed ASD index children aged 8 to 12 years. Genomic DNA from blood samples was prepared and analyzed using microarray analysis and Next-Generation Sequencing (NGS). The resulting data were subjected to a series of integrative analyses, including bioinformatics, statistical, computational biology, and pathway analyses, to uncover the relationship between genomic variants and ASD. Our aim was to elucidate the pathophysiology of ASD and identify new therapeutic targets for its prevention.Microarray analysis detected copy number variations (CNVs) with duplications smaller than 1MB in critical chromosomal regions such as 6p21.32, 7p21.2, 8p23.3, 14q12.14, and Xp22.11. Whole-exome sequencing (WES) revealed single nucleotide variants (SNVs) in several genes, including CDK5RAP2, CC2D1A, AP4M1, SYNGAP1, STXBP1, ACSL4, ASTN1, MBD5, CNTNAP2, SCN8A, SATB1, CREBBP, SHANK3, SHANK2, GNG13, and PPP3CB. While many of these variants are novel and may contribute to ASD development, a few have been previously linked to synaptic function, neuronal connectivity, and neurodevelopmental processes.Pathway and functional enrichment analyses indicated a disruption in glutamatergic synaptic transmission, a crucial process for brain excitatory neurotransmission. Abnormalities in this pathway may be fundamental to ASD's pathophysiology. By integrating clinical data with comprehensive genomic analysis in trios, we can more precisely identify genetic variations associated with ASD. These findings could significantly inform genetic counseling, enabling the development of targeted therapeutic interventions and personalized treatment strategies for individuals with ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5098T: Genotype specific genetic architecture of alcohol consumption identified by a genotype-stratified GWAS in Japanese people

Authors:

M. Nakatochi[1], Y. Koyanagi[2], S. Namba[3], A. Narita[4], T. Kawaguchi[5], H. Ikezaki[6], K. Kinoshita[4], Y. Sutoh[7], A. Shimizu[7], S. Nakano[8], T. Yamaji[8], K. Wakai[1], F. Matsuda[5], Y. Okada[3], K. Matsuo[2]; [1]Nagoya Univ., Nagoya, Japan, [2]Aichi Cancer Ctr. Res. Inst., Nagoya, Japan, [3]Osaka Univ. / The Univ. Tokyo / RIKEN, Osaka, Japan, [4]Tohoku Univ., Sendai, Japan, [5]Kyoto Univ., Kyoto, Japan, [6]Kyushu Univ., Fukoka, Japan, [7]Iwate Med. Univ., Yahaba, Japan, [8]Natl. Cancer Ctr. Inst. for Cancer Control, Tokyo, Japan

Abstract:

An East Asian-specific variant on aldehyde dehydrogenase 2 (*ALDH2* rs671, G &gt A) is the major genetic determinant of alcohol consumption. We performed an rs671 genotype-stratified genome-wide association study meta-analysis of alcohol consumption in 175,672 Japanese individuals to explore gene-gene interactions with rs671 behind drinking behavior. The analysis identified three genome-wide significant loci (*GCKR*, *KLB*, and *ADH1B*) in wild-type homozygotes and six (*GCKR*, *ADH1B*, *ALDH1B1*, *ALDH1A1*, *ALDH2*, and *GOT2*) in heterozygotes, with five showing genome-wide significant interaction with rs671. We performed lookups of our identified lead SNPs in the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) data, comprising 941,280 individuals of European ancestry. While all three loci (*GCKR*, *KLB*, and *ADH1B*) identified in the rs671 GG-only analysis showed a genome-wide significant association, we did not observe any association in the other loci identified in the rs671 GA-only analysis, including two loci (*ALDH2* and *GOT2*), which have not been reported in the context of alcohol consumption.We also computed cross-ethnic genetic-impact correlation (ρ_{gi}) between Japanese and European populations using Popcorn, where ρ_{gi} represents the correlation coefficient of the population-specific allele variance-normalized SNP effect sizes. The genetic-impact correlation was defined as the correlation of effect sizes after genotypes were normalized to have mean 0 and variance 1. While Japanese individuals with the rs671 GG genotype may share a similar genetic architecture for alcohol consumption with Europeans [ρ_{gi} = 0.975, P (ρ_{gi} = 1) = 0.905], results indicated that Japanese individuals with the rs671 GA genotype may genetically differ from Europeans in terms of alcohol use [ρ_{gi} = 0.287, P (ρ_{gi} = 1) = 0.0003].Our stratified GWAS revealed that the genotype-specific genetic architecture of alcohol consumption.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5099T: Gestational diabetes mellitus shares genetic risk factors with type 1 and type 2 diabetes and is predicted by diabetes polygenic scores

Authors:

J. Winters[1,2,3], E. Jasper[4,3], J. Hellwege[4,2,5], T. Edwards[4,6], D. Velez Edwards[4,3]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Genetics Inst., Vanderbilt Univ., Nashville, TN, [3]Div. of Quantitative Sci., Dept. of Obstetrics and Gynecology, Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Div. of Epidemiology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Gestational diabetes mellitus (GDM), hyperglycemia or glucose intolerance arising during pregnancy, can result in both maternal and fetal morbidity and chronic complications. As the most prevalent metabolic disorder in pregnancy, the global incidence of GDM continues to increase. Multiple studies link GDM risk with type 2 diabetes mellitus (T2DM), but few studies have described shared genetic architecture with type 1 (T1DM). In this retrospective case-control study, our objective was to determine whether available polygenic risk score (PRS) models for T1DM (PGS002025) and T2DM (PGS003867) were associated with an individual's risk for GDM. We utilized linked electronic health records and genotype data from Vanderbilt University Medical Center's BioVU. Inclusion criteria were biologically female, pregnant persons, at least 18 years of age, and Non-Hispanic White (NHW) or Non-Hispanic Black (NHB) self-reported race and ethnicity. Cases were defined with any record of ICD-9 or ICD-10 code(s) relating to GDM and pregnancy, while controls consisted of any record of pregnancy-related ICD code(s). Both groups were created with and without the exclusion of diabetes ICD code(s), resulting in 382 cases and 5314 controls (3950|1746; NHW|NHB) and 1014 cases and 5594 controls (4613|1995; NHW|NHB), respectively.[GU1] Associations between each PRS and GDM diagnosis were modeled using logistic regression with median BMI, age at first GDM diagnosis (cases) or last pregnancy classification (controls), and top 10 principal components of ancestry as covariates. T1DM PRS was validated in NHW individuals when diabetics were included but exhibited no significant ($p<0.05$) results upon exclusion. The T2DM PRS was validated in both NHW and NHB individuals when diabetics were included but only displayed significant associations with GDM in NHW individuals after exclusion (AUC=0.73; OR=14.58; SE=0.82). We observed significant genetic correlations between GDM and both T2DM (RG=0.76; SE=0.06) and T1DM (RG=0.73; SE=0.12). These results suggest that other diabetic PRS models could be useful predictors of GDM, particularly regarding T2DM and NHW individuals. Our findings indicate that

genetic correlations exist between GDM and both T1DM and T2DM, demonstrating that GDM may share genetic risk factors with multiple types of diabetes.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5100T: GWAS in Pacific Islanders identified established and novel risk loci for four serum lipid measures

Authors:

W-C. Hsueh[1], S. Safabakhsh[2], S. Kobes[1], E. Vazquez Arreola[1], J. E. Curran[3], R. L. Hanson[1]; [1]NIDDK, Phoenix, AZ, [2]Micronesian Inst. for Disease Prevention and Res., Sinajana, GU, [3]Univ. of Texas Rio Grande Valley, Brownsville, TX

Abstract:

Findings from epidemiological studies suggest that Pacific Islanders (PI) may be at a higher risk for metabolic and cardiovascular diseases, yet their underlying genetic and non-genetic causes remain unclear. We conducted a community-based study aimed at identifying genetic risk factors for diabetes and related traits in PI from Guam and Saipan. Our sample of 2,022 subjects (mean age=43 years, 54% female, and 41% with diabetes) were genotyped using the OmniExpress Core Exome array (Illumina, San Diego, CA) which yielded data on 260,761 variants. We further performed whole genome imputation analysis using a reference panel comprised of 2,504 samples from Phase 3 of the 1K Genome Project and 40 PI from our own study who had undergone whole genome sequencing. Using the imputed data of 8.4 million SNPs which passed quality control measures, we conducted genome-wide association studies (GWAS) for 4 serum lipid measures including total cholesterol, high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C, calculated by the Friedewald equation), and triglycerides (TG). The lipid levels were log transformed and effects of other notable risk factors (age, sex, diabetes status, and the first 4 genetic principal components, to account for ancestry) were adjusted. P values for SNP effects were further adjusted using genomic control approach.

We identified significant SNP associations at the genome-wide level (p<E-8) with all 4 traits (n=1,320-1,602). The strongest association was observed with LDL-C (rs1065853, β=-0.62 s.d. unit (SDU), p=4.3*E-13 on chr 19 near *APOE*). A SNP in the *APOA5*/*APOC3* region on chr 11 was associated with TG (rs651821, β=-0.29 SDU, p=E-11). An *APOE* nsSNP was associated with total cholesterol (rs7412, β=-0.46 SDU, p=7.6*E-9). For HDL-C, 2 SNP associations were of borderline significance (rs3908772, β=-0.81 SDU, p=4.9*E-8 on chr 1 in *DAB1*; and rs111405191, β=0.42 SDU, p=5.7*E-8 on chr 2 in *THSD7B*). In summary, we identified 2 well-established genomic

regions in other populations (*APOE* for LDL-C/total cholesterol and *APOA5/APOC3* for TG), which suggests that these loci also have strong effects in Pacific Islanders with a diverse genetic background. Two novel genetic associations for HDL-C will require further studies with larger samples. In addition, a *CETP* nsSNP (rs1597000001) was recently reported to alter lipid levels in Polynesians; and we also observed a suggestive SNP association with HDL-C in the same gene (rs12720922, an intronic SNP, β=-0.24 SDU, p=2.5*E-6). Through a network of studies in PI and Native Hawaiians, we plan to conduct replication and meta-analyses to refine and validate these findings.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5101T: GWAS uncovers the neuronal contribution to multiple sclerosis susceptibility ⭐

Authors:

L. Zeng[1], A. Khan[1], M. Fujita[2], F. Zipp[3], G. Wang[1], K. Kirlyuk[1], P. De Jager[4]; [1]Columbia Univ., New York, NY, [2]Columbia Univ. Irving Med. Ctr., New York, NY, [3]Univ. Med. Ctr. of the Johannes Gutenberg Univ. Mainz, Mainz, Germany, [4]Columbia Univ Med Ctr, New York, NY

Abstract:

Multiple Sclerosis (MS) is a chronic, inflammatory and neurodegenerative disease affecting the central nervous system. Recent GWAS have identified many risk loci, but pinpointing the causal genes and variants within these loci and understanding the underlying mechanisms behind MS susceptibility remain major challenges. Prior work suggested that genetic susceptibility to MS primarily manifested as altered gene expression in a variety of peripheral immune cells.Here, we enlarged the latest MS GWAS, including 19,865 MS cases and 623,043 controls of European ancestry (EUR), and we identified 74 novel susceptibility loci that reached genome-wide significance. We also performed a multi-ancestry meta-analysis, with 614 MS cases and 62,044 controls of African American ancestry (AFR), as well as 352 MS cases and 44,133 controls of Latino American ancestry (AMR).Integrating single cell-derived cis-eQTL data from both peripheral blood mononuclear cells and human brain, we implicate 73 genes whose expression is altered by the MS risk variant: significant colocalization of GWAS & eQTL results (PP4>0.8). Consistent with prior results, CD 4+ T cells exhibited the most colocalization results, followed by CD8+ T cells. However, brain snucRNA- for the first time and found evidence of colocalization in multiple CNS cell types, with inhibitory neurons harboring the most colocalized genes with MS risk, followed by excitatory neurons. Our results highlight specific targets such as *AHI1*, *CD37* and *IFITM3* across multiple cell types, and *IL7* and *STAT3* only in inhibitory

neurons.PRS-CSx showed that individuals within the top 1% of PRS had a more than 7-fold increased risk of MS in EUR (95% CI: 4.4-12.0, P=1.6×10^{-14}); however, the PRS also worked to some degree in other populations, with a 9-fold higher risk in AMR (95% CI: 2.8-35.5, P=4.4×10^{-4}) and a 2.7-fold increased risk in AFR (95% CI: 2.8-35.5, P=0.06). While the current PRS is probably over-fitted for European populations, it provides a useful initial tool for non-European populations as well.Together, these results advance our understanding of the biological etiology of MS, refocusing our efforts in understanding the onset of the disease to include the brain. While most loci have functional consequences in a variety of immune cell types, our study prioritizes understanding the unsuspected neuronal contribution to the onset of MS. They alter our conceptualization and approach to primary prevention in MS, which will have to include interventions targeting central nervous system pathways to develop potent, preventive therapeutic approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5102T: GWAS-informed data integration and non-coding CRISPRi screen illuminate genetic etiology of bone mineral density

Authors:

M. Conery[1,2], J. A. Pippin[2], Y. Wagley[3], K. Trang[2], M. C. Pahl[2], D. A. Villani[4], L. J. Favazzo[4], C. L. Ackert-Bicknell[4], M. J. Zuscik[4], E. Katsevich[1], A. D. Wells[5], B. S. Zemel[2], B. F. Voight[1], K. D. Hankenson[3], A. Chesi[1], S. F. A. Grant[5,1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Univ. of Michigan Med. Sch., Ann Arbor, MI, [4]Univ. of Colorado Anschutz, Aurora, CO, [5]Children's Hosp. of Philadelphia, PHILADELPHIA, PA

Abstract:

Over 1,100 independent signals have been identified with genome-wide association studies (GWAS) for bone mineral density (BMD), a key risk factor for mortality-increasing fragility fractures; however, the effector gene(s) for most remain unknown. Informed by a variant-to-gene mapping strategy implicating 89 non-coding elements predicted to regulate osteoblast gene expression at BMD GWAS loci, we executed a single-cell CRISPRi screen in human fetal osteoblast 1.19 cells (hFOBs). The BMD relevance of hFOBs was supported by heritability enrichment from cross-cell type stratified LD-score regression involving 98 cell types grouped into 15 tissues. 23 genes showed perturbation in the screen, with four (*ARID5B*, *CC2D1B*, *EIF4G2*, and *NCOA3*) exhibiting consistent effects upon siRNA knockdown on three measures of osteoblast maturation and mineralization. Lastly, additional heritability enrichments, genetic correlations, and multi-trait fine-mapping revealed unexpectedly that

many BMD GWAS signals are pleiotropic and likely mediate their effects via non-bone tissues that warrant attention in future screens.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5103T: Heterogeneity in the Genetic Architecture of AD across Ancestry Groups

Authors:

P. Benchek[1], C. Reitz[2], B. Kunkle[3], S-T. Dietrich[3], N. Ray[2], K. Hamilton-Nelson[3], G. Beecham[3], J. Vance[3], G. S. Byrd[4], J. Thulaseedhara[2], A. Naj[5], G. Tosto[2], R. Akinyemi[6], J. Haines[1], M. Pericak-Vance[3], W. Bush[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Columbia Univ., New York, NY, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]Wake Forest Univ. Sch. of Med., Winston-Salem, NC, [5]Univ. of Pennsylvania, Philadelphia, PA, [6]Univ. of Ibadan, Ibadan, Nigeria

Abstract:

Most assessments of the genetic component of Alzheimer's disease have focused on Non-Hispanic White (NHW) individuals. However, African American (AA), Asian American (AsA), and Hispanic/Latino (HL) populations account for 13.6%, 6.1%, and 18.9% of the US population respectively. Epidemiological evidence implies substantial differences in AD across populations (e.g. HL are twice as likely as NHW to have AD and the incidence of new AD cases in Caribbean HL families is three times that in NHW families). There are also paradoxical differences in the effect size of *APOE* by ancestry, with *APOE*-e4 risk generally lower in populations with increased overall disease prevalence. Other genes (*SORL1*, *ABCA7*, and *ACE*) similarly have variants that contribute to AD risk in an ancestry-specific manner. By exploring ancestry-specific heritability estimates we sought to understand the extent to which we expect the genetic architecture of AD is shared and differs by ancestry. We used the Alzheimer's Disease Genetics Consortium multi-ancestry TOPMED-imputed dataset to perform GCTA-based analyses to estimate the heritability of AD in cohorts ascertained for AA (N=6409), AsA (N=3216), NHW (N=32,479) and HL (N=1758) participants. Consistent with the approach of Guo et al, 2021, we extracted overlapping SNPs that were common (MAF > 1%) and well-imputed ($R^2 > 0.8$) across all populations, and further restricted variants to those well-genotyped in HapMap 3. We calculated a pairwise, joint genetic relatedness matrix (GRM), standardizing with respect to ancestry specific allele frequencies, and conducted a bivariate GREML analysis, adjusting for sex, age, *APOE*-e4 dosage and principal components to assess the degree of genetic heterogeneity in AD across the ancestry pairs. Estimates of genetic correlation (r_g) were moderate for NHW and AA (r_g=0.50, se=0.30, p {r_g = 1} = 0.14) and for NHW and AsA (r_g=0.56, se=0.18, p {r_g = 1} = 0.04).

Estimates were lower for AA and AsA (r_g =0.22, *se*=0.62, *p* {r_g = 1} = 0.23), for AA and HL (r_g =0.40, *se*=0.59, *p* {r_g = 1} = 0.25), for HL and AsA (r_g =-0.28, *se*=0.52, *p* {r_g = 1} = 0.13) and for NHW and HL (r_g =0.31, *se*=0.15, *p* {r_g = 1} =8x10$^{-4)}$. These estimates do not appear to be highly sensitive to differences in linkage disequilibrium or Fst. Our findings suggest that there is a low to moderate amount of overlap in the genetic architecture of AD across ancestral populations. Thus, there are likely ancestry-specific genetic contributors to AD for these populations. We are in the process of including participants from the African Dementia Consortium as an additional contrast, and continued efforts to increase diversity in future genetic studies is critical.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5104T: Higher gene expression of dynein heavy chains in the dorsolateral prefrontal cortex predict lower neuropathology and better cognitive outcomes in individuals with Alzheimer's disease

Authors:

E. Lucero[1], T. Brunetti[1], H. Chial[1], H. Potter[2], C. Gignoux[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ Colorado Anschutz Med. Ctr., Aurora, CO

Abstract:

Motor proteins play a key role in neuronal functions and morphology that are important for learning and memory. We have previously reported that increased expression *KIF11/*Kinesin-5 overrides amyloid-mediated effects on dendritic spine density and long-term potentiation in a mouse model of Alzheimer's disease (AD), effectively maintaining cognitive function in the face of amyloid beta pathology. Here, we evaluated the association of key AD phenotypes with mRNA expression levels of a select set of Dynein motor proteins.Here, We utilized measurements of gene expression, AD neuropathology burden, and cognition provided by the ROS/MAP study to determine whether an association exists between AD phenotypes and expression of genes for cytoplasmic and axonemal dynein heavy chains. Neuropathology burden was determined through immunohistochemistry. Z-scores from the raw scores of 19 cognitive tests were used to determine global cognition. Neuropathology burden and global cognition measurements were provided by Rush Alzheimer's Disease Center. Measurements of gene expression in the dorsolateral prefrontal cortex (DLPFC) (n=634) were determined through an established RNAseq analysis pipeline (Logsdon et al., 2019, syn8456638; syn8456629) and made available to us through the Accelerating Medicines Partnership in Alzheimer's Disease Target Discovery and Preclinical Validation knowledge portal (syn3219045). Associations of gene expression with neuropathology and cognition were determined through multiple linear

regression and mixed effects models covarying for age, sex, education, and post-mortem interval. In participants with AD (CERAD score > 2), higher gene expression levels of *DYNC1H1* and *DNAH1* in the DLPFC predicted better cognitive performance longitudinally (p=0.03 and p=0.00197, respectively) and at the last visit prior to death (p=8.701e-05 and p=3.571e-05, respectively). Higher expression of *DYNC1H1* and *DNAH1* were also associated with lower amyloid pathology (p=0.0001257 and p=2.356e-07, respectively) and tau tangles (p=2.356e-07 and p=2.659e-09, respectively). Our finding that AD participants with higher expression levels of *DYC1H1* and *DNAH1* show reduced cognitive decline and decreased AD pathology suggest a potential beneficial effect of these motor proteins on AD progression. Based on the functions of these motor proteins, these findings warrant further studies to identify potential mechanisms underlying this association. Potential mechanisms include facilitation of higher rates of phagocytosis or neuronal lysosomal degradation of amyloid beta.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5105T: Identification of a rare missense variant in *VSX2* as a novel risk locus for retinal detachment ⭐

Authors:

V. Ravanmehr[1], D. Brock[1], F. Hu[2], Q. Wang[3], S. Petrovski[4], R. Dhindsa[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Astrazeneca, Cambridge, United Kingdom, [3]AstraZeneca, Chapel Hill, NC, [4]AstraZeneca, Fulbourn, United Kingdom

Abstract:

Retinal detachment (RD) is an emergency that can cause blindness if not treated promptly. While surgical treatment can achieve reattachment of the retina in 90% to 96% of cases, it is preferable to focus on the prevention of RD. Genetic identification of RD can reveal molecular events causing a detached retina and highlight shared risks among certain eye conditions. However, identifying genetic variants contributing to RD has been challenging due to the small sample size of diagnosed cases. So far, most efforts for finding genetic variants associated with RD have relied on genome-wide association studies (GWAS), which focus on common variants. Nevertheless, common variants tend to have small effect sizes and can be difficult to map to causal genes. Here, we analyzed whole-genome sequencing data from 7,276 cases and 236,741 controls in the UK Biobank to perform gene-level collapsing analysis and a variant-level exome-wide association study (ExWAS). There were two study-wide significant associations in the collapsing analysis. The first hit was *COL2A1* ($p=3.45x10^{-14}$; OR=162.9, 95% CI [35.68, 743.65]), which is a well-established risk factor for Stickler Syndrome, one of the most common and

strongest genetic risk factors for RD. The second significant hit emerged from the gene *VSX2* ($p = 2.39x10^{-10}$; OR = 2.7846; 95% CI [2.1, 3.68]), representing a novel association with RD. The ExWAS also detected several positive control loci, including *FAT3* and *RDH5*, previously discovered via GWAS. However, a single missense variant in *VSX2* was also significantly associated with RD ($p=9.29x10^{-10}$; OR = 5.9; 95% CI [3.7, 9.4]; p.Glu218Asp). This variant is quite rare in the European population, with a control allele frequency of 0.0002. We replicated this association in FinnGen (release 10), in which a separate missense variant (p.Pro45Arg) was also nominally associated with traction RD (p=0.0099; OR=25.102) among 239 cases and 376,650 controls. *VSX2* encodes a homeobox protein originally described as a retina-specific transcription factor. Autosomal recessive mutations in *VSX2* are associated with a substantially more severe phenotype, characterized by microphthalmia, juvenile cataract and iris abnormalities. To the best of our knowledge, heterozygous mutations in this gene have never been associated with retinal detachment or other diseases. Overall, this work highlights a novel risk factor for RD and provides a prime example of how analyzing UK Biobank data can broaden our understanding of complex dosage effects of Mendelian variants.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5106T: Identification of genes influencing rhinovirus and respiratory syncytial virus IgG responses in children and adults

Authors:

E. Bouzigon[1,2], R. Vernet[1,2], J. Wenta[1,2], A. Guillien[3,4], A. Boudier[3,4], A. Estermann[1,2], M. Mauclin[3,4], C. Linhard[1,2], F. Demenais[1,2], K. Niespodziana[5], R. Valenta[5,6], V. Siroux[3,4]; [1]INSERM, UMRS-1124, Paris, France, [2]Université Paris Cité, Paris, France, [3]INSERM U1209, Grenoble, France, [4]Univ. Grenoble Alpes, Grenoble, France, [5]Med. Univ. of Vienna, Vienna, Austria, [6]Karl Landsteiner Univ., Krems, Austria

Abstract:

Respiratory viral infections, especially from rhinovirus (RV) and respiratory syncytial virus (RSV) are associated with the occurrence and exacerbation of asthma. However, the genetic factors influencing the levels of RV- and RSV-specific antibodies in childhood or in adulthood have been poorly investigated. Moreover, it has been reported that innate and adaptive immune responses to numerous viruses may differ between men and women.
We aimed to identify genetic variants associated with RSV- and RV-specific IgG responses in both children and adults while accounting for differential sex effect.
Cumulative RV-specific IgG levels (species A, B and C) and IgG levels to RSV-G protein were

measured using micro-array technology applied to blood samples of subjects from the EGEA study on asthma: 507 samples of children (EGEA1) and 1153 samples of adults (EGEA2) of European ancestry. We conducted genome-wide association study of RSV- and RV-specific IgG levels stratified on sex by using a linear mixed model applied separately to children and adult individuals (totaling 292 boys and 215 girls; and 555 men and 598 women). Then, we performed two tests in children and adults separately: 1) a joint test of the genetic main effect (G) and the gene-by-sex interaction effect (GxSex) and 2) a test of GxSex interaction alone.
We identified four genome-wide significant loci associated with RSV- or RV-specific IgG levels (Joint test $P < 5x10^{-8}$). Two of them were detected in children: one was located in *GALNT15* and associated with RV-C-specific IgG levels independently of the sex ($P_{joint}= 1.0x10^{-9}$, $P_{male} = 2.3x10^{-8}$ and $P_{female} = 5.3x10^{-3}$), while the other one, located in *GABRG3*, was associated with RSV-specific IgG levels in females only ($P_{female} = 2.4x10^{-8}$ *versus* P_{male}=0.67 for rs12324881). In adults, one locus located in an intergenic region between *MAP3K7* and *EPHA7*, was associated with RV-A-specific IgG levels in females only ($P_{female} = 4.4x10^{-9}$ *versus* P_{male}=0.79 for rs768382) whereas the other one, located between *TRPA1* and *KCNB2*, was associated with RV-B-specific IgG levels in males only ($P_{male} = 5x10^{-8}$ *versus* P_{female} =0.34 for rs10542502).
By testing gene-by-sex interactions, we identified, for the first time, four genome-wide significant loci influencing RSV- and RV-specific IgG levels with three of them in a sex-specific manner. Multivariate analyses will enable to assess whether immune responses to various RSV and RV share common genetic determinants.
Funding: ANR-19-CE36-0005-NIRVANA

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5108T: Identifying common, rare, and structural variants associated with cardiac structure and function by whole genome sequencing

Authors:

S. Chung[1], T. Alkis[1], T. M. Bartz[2], E. Boerwinkle[1,3], J. A. Brody[4], S. Cheng[5], M. Fornage[6], M. E. Hall[7], B. A. Hidalgo[8], R. C. Kaplan[9,10], D. M. Lloyd-Jones[11], A. C. Morrison[1], B. M. Psaty[4,12], L. M. Raffield[13], C. J. Rodriguez[9,14], J. I. Rotter[15], S. J. Shah[16], A. Shah[17], K. D. Taylor[15], V. Xanthakis[18,19], R. S. Vasan[19,20], B. Yu[1]; [1]Human Genetics Ctr., Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Dept.s of Biostatistics and Med., Univ. of Washington, Seattle, WA, [3]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [4]Cardiovascular Hlth.Res. Unit, Dept. of Med., Univ. of Washington, Seattle, WA, [5]Smidt Heart Inst., Cedars-Sinai Med. Ctr., Los Angeles, CA, [6]Inst. of Molecular Med., McGovern Med.

Sch., the Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [7]Dept. of Med., Univ. of Mississippi Med. Ctr., Jackson, MS, [8]The Univ. of Alabama at Birmingham, Birmingham, AL, [9]Dept. of Epidemiology and Population Hlth., Albert Einstein Coll. of Med., Bronx, NY, [10]Publ. Hlth.Sci. Div., Fred Hutchinson Cancer Ctr., Seattle, WA, [11]Dept. of Preventive Med., Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [12]Dept.s of Med., Epidemiology and Hlth.Systems and Population Hlth., Univ. of Washington, Seattle, WA, [13]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, CHAPEL HILL, NC, [14]Dept. of Med., Albert Einstein Coll. of Med., Bronx, NY, [15]Inst. for Translational Genomics and Population Sci., Dept. of Pediatrics, Lundquist Inst. at Harbor-UCLA Med. Ctr., Torrance, CA, [16]Div. of Cardiology, Dept. of Med., Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [17]Div. of Cardiology, Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [18]Dept. of Med., Boston Univ. Chobanian and Avedisian Sch. of Med., Boston, MA, [19]Framingham Heart Study of the NHLBI and Boston Univ. Sch. of Med., Framingham, MA, [20]Sch. of Publ. Hlth., Univ. of Texas, San Antonio, TX

Abstract:

Background Echocardiography (echo) assesses cardiac structure and function and is used clinically to help diagnose heart failure. However, genetic characterization of echo measures is still limited. This study aims to identify common, rare, and structural variants (SVs) associated with cardiac structure and function in general populations.**Method** Single nucleotide variant (SNV), gene-centric coding, and SV analyses were performed on 18,611 ancestrally diverse participants from the Trans-Omics for Precision Medicine (TOPMed) Program (50.4% non-European ancestry) with available whole genome sequencing (WGS) and 37 cross-sectionally measured echo variables, adjusting for study, center, age, sex, height, genetic ancestry, and genetic relationship matrix. The significant SNV, gene-centric, and SV associations were identified as p-value <5.0e-8, 2.5e-6, and 1.5e-6, accounting for the number of tests performed. For eighteen echo measures with loci that reached significance in the cross-sectional analysis, we further examined the genetic impact on longitudinal changes of those measures in 34,738 observations on 20,237 participants from the same studies. **Results** The SNV analysis (MAF ≥1%) identified 19 significant echo loci, including *CXCL14* and *HSPB7*, known loci associated with heart failure. In the gene-centric coding analysis (MAF <1%), four significant genes were identified: *CIB3* with left ventricular (LV) end-diastolic diameter, *EMSY* with LV wall thickness, *EXTL2* with e' lateral mitral annulus, and *BCL9* with LV ejection fraction. A 997 bp deletion overlapping *TMSB15B* on chromosome X was associated with left atrial diameter. When investigating the longitudinal change of echo, four echo loci identified from the cross-sectional analysis remained significant, i.e., *HSPB7*, a gene highly expressed in the heart, was consistently associated with LV end-systolic diameter (LVESD) cross-sectionally and longitudinally. *HSPB7* is essential for heart development, and its association with LVESD was replicated using Japan Biobank data. The *CIB3* and *EMSY* genes remained significant in the gene-

centric coding analysis in the longitudinal echo analyses. Twelve additional echo-gene pairs were identified, including the *RNASEH1* with right ventricular end-systolic area, a gene involved with normal heart function in animal models. **Conclusion** Our WGS analyses identified novel common, rare, and structural variants associated with echo measures and their change in diverse populations, providing additional insight into the genetic architecture of cardiac structure and function.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5109T: Identifying of 10 Genes of Interest in 17 Families with dyslexia: The Influence of Simultaneously Present Variants on Individual Phonemic Decoding Ability

Authors:

Y. Kim[1], E. Kim[1], A. Lood[1], B. Peter[2]; [1]Arizona State Univ., Tempe, AZ, [2]Arizona State Univ, Tempe, AZ

Abstract:

Dyslexia is a common learning disorder affecting 5-10% of the population, characterized by difficulties in accurate and/or fluent word recognition, poor spelling, and decoding abilities. Previous studies have identified several potential dyslexia-associated genes and variants, but their phenotypic impacts require further investigation. This research aims to identify dyslexia-related genes through whole exome sequencing (WES) and determine whether families and individuals in our database share variants in the same genes with each other and with those listed in published sources. Additionally, it examines the effect of multiple variants in the same individuals on their dyslexia severity.WES was performed on DNA from 45 individuals with dyslexia representing 17 families. The exome data were filtered based on: (1) absence in gnomAD or presence with a minor allele frequency (MAF) <0.05; (2) CADD score >15; (3) annotation as nonsense, essential splice site, frameshift, missense, in frame, 5’ UTR, 3’ UTR, and regulatory region elements; (4) predicted “deleterious” or “damaging” by PolyPhen-2 and SIFT scores. Variants were visually validated using the Integrated Genome Viewer. After filtering, 1,071 genes shared by at least two families were identified and cross-referenced with 83 genes implicated in communication disorders (Guerra & Cacabelos, 2019) and 42 genes identified in a dyslexia GWAS study (Doust et al., 2021). Phenotypic data, including sentence recall, verbal and nonverbal indices, nonword repetition, finger tapping, rapid automatized naming, word and nonword reading, spelling, and phonological processing ability, were obtained through standardized assessments. Statistical analysis evaluated the correlations between number of

variants and phenotype.Filtering and cross-referencing revealed 7 overlapping genes from Guerra & Cacabelos (*PCDH11X, NEDD4L, MET, GNPTAB, GCFC2, FOXP2, CTTNBP2*) and 3 from Doust et al. (*PPP2R3A, PPP2R1B, TBC1D5*), resulting in 10 genes and 26 variants of interest from 18 individuals from 12 families, with number of variants ranging from 1 to 6. A Pearson correlation analysis showed a significant correlation (r = -.253, p = .047*) between the number of genetic variants and phonemic decoding ability. This study supports the hypothesis that the number of variants in individuals with dyslexia influences their reading ability, consistent with polygenic causes of dyslexia not only between, but especially within individuals. Further research will investigate functional roles of identified genes in dyslexia.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5110T: Identifying rare and less common SNPs associated with type 2 diabetes mellitus in multiplex families from the multiethnic GENNID study

Authors:

L. Simon, J. Wan, B. Wu, K. Edwards; Univ. of California, Irvine, Irvine, CA

Abstract:

Objective: Type 2 diabetes mellitus (T2DM) is a major global health problem. T2DM is a complex, polygenic disease with a strong genetic component, and GWAS have identified common SNPs that only explain a small proportion of T2DM heritability. Rare and less common SNPs contribute to the missing heritability but are difficult to study in the population. Given that rare SNPs in the population can be enriched in families, we aim to use the GENNID family-based multiethnic study to identify rare and less common T2DM-associated SNPs. **Methods:** The GENNID study sample consists of 1651 subjects in 340 T2DM multiplex families across four groups of genetic similarity: African American (AA), European American (EA), Japanese American (JA) and Mexican American (MA). We identified rare and less common SNPs by parsing through GENNID data and compiling lists of SNPs present in each GENNID group with population-specific gnomAD MAF<1% (rare) and 1%≤MAF<5% (less common). GCTA was used to test each SNP's association with T2DM within each GENNID group using a generalized linear mixed model, accounting for family relationship structure and adjusting for age, sex and BMI. In our association analyses, significance at p<0.05 was used as an initial threshold, followed by p<0.01 and p<0.001. **Results:** In AA, 54.6% of rare and less common SNPs were enriched in GENNID compared to the population (i.e., MAF in GENNID>0.05 while MAF in gnomAD<0.05); of 10883 SNPs significant at p<0.05, 2329 (21.4%) and 195 (1.8%) were significant at p<0.01 and

p<0.001, respectively. In CA, 54.9% of SNPs were enriched in GENNID compared to the population; of 6659 SNPs significant at p<0.05, 1197 (18%) and 96 (1.4%) were significant at p<0.01, and p<0.001. In JA, 58.7% of SNPs were enriched in GENNID compared to the population; of 4148 SNPs significant at p<0.05, 541 (13%) and 20 (0.5%) were significant at p<0.01, and p<0.001. In MA, 60.4% of SNPs were enriched in GENNID compared to the population; of 9022 SNPs significant at p<0.05, 1759 (19.5%) and 165 (1.8%) were significant at p<0.01 and p<0.001. Of the 10 most significant SNPs (and their genes) identified in each GENNID group, no overlapping was present. **Conclusions:** The GENNID family study was able to capture more than 50% of rare and less common SNPs on the MegaEx Illumina array that would have been omitted in common-variant GWAS analyses. Significant rare and less common SNPs were identified in each group with minimal overlap, and a large proportion were enriched in GENNID. Future research will utilize gene-based analyses to test the aggregate effect of multiple rare and less common SNPs.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5111T: Identifying splicing quantitative trait loci (sQTL) in colon tissue of IBD patients

Authors:

Y-C. Huang[1], N. Nishiyama[1], S. Silverstein[1], A. Bharadwaj[1], R. Ganesan[2], K. Clough[1], K. Darlington[1], G. Lian[1], M. Schaner[1], W. A. Faubion, Jr.[3], S. Sheikh[1], T. Furey[1]; [1]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [2]Carnegie Mellon Univ., Pittsburgh, PA, [3]Mayo Clinic, Phoenix, AZ

Abstract:

Inflammatory Bowel Disease (IBD) is a multi-factorial complex disease for which genetic susceptibility is a primary contributing factor. Although there are more than 300 loci have been associated with IBD in genome-wide association studies (GWAS), the underlying mechanism of how variants in the loci contribute to the disease is unclear. Alternative splicing events can produce isoforms that modify gene function. Genetic variation can modulate the frequency in which certain isoforms are produced by changing splicing signals around donor and/or acceptor sites increasing or decreasing splicing events at these locations. This study aims at finding genetic variants linked to changes in the relative presence of alternative isoforms which may uncover molecular mechanisms underlying disease. Using RNA-seq data generated in the colon tissue of 247 IBD patients, we detected 95,695 splice junctions within 10,062 genes using leafcutter. After correcting for known and unknown covariates, we detected a total of 15,717

sQTL at splice sites within 1,626 spliced genes (sGenes) (FDR <5%). Notably, our study identified *NOD1* as a sGene. Previous studies have associated mutations in the *NOD1* (nuclear organization domain 1) gene with Ulcerative Colitis. Both NOD1 and NOD2, the latter having the strongest genetic association in IBD, are pattern-recognition receptors that bind bacterial peptidoglycan initiating an inflammatory response. Recent research indicates that NOD1 isoforms are unable to transduce NF-kappaB activation, potentially discover the molecular mechanisms by which NOD1 detects bacterial peptidoglycan. To our knowledge, the lead variant for this sQTL (rs259440, chr7 C:T, $p < 1.19X10^{-19}$), which is in the first intron of *NOD1*, is a novel signal that has not been previously associated splicing of *NOD1* or directly with IBD. Together, these results underscore splicing events and candidate mechanisms that may influence gene function at GWAS loci.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5112T: Identifying ultra-rare genetic variants in Alzheimer's disease

Authors:

P-L. Cheng, W-P. Lee, T-T. Chung; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Although genome-wide association test (GWAS) systematically evaluates common genetic variants (minor allele frequency [MAF] > 1%) for complex diseases, such as Alzheimer’s disease (AD), additional disease risks could be explained by rare variants (MAF < 0.1%). However, performing traditional single-variant analysis on rare variants is often underpowered with a typical sample size. Aggregative tests that evaluate association of multiple variants in a biologically relevant region, such as burden test and SKAT-O, is an alternative method to study the association between AD and rare variants. While researchers try to find variants that frequently occur in disease or healthy samples, variants that only exist in disease or healthy samples are disease associated. In this study, we focus on identifying variants that severely impair protein function and only can be found in either disease or healthy samples with large sample sizes. We first focused on rare exonic variants exist in either case or control samples and are predicted by AlphaMissense as likely pathogenic in ADSP-r4 dataset (36,361 whole-genome-sequencing samples) and in another independent dataset (ADGC with 38,271 genotype samples imputed by TOPMed). We further restricted variants with the same direction in both datasets and found 1,940 risk variants located in 1,738 genes, and 4,490 protective variants located in 3,531 genes. In these candidate genes, we found 11 and 19 known AD associated genes for risk

and protective variants respectively. We also identified several novel rare variants in these known AD related genes. For example, a rare risk variant, rs1211631545 (AC = 4 for ADSP-r4 and AC = 2 for ADGC), in PSEN2 which is related to early onset Alzheimer's disease. The other example was a protective variant, rs202218688 (AC = 2 for ADSP-r4 and AC = 1 for ADGC), in APP that encodes amyloid-beta precursor. The discovery of risk and protective variants in known AD-related genes underscores the importance of studying ultra-rare variants in the genetic architecture of AD. Future work will focus on the functional validation of these candidate variants to better understand their effect in AD pathology.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5114T: Impulsivity facets and substance use involvement: Insights from genomic structural equation modeling

Authors:

L. Vilar-Ribo[1], A. Hatoum[2], A. Grotzinger[3], 23andMe Research Team, S. Elson[4], P. Fontanillas[5], A. Palmer[6], D. Gustavson[3], S. Sanchez-Roige[7]; [1]Univ. of California, San Diego (UCSD), San Diego, CA, [2]Dept. of Psychological and Brain Sci., Washington Univ. in St. Louis, St. Louis, Missouri, MO, [3]Inst. for Behavioral Genetics, Univ. of Colorado Boulder, Boulder, CO, [4]23andMe, Mountain View, CA, [5]23andMe, Sunnyvale, CA, [6]Univ. of California San Diego, La Jolla, CA, [7]Dept. of Psychiatry, Univ. of California San Diego, San Diego, CA

Abstract:

Impulsivity is a multidimensional trait associated with substance use disorders (SUDs), but the relationship between distinct impulsivity facets and stages of substance use involvement remains unclear. Using genomic structural equation modeling (SEM), we examined the latent genetic architecture between six impulsivity facets (based on the UPPS-P, BIS-11, and Monetary Choice Questionnaire) and two factors capturing normative substance use (based on lifetime and quantity/frequency measures; SU) and SUD using European-ancestry GWAS (N=79,729-632,802). SU and SUD factors were strongly genetically correlated but exhibited distinct patterns of association with impulsivity facets. Lack of premeditation showed positive genetic correlations with both SU and SUD factors. Sensation seeking and lack of perseverance were more strongly genetically correlated with SU than SUD. Conversely, delay discounting was more strongly genetically correlated with SUD than SU. After controlling for the shared genetic overlap between SU/SUD, we identified that sensation seeking, positive urgency, and lack of perseverance were independently genetically associated with SU, whereas negative urgency and delay discounting were independently genetically associated with SUD. Our findings suggest

that different impulsivity facets may confer risk for distinct stages of substance involvement through divergent mechanisms, with potential implications for SUD prevention and treatment.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5115T: Integrating massive-scale GWAS and multi-omic datasets reveals disease relevant cell types and effector genes for inflammatory bowel disease ⭐

Authors:

L. Fachal, International IBD Genetics Consortium; Wellcome Sanger Inst., Hinxton, Saffron Walden, United Kingdom

Abstract:

Genome-wide association studies (GWAS) have identified 370 independent signals, at 320 loci, associated with inflammatory bowel disease (IBD) and its two main subtypes, Crohn's disease (CD) and Ulcerative Colitis (UC). For the majority of these loci, the disease-effector genes are unknown, but ~15% contain genes in pathways targeted by advanced therapies for IBD. Better identification of effector genes, and the identification of additional loci, could thus significantly improve our biological understanding of the disease and identify new candidate drug targets.
To this end, we performed a multi-ancestry GWAS meta-analysis of 67 cohorts, from European, East and South Asian populations. The final dataset comprised 125,992 IBD patients (59,734 with CD; 57,565 with UC) and >1.2 million population controls. We identified 290 new independent disease association signals, 125 of which were located >1Mb from any previously reported loci. While most signals were shared between CD and UC, ~38% were subtype-specific (148 CD; 103 UC).
We used stratified LD score regression to decipher cell-disease susceptibility relationships, by integrating cell type specific, single cell (sc) and single nucleus, ATAC-seq annotations. Overall, the largest enrichment in IBD susceptibility variants was found at blood and tissue resident T cells, macrophages, B cells and NK cells. Despite most signals being shared between CD and UC, we observed disease subtype-specific enrichment of UC variants in enterocytes, goblet, paneth and tuft cells.
To nominate effector genes we followed two complementary approaches. We first ran colocalization analyses with cis expression QTL (eQTL) and sc-eQTL data for >30 relevant tissues and cell types. This approach prioritised 498 genes whose expression is likely to be perturbed by causal variants at 259 association signals. Integrating cis and trans protein QTL datasets via Mendelian Randomization prioritised 270 proteins with no evidence of reverse causality.

Two prioritised genes, *BCL2* (P_{UC} = 1.7E^{-8}) and *BCL2L11* (P_{UC} = 2.2E^{-13}), play a central role in anoikis, a form of enterocyte apoptosis key to maintaining the homeostasis of the intestinal epithelium. The two UC-risk association signals, located at cis candidate regulatory elements in enterocytes, are associated with higher levels of expression of these two genes, independently. These results give mechanistic insights into the observed effect of BCL2 inhibitors ameliorating colitis in IL10$^{-/-}$ mice.
In summary, this study provides new insights into the cellular types by which genetic variants influence disease, and nominates genes regulated by disease susceptibility variants.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5116T: Integrative analysis of tandem repeat expansions in pediatric heart diseases

Authors:

A. Mitina[1], M. Khan[1,2], L. Robert[1], Y. Yin[1], W. Engchuan[1,3], O. Hamdan[1,3], G. Pellecchia[1,3], B. Trost[1,4], I. Backstrom[1], Z. Wang[1,3], T. Nalpathamkalam[1,3], B. Thiruvahindrapuram[1,3], T. Papaz[5,6], J. Bouwmeester[1], C. E. Pearson[1,2], S. W. Scherer[1,3,7], J. Lougheed[8], T. Mondal[9], J. Smythe[10], L. Altamirano-Diaz[11], E. Oechslin[12], S. Mital[1,5,6], **R. K. C. Yuen**[1,2,3]; [1]Genetics and Genome Biology, The Hosp. for Sick Children, Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [4]Molecular Med., The Hosp. for Sick Children, Toronto, ON, Canada, [5]Ted Rogers Ctr. for Heart Res., Toronto, ON, Canada, [6]Div. of Cardiology, Dept. of Pediatrics, The Hosp. for Sick Children, Toronto, ON, Canada, [7]Dept. of Molecular Genetics and McLaughlin Ctr., Univ. of Toronto, Toronto, ON, Canada, [8]Div. of Cardiology, Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [9]Div. of Cardiology, Dept. of Pediatrics, McMaster Children's Hosp., Hamilton, ON, Canada, [10]Div. of Cardiology, Dept. of Pediatrics, Kingston Gen. Hosp., Kingston, ON, Canada, [11]Div. of Cardiology, Dept. of Pediatrics, London Hlth.Sci. Ctr., London, ON, Canada, [12]Div. of Cardiology, Toronto Adult Congenital Heart Disease Program at Peter Munk Cardiac Ctr., Dept. of Med., Univ. Hlth.Network, and Univ. of Toronto, Toronto, ON, Canada

Abstract:

Congenital heart disease (CHD) and cardiomyopathy (CMP) are the major cause of heart failure in infants, accounting for ~1% of live births per year. Both CHD and CMP vary by disease type and are clinically and genetically heterogeneous. The majority of the cases (50-70%) in these conditions remain gene elusive. Rare single nucleotide variants in genes encoding heart contractile proteins, such as *MYH7*, can be associated with both CMP and CHD, suggesting

shared pathophysiological mechanisms. At the same time, phenotypic manifestation can vary among the family members carrying the same disease-associated mutation, which implies the role for genetic modifiers in disease development.
Tandem repeat expansions (TREs), which are linked to epigenetic changes and can act as genetic modifiers, were shown to contribute to the risk of complex human diseases. Our group had previously demonstrated that hypermethylation of enhancer proximal TREs can lead to gene silencing in CMP. Here, we used the same approach to show that rare TREs in patients with CHD act through similar mechanisms. We analyzed whole genome sequencing (WGS) data from 563 individuals with tetralogy of Fallot (TOF) and 382 individuals with transposition of great arteries (TGA) representing Asian (15-17%), Black (2-7%), White (72-80%) and other ethnic groups including Aboriginal, Hispanic and Indocaribbean (3-8.5%). We found that rare TREs are disproportionately present in constrained genes near transcriptional start sites and have high GC content in all three examined cohorts. Rare TREs affect exonic regions in CMP (odds ratio=4.3; $p=1.7 \times 10^{-3}$) and in TGA (odds ratio=3.1; $p=1.4 \times 10^{-2}$), and splice sites in TOF (odds ratio=2.4; $p=1.8 \times 10^{-2}$). While rare TREs in CMP are enriched near active enhancer H3K27ac marks in the left ventricle, we found enrichment of rare TREs in TGA near H3K27me3 in the right ventricle.
We show that rare TREs are found in genes that are highly expressed in human cardiomyocytes and are differentially expressed in the left ventricle of the heart in individuals with cardiomyopathy, while rare TREs are found in genes that are expressed in endothelial cells in individuals with TOF, and in pericytes in individuals with TGA. Two genes, *DIP2B* and *RALGAPA2,* had rare repeat expansions in all three examined cohorts. We demonstrate the gene silencing effect of expanded CGG tandem repeats in *DIP2B* through promoter hypermethylation using Oxford Nanopore Technology long-read sequencing.
Our findings highlight the underrecognized contribution of rare tandem repeat expansions to the risk of pediatric heart diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5117T: Inter-eye clinical asymmetry and molecular discrepancies in the corneal epithelium in Keratoconus

Authors:

M. Gajecka[1,2], A. Wysocka[2], E. Matuszewska[1], M. Maleszka-Kurpiel[1,3], J. Wozniak[1,4], R. Ploski[5], J. Matysiak[1], M. Rydzanicz[5], K. Jaskiewicz[2]; [1]Poznan Univ. of Med. Sci., Poznan, Poland, [2]Inst. of Human Genetics Polish Academy of Sci., Poznan, Poland, [3]Optegra Eye Hlth.Care Clinic, Poznan, Poland, [4]Poznan Univ. of Life Sci., Poznan, Poland, [5]Med. Univ. of Warsaw, Warsaw, Poland

Abstract:

Introduction: Keratoconus (KTCN) is the most often occurring corneal ectasia (1:2000), a bilateral condition, characterized by progressive thinning of the cornea, leading to irregular astigmatism and gradual impairment of vision. One eye in a patient with KTCN is typically more severely affected than the other. To elucidate the molecular inter-eye differences, we used a multi-omic approach. **Materials and Methods:** The 126 experimental samples of the corneal epithelium (CE) of 21 KTCN patients and 26 experimental samples of CE of 6 controls (mild myopia patients) were assessed. The RNA and proteins were extracted using RNA/DNA/Protein PurificationPlus MicroKit (Norgen Biotek), NGS libraries were prepared using TruSeq Stranded TotalRNA LibraryPrep Gold (Illumina) and sequenced on Novaseq platform (60mln read pairs/sample). MALDI-TOF/TOF Tandem Mass Spectrometry was performed with an UltrafleXtreme mass spectrometer (Bruker Daltonics). The molecular data, concerning *central*, *middle*, and *peripheral* CE *topographic* regions (CETRs, our novel experimental model: PMID36811882) were interpreted within eye pairs, considering clinical data. **Results:** In KTCN patients inter-eye asymmetry was observed in clinical parameters such as epithelial thickness in *central* and *middle* CETRs (adjusted p-value=0.003 and 0.039, respectively), epithelial minimal thickness (adj. p-value=0.031), thinnest corneal thickness (adj. p-value=0.006), K1 (adj. p-value=0.001), K2 (adj. p-value<0.001), Kmax (adj. p-value<0.001), anterior elevation (adj. p-value<0.001), and posterior elevation (adj. p-value<0.001). 42 out of 85 discriminating proteins, were dysregulated in all three CETRs of KTCN patients (adj. p-value ≤0.05). GTF2IRD2, CWC25, PTPN1, DOCK8, and ZFP106 were specifically dysregulated in the *central* CETR of both patient's eyes (adj. p-value≤0.05). Whereas, only two proteins, TBC1D4 and HSPA1A, were upregulated solely in the *central* CETR of the eye with a more advanced disease state. **Conclusions:** The inter-eye differences found in ophthalmological parameters in patients with KTCN were poorly reflected in the molecular data, suggesting that molecular changes in the fellow eye precede clinical abnormalities. **Support:** The National Science Centre in Poland, grant no. 2021/41/B/NZ5/02245.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5119T: Investigation of rare recessive effects on quantitative and binary phenotypes in a cohort of 39,152 British Pakistanis and Bangladeshis enriched for autozygosity

Authors:

G. Kalantzis[1], F. H. Lassen[2], S. Hodgson[3], T. Heng[1], K. Walter[1], Genes & Health Research Team, D. A. van Heel[3], D. Palmer[4], H. C. Martin[1]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [3]Queen Mary Univ. of London, London, United Kingdom, [4]Univ. of Oxford, Big Data Inst., Oxford, United Kingdom

Abstract:

Previous common and rare variant-based genetic association studies have predominantly concentrated on detecting additive effects, while a notable gap remains in our understanding of recessiveness. We aim to detect rare recessive effects in the Genes & Health cohort of 39,152 British Bangladeshis and Pakistanis using exome-sequence and genotype chip data, using recent advances in statistical phasing to aggregate homozygous with compound heterozygous genotypes.

By utilizing sequenced trios, we empirically verified that our phasing framework is accurate, even after considering singletons (MAF=$1.3x10^{-5}$), achieving an average switch error rate of 0.55%, which drops to 0.34% when applying stricter filtering. Due to the prevalence of consanguinity in the cohort - 33% are offspring of second cousins or closer - our sample shows a higher ratio of homozygous to compound heterozygous genotypes at rare variants (MAF5%; predicted LOF or damaging variants) than UK Biobank (5.3:1 versus 1.6:1; Lassen et al., 2023). Next, we combined all available deleterious biallelic genotypes and tested for gene-based recessive association with a set of 471 binary and 66 quantitative phenotypes, using logistic and linear mixed models. We detected 14 significant associations passing our Bonferroni threshold ($p<5.7x10^{-9}$), the majority of which were with *HBB*, such as a well-documented recessive effect on Thalassemia ($p=6.2x10^{-15}$; OR=1.78), and a novel association with HDL cholesterol ($p=1.5x10^{-11}$; beta=-0.59). Moreover, *CATSPER2* was significantly associated with immature granulocytes ($p=5.4x10^{-10}$; beta=0.97) (a novel association), and *CLYBL* with vitamin B12 ($p=2.6x10^{-9}$; beta=-0.49), with a mechanism that has been described previously (Shen et al., 2017). We tested for dominance deviation, based on a smaller subset of unrelated individuals, and found that most of the rest associations appear to be partly recessive with some effect from heterozygotes. Lastly, we observed plausible associations close to our multiple-testing threshold, such as other hereditary anemias and *HBB* ($p=7.9x10^{-9}$; OR=1.49), atrial fibrillation and *CNTN5* ($p=7.0x10^{-6}$; OR=2.22), inflammatory bowel disease and *SLC15A3* ($p=7.0x10^{-6}$; OR=1.4) or *NCF4* ($p=7.5x10^{-5}$; OR=2.11; also indicated by Bolton et al., 2022), or Type I diabetes and *AMBN* ($8.4x10^{-6}$; OR=3.0), each of which had nominally significant p-values for dominance deviation.

We are in the process of developing a variance-component test for recessive effects to enhance power. This work is part of a meta-analysis with other diverse biobanks, all following the same pipelines, within the Biobanks Rare Variant Analysis (BRaVa) consortium.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5120T: Large scale analysis of patient-derived induced neurons to identify early developmental changes in schizophrenia ⭐

Authors:

B. Rowan[1,2], C. Seah[3,4], E. Drapeau[1,2], T. Rusielewicz[5], S. Bheda[1,2], M. Breen[1], S. Dobariya[5], J. Goldberg[5], B. McCarthy[5], L. Sloofman[1,2], X. Wang[1,2], Schizophrenia Stem Cell Resource, NYSCF Global Stem Cell Array® Team, K. Davis[2], J. Friedman[2], L. Huckins[6], D. Paull[5], K. J. Brennand[6], J. D. Buxbaum[1,2]; [1]Seaver Autism Ctr. for Res. and Treatment, Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Psychiatry, Icahn Sch. of Med., New York, NY, [3]Dept. of NeuroSci., Icahn Sch. of Med., New YorK, NY, [4]MSTP, Mount Sinai Medical Science Training Program, NY, [5]New York Stem Cell Fndn. Inst., New York, NY, [6]Yale Univ., New Haven, CT

Abstract:

Introduction: Schizophrenia is a debilitating psychiatric disorder with an incidence of 1% worldwide and with limited recent advances in therapeutics. Schizophrenia shows high heritability and both common and rare variation have been implicated. While GWAS studies have been very successful, the identified risk variants and loci remain poorly understood. We focus on transcriptional analyses from an ancestrally diverse iPSC dataset in order to better understand how early developmental processes shed light on schizophrenia risk.
Methods: We collected and reprogrammed ~100 donor fibroblasts and derived induced glutamatergic (iGlut) and GABAergic (iGABA) neurons from one or more clones per donor. Polygenic risk scores (PRS) were calculated for each donor using PRSice, utilizing the best fit PRS model based on the Psychiatric Genomic Consortium 3 (PGC3) Schizophrenia GWAS. For gene expression analysis, we used limma to perform weighted least square regression in iGABA cells and linear mixed regression in iGlut cells (because of multiple clones). Statistical significance was set at FDR<0.05. Expression cis-quantitative trait loci (eQTL) analysis was conducted using matrix QTL and colocalization analysis was performed using coloc.abf, with a posterior probability greater than 0.70 indicating statistical significance for hypothesis 4. All analyses were restricted to autosomes.
Results: We found 537 genes in iGlut and 38 genes in iGABA where expression was significantly correlated with PRS (ePRS). Twenty-nine ePRS were found in both analyses, with the same direction of effect. We found 102,151 eQTLs in iGlut, and 71,534 eQTLs in iGABA, representing 2,595 and 3,836 eGenes, respectively, with 1,216 seen in both cell types. The majority of eGenes identified were cell-type specific (75%). Strong replication rates were observed for iGlut ($\pi1$ =.856) and iGABA ($\pi1$ =.866) when comparing cell-type eQTLs to a large-scale, publicly available tissue-level cortical eQTL study. 289 iGlut ePRS overlapped with the iGlut eGenes, and

19 iGABA ePRS overlapped with iGABA eGenes. We found 7 iGLut eQTLs and 8 iGABA eQTLs to colocalize with PGC3 schizophrenia loci, with one locus identified in both iGlut and iGABA eQTLs.
Conclusion: We developed a large, ancestrally diverse, Schizophrenia Stem Cell Resource (SSCR) including fibroblasts, iPSCs, genotyping and transcriptomes from iGlut and iGABA. SSCR is proving useful for relating genetics to processes in early development of neuronal subtypes. Future work will use this dataset to generate transcriptome wide analysis (TWAS).

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5121T: Large scale exome sequencing in South-Asian populations identifies an enriched missense variant in *HNF4A* protective from type 2 diabetes

Authors:

M. Riaz[1], G. Pearson[2], J. Backman[3], P. Dornbos[1], J. Bovijn[4], S. Malhotra[5], Regeneron Genetics Center, the BELIEVE study group, S. Choudhury[6], F. Khan[7], M. Khalequzzama[8], A. Butterworth[9], L. Lotta[1], J. Marchini[1], M. Cantor[1], A. Baras[1], G. Abecasis[1], V. Gusarova[2], M. Sleeman[2], A. Shuldiner[10]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron Pharmaceuticals, Tarrytown, NY, [3]Regeneron, Tarrytown, NY, [4]Regeneron Genetics Ctr., White Plains, NY, [5]Regeneron, New York, NY, [6]Natl. Heart Fndn. Hosp. & Res. Inst., Dhaka, Bangladesh, [7]ICCDR, B, Dhaka, Bangladesh, [8]Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, [9]Univ. of Cambridge, Cambridge, United Kingdom, [10]Regeneron Genetics Ctr., New York, NY

Abstract:

South Asians (SA) have a fourfold higher risk of type 2 diabetes (T2D) compared to Europeans. Protein-coding genetic variants that affect disease risk/protection can provide clues to disease pathogenesis or therapeutic targets. We present the largest T2D exome wide association study of SA ancestry to date. The genome and exome-wide association analyses was conducted in two SA cohorts (the BangladEsh Longitudinal Investigation of Emerging Vascular Events [BELIEVE] study and UK Biobank South Asians) using additive model adjusted for age, sex and 20 genetic principal components in REGENIE. Inverse-variance weighted meta-analysis was conducted across these cohorts, and summary statistics of publicly available T2D association data from SA East London Genes and Health study was added, making a total effective sample size of 25,441 T2D cases and 100,419 controls in the pooled meta-analysis. HNF4A expression (RNA by qPCR and protein by immunohistochemistry) was measured in Huh7 cells transiently transfected with mutant (p.Pro437Ser or P>S) or wild type (WT) constructs. Fifty-three genome-wide significant

(P < 5e-08) index SNPs were associated with T2D in SA, including 5 novel (in *HNF4A (2 SNPs), UBE2E2, DUSP9* and *STK3* gene) and 48 known common variants that were previously identified in GWAS of T2D and related conditions. We observed a novel protective association of a missense variant rs150776703 (p.Pro437Ser) in Hepatocyte Nuclear Factor 4 Alpha (*HNF4A)* (P= 8.21e-19, OR = 0.53). Rs150776703 is markedly enriched in the Bangladesh (MAF 0.01) compared to Europeans (MAF 1.16E-06). Also, this variant showed nominal protective association of clinically diagnosed or self-reported chronic kidney disease (P= 4.30E-03 ;OR 0.63), remain significant after adjustment with T2D (P = 0.01; OR = 0.58). Transient overexpression of either P>S (rs150776703) mutant or WT HNF4A showed similar levels of HNF4A RNA transcript expression; however, P>S mutant HNF4A protein was increased compared to WT HNF4A (P<0.05). HNF4A is a transcription factor and plays vital role in the development and cellular function, including beta cell insulin secretion. While loss-of-function mutations in *HNF4A* are known to cause maturity-onset diabetes of the young, here we report, first missense variant in *HNF4A* that is markedly enriched in South Asians and confers protection from T2D and possibly CKD. Increased protein level of P>S may be due to increased protein stability, reduced degradation or post-translational modifications, consistent with effective gain-of-function. Our findings provide first human genetic evidence for activation of the HNF4A pathway as a novel therapeutic approach for T2D.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5122T: Latent classes analyses as a tool for revealing variable genetic association patterns in COVID-19 clinical subgroups

Authors:

N. Avramidis[1,2], S. Seth[3], E. Castineira, Pairo[1,2], K. Rawlik[1,2], N. Pirastu[4], K. Baillie[1,2,5,6]; [1]Pandemic Sci. Hub, Ctr. for Inflammation Res., Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Roslin Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Sch. of Informatics, Univ. of Edinburgh, Edinburgh, United Kingdom, [4]Human Technopole, Milan, Italy, [5]MRC Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh Western Gen. Hosp.,, Edinburgh, United Kingdom, [6]Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom

Abstract:

There is evidence that critically ill COVID-19 patients exhibit a range of clinical phenotypes, which can affect their response to treatment. Understanding the underlying biological reasons for this variation can contribute to the development of advanced therapeutic interventions. Using electronic healthcare records and genotyped data from individuals who participated in

both the Genetics of Mortality in Critical Care (GenOMICC) and the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C) studies, we aimed to identify distinct subgroups of critical illness in COVID-19. We analyzed 7,609 patients, clustering them into subgroups based on the presence of 25 symptoms reported upon hospitalization, using the latent class analysis method. We conducted a Genome-Wide Association Study (GWAS) in a one-vs-rest model. Each cluster was alternately treated as a case group, with remaining clusters as controls.The identified clusters are stable across models, ranging from 2 to 9 clusters, with the minimum identifiable Odds Ratio per cluster ranging between 1.27 and 2.05. As the model progresses to higher-level solutions, new clusters are formed by partitioning existing clusters, with minimal sample movement among the existing ones. Most clusters exhibit three core symptoms (fever, cough, and shortness of breath). Cluster distinction is based on the presence of less common symptoms, such as myalgia, loss of taste and smell, and gastrointestinal symptoms. Accounting for cluster stability, phenotypic differentiation, and statistical power, a one-vs-rest GWAS, using 5 clusters, was conducted. The GWAS results suggest genetic heterogeneity among the subgroups. Examining 49 previously identified genetic susceptibility loci, we found that cluster-specific effect sizes exhibit variability in the association patterns. The therapeutic implications of these findings are being explored through downstream transcriptome-wide association study models.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5123T: Leveraging EHR Phenotyping and Genome-Wide Association Studies to Identify Genetic Predisposition Factors for Systemic Lupus Erythematosus Within *All of Us* Research Program

Authors:

D. Shittu[1], A. Awan[1], P. Sauer[1], M. Nelson[2], M. Horton[2], T. Ferrara[2], J. Denny[3], H. Mo[3]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]NIH, Bethesda, MD, [3]NIH, Bethesda, MD

Abstract:

Systemic lupus erythematosus (SLE) is a complex autoimmune disease with substantial genetic predisposition. The heterogeneous clinical presentation of SLE complicates its diagnosis and characterization. Previous genome-wide association studies (GWAS) have identified over 120 genetic loci associated with SLE risk. Most GWAS have been conducted in Asian or European populations, leaving people who are historically understudied and often have worse SLE outcomes unrepresented. This study aims to identify genetic predisposition factors for SLE using electronic health record (EHR) phenotyping and GWAS within the *All of Us* Research Program, a

diverse US cohort which includes about 181,000 participants with both whole genome sequencing and EHR data.We used rule-based algorithms to select 1018 cases and 10180 controls. Cases were defined as those satisfying at least 4 of the 17 SLE criteria, including the SLE ICD code, and controls as those without any of the 17 criteria or the SLE ICD code, maintaining a control to case ratio of 10:1. Genome-wide significance thresholds were defined as 5×10^{-8}.Our trans-ancestry GWAS identified 59 SLE-associated significant SNPs after adjusting for the first 16 principal components. The result shows that the HLA region (6p23-6p21) is greatly enriched for significant SNPs, mainly found in the HLA-B gene, consistent with previous GWAS results. Known associations in previous studies were observed in our study, included *STAT4* rs7574865-G ($P = 1.18\times10^{-6}$, OR 0.72), *IRF5* rs4728142-A ($P = 1.80\times10^{-4}$, OR 1.25), *TNPO3* rs10488631-C ($P = 1.38\times10^{-6}$, OR 1.57), *ITGAM* rs9888739-T ($P = 3.39\times10^{-4}$, OR 1.28), and *PLCL1* rs11679484-A ($P = 2.97\times10^{-3}$, OR 1.22). Some loci identified as significant in previous studies were not significant in our GWAS, and many significant loci were either not previously associated with SLE or not found in the GWAS catalog.Given the higher severity and prevalence of SLE in the African ancestry, we conducted an African ancestry specific GWAS. We observed a significant association for *STAT4* rs7574865-G ($P = 1.19\times10^{-4}$, OR 0.76), *BANK1* rs10516487-A ($P = 4.66\times10^{-3}$, OR 0.83), *PHRF1* rs4963128-C ($P = 6.79\times10^{-3}$, OR 1.16), and *ITGAM* rs1143679-A ($P = 3.39\times10^{-4,}$ OR 1.30). The identification of both novel and previously recognized loci associated with SLE in a diverse cohort underscores the importance of studying varied populations to fully understand the genetic architecture of SLE. This study's result emphasizes the need for further validation in independent cohorts, especially in African ancestry populations, and functional validation of these variants' roles in SLE.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5124T: Longitudinal blood pressure is associated with risk of all-cause dementia

Authors:

G. Chittoor[1], A. E. Justice[1], G. Finney[1], B. Keenan[2], J. Robishaw[3], L. Kirchner[1], A. Pack[2]; [1]Geisinger, Danville, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Florida Atlantic Univ., Boca Raton, FL

Abstract:

Alzheimer's disease (AD) can be identified pre-clinically, but there are no specific therapies to notably delay progression. A potential approach is to target modifiable conditions that increase AD risk or, more broadly, mild cognitive impairment and dementia (DEM) to delay progression.

Adults with high blood pressure (BP) exhibit a greater risk for DEM and longitudinal decline in neuropsychological measures. Recent evidence suggests a shared genetic etiology between BP and AD. As such, we assessed the association of longitudinal BP with AD and DEM risk while accounting for genetic predisposition to AD and DEM. We leveraged the rich genotypic and longitudinal electronic health records (EHR) data in Geisinger's MyCode Study including up to 36k unrelated self-identified and EHR-derived White/European American individuals. We developed and validated, through chart review, a phenotypic algorithm to identify individuals with AD and general DEM from the EHR. Also, we derived three longitudinal measures of systolic (SBP) and diastolic (DBP): 1) median BP, 2) variance in BP, and 3) area under the curve (AUC). BP was right censored upon diagnosis for cases. We used publicly available GWAS statistics of AD to identify the best predictive polygenic risk score (PRS) using LDpred. We implemented cox proportional hazard regression (CoxPH) analyses for age-at-diagnosis of AD/DEM on BP measurements, adjusted for median age at BP measure, sex, median BMI, and with/without PRS. We also used logistic regression to examine associations of AD/DEM with the same covariates and including interaction of BP with PRS. We identified a total of 218 AD cases (56% female), 897 DEM cases (54% female), AD cases overlap with DEM. In our CoxPH analyses, we did not observe significant associations between BP measures and AD or DEM, but the PRS was significantly associated ($P<0.05/18$ tests=2.8e-3) in all models with increased risk (Hazard Ratios [95%CI] range from 2.10 [1.69-2.61] to 2.15 [1.73-2.67] and 1.46 [1.31-1.64] to 1.47 [1.32-1.65] for AD and DEM, respectively). Our case/control analysis showed that median and AUC of SBP and DBP were significantly ($P<2.8e-3$) associated with increased risk to DEM, but not AD, and remain significant with the addition of the PRS. We did not identify any significant PRSxBP interactions on AD or DEM. Polygenic risk of AD and DEM is significantly and robustly associated with higher risk at an earlier age-at-diagnosis. Longitudinal BP measures may play a role in overall risk for DEM, but not AD, nor age of onset. AD and all-cause DEM progression needs further study as more evidence is needed to reveal the potential therapeutic implications of interventions on higher BP.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5125T: Lower fasting insulin and higher vigorous physical activity independently associate with the same direction of cell-type-specific gene expression in human skeletal muscle fibers ⭐

Authors:

D. Ciotlos[1], S. Hanks[1], A. Varshney[1], M. Erdos[2], N. Manickam[1], H. Stringham[1], P. Orchard[1], E. Hill-Burns[1], N. Narisu[2], L. Bonnycastle[2], M. Sweeney[1], M. Laakso[3], J. Tuomilehto[4], T. Lakka[5], K.

Mohlke[6], M. Boehnke[1], H. Koistinen[4], F. Collins[2], S. Parker[1], L. Scott[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]NIH, Bethesda, MD, [3]Inst. of Clinical Med./Kuopio Univ. Hosp., Kuopio, Finland, [4]Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, [5]Inst. of Clinical Med., Kuopio, Finland, [6]Univ North Carolina, Chapel Hill, NC

Abstract:

Skeletal muscle plays a role in the regulation of glucose metabolism, and physical activity is a preventive factor for type 2 diabetes (T2D). To understand the relationships of T2D-related traits and physical activity with gene expression and chromatin regulation in skeletal muscle cell types, we analyzed skeletal muscle single nucleus RNA-sequencing (180,583 nuclei) and ATAC-sequencing (268,543 nuclei) data from 279 Finnish adults from the FUSION Tissue Biopsy Study. Participants were 60.2±7.4 years of age, 26% had T2D, and 41% were female. We used negative binomial models to test for associations between fasting plasma insulin (INS) or vigorous physical activity (PA) and cell-type composition, gene expression, or chromatin accessibility (23,841 genes; 924,519 peaks). Of 12 cell types including three muscle fiber types (1, 2a, 2x), individuals with higher INS had a higher proportion of Type 2x muscle fibers (log FC=0.13, p=1.9e-6), and a lower proportion of Type 1 muscle fibers (log FC=-0.07, p=1.7e-4), but no change in the proportion of total muscle fiber. Similar significant changes in the proportion of muscle fibers were associated with lower PA.

Across the three muscle fiber types, we found 376-566 genes associated with INS and 28-326 with PA. Within each fiber type, we found concordant directions of gene expression associated with higher PA and lower INS; these persisted when adjusting for the other trait. Using gene set analysis, we found concordant directions of gene set enrichments for lower INS and higher PA, for example, with higher expression of cellular respiration genes in Type 2a muscle fiber (PA: OR=1.77, p=1.4e-9, INS: OR=0.79, p=8.4e-5).

In the three muscle fiber types, we found 944-3,084 ATAC-seq peaks associated with INS and 231-2,640 with PA. To examine the relationships between gene expression and chromatin accessibility, we compared the direction of differential accessibility of peaks to the direction of expression of nearby genes. In Type 1 muscle fibers, of 143 genes with a significant peak for INS within 5 kb upstream of the gene's TSS, 92 genes (64%) had gene expression in the same direction as the chromatin accessibility; 16 of these genes had significant differential expression. For example, lower INS was associated with higher expression of PREX1 (log FC=-0.11, p=8.5e-5) and with higher chromatin accessibility at two PREX1 promoter peaks, one overlapping the TSS (log FC=-0.07, p=8.7e54).

The consistency of gene-level associations, with or without adjusting for both PA and INS, suggests that lower fasting insulin and higher vigorous physical activity independently influence gene expression in the same direction in similar pathways.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5126T: Mapping RNA isoform diversity and dysregulation in neuropsychiatric and control human brains

Authors:

D. Pinto[1], X. Lin[1], Y. Hadas[1], E. Hadjimichael[1], L. Li[2], E. Monte[3], E. Koornstra[1], A. Shie[4], R. Sebra[1], H. van Bakel[1], M. Snyder[3], J. Hallmayer[3], X. Wang[2], C. Liu[4], A. Urban[3], PsychENCODE Consortium; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Univ. of North Dakota, Grand Forks, ND, [3]Stanford Univ. Sch. of Med., Palo Alto, CA, [4]Upstate Med. Univ., New York, NY

Abstract:

Regulation of transcript structure generates transcript isoform and splicing diversity and plays an important role in health and disease. Differences in isoform expression account for substantial effect sizes and genetic enrichments in neuropsychiatric disorders such as ASD and schizophrenia; however, a systematic characterization of isoform expression in the human brain is lacking as most isoforms are yet to be directly profiled at the tissue or cellular level. Here, we performed deep single-molecule real-time (SMRT) isoform long-read sequencing of prefrontal cortex (PFC), cerebellum (CBL), hippocampus (HIP) and dorsal anterior cingulate cortex (dACC) from >50 individuals and 7 pools, totaling >60 libraries with >100 SMRT cells and >250 M long-reads. This high-resolution map of >350k isoforms greatly expands the transcriptional landscape of brain-expressed genes with >150k novel unannotated complete isoforms (54%), extends known coding and noncoding gene annotations, and uncovers hundreds of novel or reconstructed gene loci that can show cell-type specific expression. Orthogonal data, including, ribosome sequencing and mass-spectrometry data from >280 prefrontal cortical samples, 5'-end TSS profiling, and nanopore long-read sequencing, were used to further characterize the structure of novel isoforms and open reading frames. We further leveraged PsychENCODE single-nuclei transcriptomic and chromatin data to annotate cell-type-specific regulatory interactions for novel isoforms, reconstructed loci, and novel genes. Multi-platform analysis of single-nuclei full-length transcripts across PFC, HIP and striatum further revealed that isoform diversity seen in brain tissues is often due to underlying cell-type-specific patterns. Finally, we leveraged our reference map of full-length isoforms to characterize medically relevant genes, assess the functional consequence of rare genetic variants from a compendium of disease studies, and to improve quantitative trait analyses and refine the prioritization of candidate risk genes. Taken together, we provide a comprehensive resource to inform about diversity and dysregulation of isoform expression in neuropsychiatric disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5127T: Metabolic Syndrome-Associated Variants and Orofacial Cleft Risk in Puerto Rican Hispanics: Impact of Maternal BMI

Authors:

C. J. Buxo[1], M. Soto-Ortiz[1], N. Mukhopadhyay[2], T. Busch[3], M. J. Young[3], A. L. Petrin[4], L. M. Moreno Uribe[4], W. O. Awotoye[5], E. M. Albino[6], M. I. Salcedo[1], N. Debs[1], L. M. López-Del Valle[1], J. C. Murray[7], R. G. Munger[8], P. A. Mossey[9], M. L. Marazita[10], A. Butali[3]; [1]Dental and Craniofacial Genomics Core, Univ. of Puerto Rico Sch. of Dental Med., San Juan, PR, [2]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Univ. of Pittsburgh Sch. of Dental Med.,, Pittsburgh, PA, [3]Dept. of Oral Pathology Radiology and Med., Univ. of Iowa Coll. of Dentistry, Iowa City, IA, [4]Iowa Inst. for Oral Hlth.Res. and Dept. of Orthodontics, Univ. of Iowa Coll. of Dentistry, Iowa City, IA, [5]Iowa Inst. of Dental Res., Univ. of Iowa Coll. of Dentistry, Iowa, IA, [6]Univ. of Puerto Rico Sch. of Hlth.Professions, San Juan, PR, [7]Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA, [8]Dept. of Nutrition and Food Sci., Utah State Univ. Ctr. for Epidemiologic Studies, Logan, UT, [9]Dundee Dental Sch., Univ. of Dundee, Dundee, United Kingdom, [10]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., and Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Orofacial clefts (OFCs) are among the most common congenital anomalies with a multifactorial etiology of genetic and environmental factors. Previous studies have implicated components of metabolic syndrome (MetS) [e.g., type 2 diabetes mellitus (T2DM), gestational diabetes mellitus (GDM) and obesity/Body Mass Index (BMI)] as independent risk factors of OFC. This study investigated the contribution of metabolic syndrome (MetS)-associated variants to OFCs risk in Puerto Ricans. We conducted a case-control association study of children aged 0-18 including 226 non-syndromic OFC cases and 220 controls. Forty-eight strongly associated single nucleotide polymorphisms (SNPs) were selected from recent MetS GWAS consisting of 24 SNPs that were associated with T2DM risk, and another 24 SNPs known to affect BMI. Five phenotypes were analyzed: all cleft types combined, cleft lip with or without cleft palate (CL/P), cleft lip with a cleft palate (CLP), cleft lip only (CL), and cleft palate only (CP). Genotyping was carried out using Fluidigm. Association was run first without any covariates, followed by analyses to study the impact of mothers' BMI on the child's cleft risk. We examined both first-order and second-order (i.e. GXE) effects including the mothers' BMI measurements as covariates. Logistic regression was run using PLINK software.Significant associations ($p<0.05$) with and without inclusion of maternal BMI were observed for several phenotypes. In the SNP

only analysis, CL/P and CLP showed association to MetS-associated SNPs: rs13389219 (COBLL1, CL/P p-value 0.002, CLP p-value 0.0004) and rs7756992 (CDKAL1, CL/P p-value 0.002, CLP p-value 0.004). The main-effects only analysis using mother's BMI as the covariate showed association of the CL phenotype with SNP rs12571751 near ZMIZ1. ZMIZ1 has been previously reported in connection with facial dysmorphism (most significant p-value 0.026). In the GXE interaction using maternal BMI as a covariate, we observed significant associations at SNP rs1111875 with the four cleft phenotypes involving a cleft lip (gene HHEX, p-values ranging between 0.004-0.026). We are currently looking at confirmation of our significant MetS-associated SNPs with recent OFC GWAS in diverse populations (e.g., Africa and Central/South America). In conclusion, our study provides evidence that MetS-associated variants affect CL/P risk, and that maternal BMI may modulate this effect. Larger epidemiological studies are needed to confirm the connection between maternal phenotypes, such as T2DM and BMI status, and the risk of CL/P in the offspring.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5128T: Metabolomic Profile of Excessive Daytime Sleepiness in the Hispanic Community Health Study/Study of Latinos

Authors:

T. Faquih[1,2]; [1]Brigham and Women's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA

Abstract:

Introduction: Excessive daytime sleepiness (EDS) is a prevalent condition estimated to affect up to 33% of the United States population. EDS is a heterogeneous and complex symptom of various sleep disorders, insufficient sleep, and other chronic conditions and medications. EDS has also been associated with increased risk of cardiovascular and neuropsychiatric diseases, potentially through bidirectional causal pathways. In this study, we used metabolomic measurements in 6071 individuals from the Hispanic Community Health Study/Study of Latinos to elucidate the metabolomic factors associated with EDS.
Methods: EDS was assessed using the Epworth Sleepiness Score (a 24-point scale) at baseline (2008), continuously modeled. Metabolomics (877 metabolites) were quantified using the Metabolon platform—from blood assayed at baseline—and were log scaled. Our primary and sex stratified models adjusted for age, body mass index, ethnicity, recruitment center, alcohol, smoking, and physical activity. Secondary analysis further adjusted for insomnia, sleep duration, and sleep apnea. Significance threshold was set to 0.00013 to account for multiple testing. Finally, the interplay between metabolites, pathways, and genes was investigated through

network and pathway enrichment analyses.
Results: Levels of seven metabolite were each associated with a decrease of the sleepiness score by approximately 0.4 points per one log unit increase in metabolite levels. Those metabolites were primarily categorized as fatty acids (Eicosadienoic acid, Docosadienoate), steroids (Pregnenolone sulphate, Tetrahydrocortisol-glucuronide), along with a sphingomyelin and two uncharacterized metabolites. Sex stratified analysis identified three additional metabolites associated with EDS in men: Tyramine O-sulfate and two phosphatidylcholines. Tyramine O-sulfate was particularly associated with an increased sleepiness score by 0.5 per log unit; with attenuation after adjusting for sleep traits.Network and pathway enrichment analysis indicated that changes in pathways related to cortisol metabolism, hormonal steroid metabolism, and melatonin metabolism were linked to EDS.
Conclusion: These results suggest that a combination of metabolomic factors, representing diet and steroid hormonal pathways dietary related metabolites, and some genetic factors contributes to the metabolic profile of EDS. Some associations appeared independent of common sleep disorders while a male-specific association was driven by sleep disturbances (insomnia).
Support: NIH R01HL153814 (to H.W.), R01HL161012 (to T.S.), and R35HL135818 (to SR).

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5129T: Meta-Prediction of Coronary Artery Disease Risk

Authors:

S-F. Chen[1], A. Khattab[2], A. Torkamani[3]; [1]Scripps Res., La Jolla, CA, [2]Scripps Res., San Diego, CA, [3]Scripps Translational Sci. Inst., La Jolla, CA

Abstract:

Coronary artery disease (CAD) remains the leading cause of mortality and morbidity worldwide. Recent advances in large-scale genome-wide association studies have highlighted the potential of genetic risk, captured as polygenic risk scores (PRS), in clinical prevention. However, the current clinical utility of PRS models is limited to identifying high-risk populations based on the top percentiles of genetic susceptibility. While some studies have attempted integrative prediction using genetic and non-genetic factors, many have been cross-sectional and focused solely on risk stratification. Our primary objective in this study was to integrate unmodifiable (age / genetics) and modifiable (clinical / biometric) risk factors into a prospective prediction framework which also produces actionable and personalized risk estimates for the purpose of CAD prevention in a heterogenous adult population.Thus, we present an integrative, omnigenic,

meta-prediction framework that effectively captures CAD risk subgroups, primarily distinguished by degree and nature of genetic risk, with distinct risk reduction profiles predicted from standard clinical interventions. Initial model development considered ~2,000 predictive features, including demographic data, lifestyle factors, physical measurements, laboratory tests, medication usage, diagnoses, and genetics. To power our meta-prediction approach, we stratified the UK Biobank (UKBB) into two primary cohorts: 1) a prevalent CAD cohort used to train baseline and prospective predictive models for contributing risk factors and diagnoses, and 2) an incident CAD cohort used to train the final CAD incident risk prediction model. The resultant 10-year incident CAD risk model is composed of 15 derived meta-features from models trained on the prevalent risk cohort, most of which are predicted baseline diagnoses with multiple embedded PRSs. This model achieved an AUC of 0.84 and a macro-averaged F1-score of 0.61, outperforming standard clinical scores and prior integrative models. We further demonstrate that individualized risk reduction profiles can be derived from this model, with genetic risk mediating the degree of risk reduction achieved by standard clinical interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5130T: Molecular analysis of a trans interaction between *Ret* and *Ednrb* in Hirschsprung disease

Authors:

R. Fine, M. Fu, H. Chubaryov, L. Fries, G. Grullon, S. Chatterjee, A. Chakravarti; New York Univ. Langone Hlth., New York, NY

Abstract:

Introduction: Hirschsprung disease (HSCR) is a human multifactorial disorder of the enteric nervous system (ENS), which is a network of neuronal and glial cells established during embryonic development. However, our lab has previously shown that this system is under the control of a gene regulatory network composed of two central nodes, *RET* and *EDNRB.* These central genes act as transcriptional rheostats through which the cumulative effects of variants on genes are quantified by cells to ultimately determine disease penetrance. **Methods**: To understand the genetic and cellular events that lead to aganglionosis (loss of ENS cells in the GI tract) in HSCR, we crossed mouse lines of *Ret* and *Ednrb* mutant alleles to generate a series of models with increasing disease penetrance. Critically, our most severe model exhibits variable length aganglionosis, which accurately models the human disorder rather than total intestinal aganglionosis common in null allele models. We performed tissue level (bulk) and single cell transcriptome profiling on these lines at two developmentally distinct timepoints, E14.5 and P0,

to build a comprehensive dataset of over 70 biologically unique samples. We then constructed a general linear regression model accounting for *Ret* and *Ednrb* expression to find novel genes. **Results:** We find that the vast majority of perturbations can be accounted for by regressing out *Ret* and *Ednrb* expression, uncovering the remaining genes of interest, including transcription factors whose levels are extremely low in a subset of *Ret* positive inhibitory motor neurons in WT animals. This suggests any perturbation, even by a small amount, could lead to phenotypic penetrance and may explain neuronal sensitivity in HSCR disease. **Conclusion:** Based on these findings, we now understand with greater resolution a complicated multifactorial disorder and can uncover the finer grain effects of cumulative loss of *Ret* and *Ednrb*, including what happens within the ENS as it inches towards phenotypic penetrance. This greatly expands our understanding of the subtleties that lead to the disease and provides a therapeutic target for future treatment.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5131T: Multi-ancestry *APOE*-stratified genome-wide association study and identification of genetically at high-risk individuals for AD among ε4 non-carriers

Authors:

R. Dacey[1], C. Zhu[1], L-S. Wang[2], R. Mayeux[3], J. Haines[4], M. Pericak-Vance[5], G. Schellenberg[2], L. Farrer[1], G. Jun[1]; [1]Dept. of Med. (BioMed. Genetics), Boston Univ. Sch. of Med., Boston, MA, [2]Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Dept. of Neurology, and the Getrude H. Sergievsky Ctr., Dept. of Neurology, Columbia Univ., New York, NY, [4]Cleveland Inst. for Computational Biology, Dept. for Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [5]John P. Hussman Inst. for Human Genomics, and the Dr. John T. Macdonald Fndn., Dept. of Human Genetics, Miller Sch. of Med., Univ. of Miami, Miami, FL

Abstract:

Background: The *APOE* ε4 allele is the strongest genetic risk factor for Alzheimer's disease (AD), but only 20-25% of the population and 40% of AD patients carry one or more ε4 alleles. Studies have shown evidence of genetic risk factors specific to *APOE* subgroups suggesting the possibility of different risk profiles by *APOE* genotypes. Identification of genetically at high-risk individuals for AD separately by ε4 carrier status will improve precision in prediction of AD progression. **Methods:** We conducted genome wide association studies (GWAS) stratified

by *APOE* ε4 carrier status within and across multiple ancestry datasets and calculated polygenic risk scores (PRS) within ε4 stratum to profile individuals with high genetic risk profiles specific to *APOE* ε4 carrier status. We used 72 studies across four ancestry populations from the Alzheimer's Disease Genetic Consortium (ADGC) using 19,006 ε4 carriers and 26,034 ε4 non-carriers, adjusting for age, sex, five principal components and genetic relatedness. Study results were meta-analyzed within and across ancestry groups. We generated *APOE*-stratified PRSs in the Framingham Heart Study (N=3,628) and autopsied ADGC participants (N=4,698) using the non-Hispanic white (NHW) ancestry GWAS summary statistics from the ADGC to determine genetically at high-risk ε4 non-carriers. **Results:** We identified a novel suggestive association with AD with rs56838477 in the *PALM2-AKAP2* gene in ε4 non-carriers from the African ancestry (β=1.49, p=4.7x10^{-7}). *PALM2-AKAP2* was previously observed to be associated with AD in East Asians. In our cross-ancestry meta-analysis, we observed stronger association with known AD-related gene *SHARPIN* in ε4 non-carriers compared in ε4 carriers (rs34173062: $\beta_{\varepsilon 4-}$=0.31, p=2.7x10^{-10}; $\beta_{\varepsilon 4+}$=-0.03, p=0.78). We generated PRSs after removing genetic variants from the *APOE* region in FHS and autopsied ADGC subjects using ε4 stratified GWAS results of NHW. We observed an odds ratio (OR) of 3.39 for AD when comparing the top 2% of PRSs with the lowest 2% in FHS ε4 non-carriers, similar to previously reported OR of ε3/ε4 referencing to ε3/ε3 genotypes. Using these low- and high-risk PRS cutoffs, we observed significant associations with BRAAK stage and CERAD scores in autopsied ADGC ε4 non-carrier participants (β_{BRAAK}=1.20, p=4.4x10^{-4}; β_{CERAD}=0.80, p=1.0x10^{-3}), similar to previously reported effect sizes in ε4/ε4 comparing to ε3/ε3 genotypes in this sample. **Conclusion:** *APOE* ε4 non-carriers exhibited distinct genetic risk profiles than ε4 carriers beyond their ε4 status. PRSs using ε4 stratified GWAS results can determine genetically at high-risk individuals for AD in ε4 non-carriers.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5132T: Multi-ancestry stratified GWAS of alcohol use disorder and comorbid depression identifies novel loci

Authors:

E. Hartwell[1,2], Z. Jinwala[3,1], B. Voight[2], D. Oslin[2,1], J. Gelernter[4], H. Kranzler[2,1], R. Kember[2,1]; [1]CMCVAMC, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Univ. of Pennsylvania, Philadephia, PA, [4]Yale Univ Sch. of Med., West Haven, CT

Abstract:

Major depressive disorder (MDD) and alcohol use disorder (AUD) are prevalent, recurrent, and frequently comorbid diagnoses, with a lifetime history of AUD increasing the risk of developing

MDD four-fold. We conducted a stratified multi-ancestry GWAS (N>400,000) of AUD with and without comorbid MDD to explore genetic drivers of this comorbidity. AUD cases in the Million Veteran Program were participants with 1+ inpatient or 2+ outpatient International Classification of Diseases (ICD) 9/10 codes for alcohol abuse or dependence (n=114,627; by genetically inferred ancestry: AFR = 31,552, EUR = 69,203, AMR = 13,114, EAS = 636, SAS =122). We then determined MDD status using 1+ inpatient or 2+ outpatient ICD codes of depression. Controls were individuals with no instances of AUD or MDD diagnoses in their electronic health record (n=288,981). Genome-wide association testing was performed using PLINK, adjusting for age, sex, and 10 principal components. We then conducted cross-ancestry meta-analyses. Polygenic scores (PGS) were calculated using PRS-CS and phenome-wide association studies (PheWAS) conducted in the Yale-Penn dataset. We identified 63 loci in the cross-ancestry meta-analysis of comorbid AUD+MDD. In the AUD+MDD ancestry specific GWAS, 3 loci were significant in AFR, 14 in EUR (9 ancestry specific), 1 in AMR, and none in EAS or SAS. In the AUD-MDD GWAS, 1 locus was significant in AFR, 2 loci in EUR, 1 in AMR, and none in EAS or SAS. In EUR, 12 loci were identified in AUD+MDD that were not significant in AUD-MDD, and one in AFR. In AFR PheWAS, PGS for AUD-MDD was associated with numerous cocaine use phenotypes which were not observed in the AUD+MDD PheWAS in AFR. With a case sample size much greater than previous investigations of this comorbidity, we found unique loci associated with these comorbid traits. These results will inform our understanding of comorbid AUD and MDD and inform precision medicine efforts.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5133T: Multi-disciplinary Genomic Studies Inform the Etiology of Congenital Heart Disease

Authors:

U. Gurusamy[1], A. Willoughby[1], B. Moreland[1], M. Marhabaie[1], S. Ramadesikan[1], I. Showpnil[1], Y-C. C. Hou[1], M. Artomov[2], B. P. Chaudhari[1,3], E. A. Varga[1], E. Mardis[1], R. Wilson[1], K. Manickam[4], V. Garg[5], P. White[1], A. M. Bigelow[6], D. Koboldt[1]; [1]1The Steve and Cindy Rasmussen Inst. for Genomic Med., The Abigail Wexner Res. Inst. at Nationwide Children's Hosp., Columbus, OH, [2]Nationwide Children's Hosp., Columbus, OH, [3]Dept. of Pediatrics, Wexner Med. Ctr., The Ohio State Univ., Columbus, OH, [4]3Div. of Genetics and Genomics, Nationwide Children s Hosp., Columbus, OH, [5]4Ctr. for Cardiovascular Res., Nationwide Children's Hosp., Columbus, OH, [6]5Dept. of Pediatrics, The Heart Ctr., Nationwide Children's Hosp., The Ohio State Univ., Columbus, OH

Abstract:

Congenital heart disease (CHD) is the most frequently occurring birth defect, affecting nearly 1% of all live births. The etiology of CHD is recognized as multifactorial, involving a combination of environmental and genetic factors. Recent advancements in sequencing technologies have facilitated the discovery of genetic variants associated with both syndromic and non-syndromic CHD by elucidating molecular pathways involved in cardiac development. However, even linking known genetic causes to specific CHD phenotypes, as well as to cardiomyopathy and arrhythmias remain challenging due to genetic heterogeneity, variable expressivity, reduced penetrance, and polygenic origins. For the past year, we have conducted an IRB-approved translational research study, Cardiac OUtcomes Research by Application of GEnomics (COURAGE) for Kids, which employs state-of-the-art genome sequencing coupled with detailed clinical phenotyping to identify novel etiologies of CHD and translate this genetic knowledge to the bedside for informed clinical decision-making. To date, we have sequenced 130 individuals from 51 families affected by structural cardiovascular defects, cardiomyopathies, aortic/venous anomalies, arrhythmia, and similar conditions. Our patient cohort encompasses both isolated and syndromic CHD, both of which have complex genetic underpinnings. Some 20% of families studied so far had diagnostic variants in known genes including JAG1 (Alagille syndrome), TTN (dilated cardiomyopathy), SON (ZTTK syndrome), MYH7 (dilated cardiomyopathy), RYR2 (ventricular tachycardia), KCNH2 (long QT syndrome), KMT2D (Kabuki syndrome), SMAD2 (congenital heart defects), and ANKRD11 (KBG syndrome). Another 17.6% of families have phenotypically relevant candidate variants with de novo (EYA4, CDH10), inherited (ACTC1, CDH10, PRKD1), and biallelic (MYH6) etiologies. Consistent with other studies, 4% of individuals sequenced harbor secondary findings in medically actionable genes. Our results so far underscore the power of genomics to uncover medically relevant variants and compelling new candidate genes associated with CHD; these in turn should improve disease prediction, personalize treatment plans, and implement preventive measures for affected families.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5134T: Multiethnic genome-wide association meta-analyses of primary open-angle glaucoma identifies novel cross-ancestry and ancestry-specific loci.

Authors:

A. Mulford[1], V. Lo Faro[2,3,4], S. Bidi[5], L. Bhatta[6], J. Shortt[7], V. Tozzo[8], Y. Wang[9,10], Q. Le Ngoc[11], K. Nam[12], H. Fry[13], E. M. Bertucci-Richter[14], L. Guare[15], B. Jacobs[5], Global Biobank Meta-analysis Initiative, B. Brumpton[6], C. Gignoux[7], B. Pasanuic[8], A. R. Martin[9,10], Y. Okada[16,17,18], S. MacGregor[11], V. Rovite[19], S. Lee[12], R. Walters[13], S. Zoellner[14], S. Setia Verma[15], D. A. van Heel[5], H.

Choquet[20], H. Snieder[4], A. R. Sanders[1,21], N. Cox[22,23], W. Zhou[24,10], J. Hirbo[23,22]; [1]Endeavor Hlth.Res. Inst., Evanston, IL, [2]Uppsala Univ., Uppsala, Sweden, [3]Amsterdam Univ. Med. Ctr., Amsterdam, Netherlands, [4]Univ. of Groningen, Groningen, Netherlands, [5]Barts and Queen Mary Univ. of London, London, United Kingdom, [6]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [7]Colorado Ctr. for Personalized Med., Aurora, CO, [8]Univ. of California, Los Angeles, Los Angeles, CA, [9]Massachusetts Gen. Hosp., Boston, MA, [10]Broad Inst. of Harvard and Massachusetts Inst. of Technology, Cambridge, MA, [11]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [12]Seoul Natl. Univ., Seoul, Korea, Republic of, [13]Univ. of Oxford, Oxford, United Kingdom, [14]Univ. of Michigan, Ann Arbor, MI, [15]Univ. of Pennsylvania, Philadelphia, PA, [16]Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [17]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [18]Univ. of Tokyo, Tokyo, Japan, [19]Latvian BioMed. Res. and Study centre, Riga, Latvia, [20]Kaiser Permanente, Div. of Res., Pleasanton, CA, [21]Univ. of Chicago, Chicago, IL, [22]Vanderbilt Univ. Med. Ctr., Nashville, TN, [23]Vanderbilt Univ. Sch. of Med., Nashville, TN, [24]Massachusetts Gen. Hosp., Chestnut Hill, MA

Abstract:

Primary open-angle glaucoma (POAG) is a chronic, progressive eye condition, characterized by loss of the retinal nerve fiber layer and optic nerve rim, causing vision impairment and ultimately blindness. Clinical manifestations and outcomes of POAG vary across ancestries, illustrating the need for studies that elucidate ancestry-specific disease mechanisms. The Global Biobank Meta-analysis Initiative (GBMI) is a collaboration aimed at providing a vast network of data to enable more powerful and diverse analyses of diseases. In a follow-up to our previous work, an expanded multiethnic and ancestry-specific sex-stratified meta-analyses of genome-wide association studies (GWAS) of POAG were conducted across 24 GBMI biobanks in a total of 1,632,971 individuals (ncases=27,500), across five ancestries: European (EUR, ncases=20,892), African (AFR, ncases=1,555), East Asian (EAS, ncases=4,331), South Asian (SAS, ncases=295) and admixed American (AMR, ncases=920). Sex-stratified meta-analyses were also performed in 536,152 males (ncases=7,640) and 626,547 females (ncases=8,852). For consistency and replicability, case definition, quality control (QC), and GWAS were conducted using a streamlined pipeline. Each biobank performed a generalized mixed model method adjusting for age, sex, 10 principal components, and any biobank specific covariates. After QC and harmonization, we ran meta-analyses based on inverse-variance weighting scheme, and identified 84 genome-wide significant loci, including 43 novel loci, 30 of which were ancestry-specific associations. Notably we identified rs144844520_*MED27* (P=2.84E-09) in EUR, rs72775769_*RBFOX1* (P=7.86E-09) in AFR, rs650208_*OLFM1* (P=4.85E-08) in EAS, rs115503338_*VWA1* (P=3.13E-09) in SAS, and rs74929283_*CACNA1C* (P=3.03E-08) in AMR; these genes have all been implicated in eye development, function, and disease. Downstream gene-based association analyses identified five novel genome-wide significant genes after

Bonferroni correction (*ISLR*, *NTHL1*, *OBP2B*, *TBC1D2*, *WNT9B*). Pathway analysis revealed four significantly enriched gene sets, including transmembrane receptor protein tyrosine kinase activity. By utilizing extensive and globally diverse biobank data, we identified many novel associations with POAG. Interpretation of these will deepen understanding of the complex mechanisms underlying POAG disease risk and progression, and help to reduce disparities in clinical manifestations as they vary across ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5135T: Multi-omic characterization of persons living with T2D in a vulnerable Hispanic/Latinx population

Authors:

E. Frankel[1], L. Petty[2], R. Roshani[1], W. Zhu[1], H-H. Chen[3], M. Yaser[4], M. Graff[5], M. Krishnan[6], V. Buchanan[7], M. Lee[8], A. Gutierrez[9], H. Highland[10], K. Young[5], J. McCormick[8], S. Fisher-Hoch[11], K. North[5], J. Below[12]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Academia Sinica, Taipei, Taiwan, [4]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [5]Univ North Carolina, Chapel Hill, NC, [6]Univ. of North Carolina, Chapel Hill, NC, [7]Univ. of North Carolina - Chapel Hill, Westfield, IN, [8]UTH SPH Brownsville, Brownsville, TX, [9]Univ. of Texas, Houston, TX, [10]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [11]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [12]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

Type 2 diabetes (T2D) is a complex metabolic condition characterized by decreased response to insulin and subsequent elevated blood glucose concentration. T2D is a major driver of global health burdens, with a disproportionate impact on historically marginalized populations, specifically, US Hispanic or Latinx (HL) populations. HL populations have an increased prevalence of T2D compared to non-Hispanic whites; however, they are often underrepresented in genetic studies and have strikingly limited representation in multi-omic studies. To address these gaps, our research leveraged a deeply phenotyped HL cohort, the Cameron County Hispanic Cohort (CCHC), in conjunction with robust cross-sectional transcriptomic, proteomic, and metabolomic data to better understand the biological mechanisms contributing to T2D risk.
We performed single mRNA, protein, and metabolite linear regressions to compare the expression and abundance of genes, proteins, and metabolite levels between individuals with T2D and controls without diabetes. We defined T2D in accordance with the American Diabetes Association criteria as of 2010 and adjusted for sex, age, the first three principal components (PC) of ancestry, and ten surrogate variables (for RNAseq and proteomics). We observed 749

genes, 348 proteins, and 158 metabolites differentially expressed or abundant in T2D cases and controls ($p < 0.05$ after FDR correction in each omic). The most significant genes identified from the transcriptomic analysis were *ABCG1*, *CPT1A*, and *KLF10*. There were thirteen genes implicated in both the transcriptomic and proteomic analyses, including *CD93*, *AOC3*, and the novel gene *TMED8*. Top finding in proteomics included LPL, PTPRS, and GCG, the precursor of GLP-1. The metabolomics analysis indicated glucose as the top finding. Additionally, the significant pathways identified included branch chain amino acid homeostasis and galactose metabolism. While some findings were novel discoveries, many of the finding reflect metabolic changes known to play central roles in T2D pathology and other downstream sequalae.
In summary, our analyses revealed multiple metabolic changes associated with TD across multiple omics characterized in the whole blood. To our knowledge, this is one of the first cross-sectional multi-omic characterization of T2D in a highly vulnerable HL population. These data demonstrate not only how multi-omic studies may elevate understanding of T2D risk in HL populations, but also implicate functional genes, proteins, and metabolites as well as a mechanism of effect into T2D risk in HL individuals.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5136T: Multi-omics profiling of angiographically confirmed coronary artery disease highlights genetic regulation of the human plasma proteome ⭐

Authors:

T. Bigdeli[1,2], C. Chatzinakos[3], P. Shwetha[1], R. Peterson[4], C. Boutin-Foster[1], M. Salifu[1], J. Marmur[1], E. Cavusoglu[1]; [1]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [2]VA New York Harbor Hlth.care System, Brooklyn, NY, [3]SUNY DOWNSATE, Brooklyn, NY, [4]SUNY Downstate Hlth.Sci. Univ.State Univ. of New York Downstate Hlth.Sci. Univ, Brooklyn, NY

Abstract:

Background: Genome-wide association studies (GWAS) have identified more than 300 risk loci for coronary artery disease (CAD). Proteome-wide association studies (PWAS) prioritize specific proteins that are regulated by disease-associated genetic polymorphisms, or protein quantitative trait loci (pQTLs). Specific pQTLs have been shown to co-localize with GWAS findings, suggesting their causal influences on CAD via the activity of specific proteins.
Methods: For 352 patients referred to our clinic for “gold standard” coronary angiography (80% African American, 55% female), we measured plasma levels of 768 proteins using the Cardiometabolic I and Inflammation I panels from Olink, Inc. Participants were genotyped on

the Illumina Global Screening Array (GSA); we performed haplotype phasing with SHAPEIT5 and imputed missing genotypes via the TOPMed (r3) reference data. We performed limma analyses to identify differentially expressed proteins in CAD, adjusting for age, gender, body mass index (BMI), genetic ancestry PCs 1-3, LDL levels, and total cholesterol. We used elastic-net regularized regression to estimate weights for single nucleotide polymorphisms (SNPs) within 1Mb of the gene encoding each protein, using the same covariates as selected for differential expression analyses along with the first 10 PEER factors. We used MetaXcan to directly impute PWAS results from a recent large-scale GWAS of CAD in the Million Veteran Program (MVP).
Results: We identified 13 CAD-associated proteins which remained significant after false discovery rate (FDR) correction; among 115 nominally significant (P<0.05) proteins, we observed a significant enrichment (P<5e-5) of targets of BACH1, RAC3, GATA3, and ETV. Applying our newly developed models to GWAS summary statistics, 11 and 26 proteins were predicted to be differentially expressed in African Americans (17,202 cases, 59,507 controls) and European Americans (95,151 cases, 197,287 controls), respectively. Ten proteins remained significant after FDR adjustment, including IL6R, LPL, and PLTP, which was more than expected by chance (P=0.0016). Follow-up gene set enrichment analysis (GSEA) using single-cell transcriptomics data revealed that CAD-associated proteins were upregulated in B cells (ρ=0.457) and downregulated in lymphatic cells (ρ=-0.329).
Conclusions: We have successfully demonstrated how high-throughput proteomics data for a modestly sized training cohort can yield powerful proteomic predictions in larger-scale GWAS results. We are now extending this integrative and cost-efficient approach to an expanded GWAS catalog of more than 2,000 diseases and lab values.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5137T: Multi-Trait Genome-Wide Association Analysis of Psychiatric Traits Identified New Loci

Authors:

S. Farhang-Sardroodi[1], **R. Tyndale**[1,2,3], M. Chenoweth[1,2,3]; [1]Dept. of Pharmacology and Toxicology, Univ. of Toronto, Toronto, ON, Canada, [2]Dept. of Psychiatry, Univ. of Toronto, Toronto, ON, Canada, [3]Campbell Family Mental Hlth.Res. Inst., CAMH, Toronto, ON, Canada

Abstract:

Tobacco use disorder and Schizophrenia are highly heritable, comorbid conditions that demonstrate substantial resistance to current treatments. Genome-wide association studies (GWASs) have proven useful for identifying biological risk pathways and potential new targets

for drug development. In large GWASs conducted separately for tobacco use and Schizophrenia, numerous risk genes were identified, some of which are overlapping. We hypothesized that a multi-GWAS approach that leverages the genetic correlation between tobacco use and Schizophrenia would identify new risk loci not captured in the single-trait GWASs. We used summary statistics from the largest GWAS conducted to date for Schizophrenia (SCZ, N=130,644; 53,386 cases; 77,258 controls; European ancestry, Wave 3 of PGC, PMID: 35396580) and for cigarette smoking behaviour, defined by the number of cigarettes smoked per day (CPD; N=337,334; European ancestry, GSCAN, PMID: 30643251). Through LD Score Regression (LDSC) and Multi-Trait Analysis of GWAS (MTAG), we revealed specific heritability estimates of h2= 0.08 (±0.01) for CPD and h2= 0.37 (±0.012) for Schizophrenia. Additionally, we identified a genetic correlation of 0.12 (P=5x10-8) between these behaviours. MTAG revealed no new loci for SCZ beyond those identified by the original SCZ GWAS but uncovered 7 new loci for CPD not found in the original CPD GWAS. The significant SNPs (at P<5x10-8) from these 7 loci were mainly clustered in intronic and intergenic regions. Using FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies), 28 genes were identified following positional, expression quantitative trait locus (eQTL), and chromatin interaction mapping. These 28 genes included NOP14 (involved in rRNA processing), TCF4 (a transcription factor involved in nervous system development), CKB (a creatine kinase involved in energy homeostasis, including in the brain), and PPP1R13B (involved in apoptosis), among others. We cross-referenced the 28 mapped genes in the GWAS Catalog. We found that while they were not identified in the original CPD GWAS, many were found in MTAG analyses combining CPD with other correlated, psychiatric traits (e.g. alcohol consumption). Our findings agree with this prior work and suggest that MTAG of CPD and genetically correlated psychiatric traits can enhance the power of genomic discovery for CPD. In addition to providing new insights into the biology of smoking behaviours, this work could eventually lead to improved treatment targets for tobacco use.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5139T: Next generation sequencing study to unravel the genetic basis of congenital bile-duct dilatation

Authors:

C. Tang[1], Q. Lin[1], P. Sham[2], P. Tam[1,3]; [1]Dept. of Surgery, The Univ. of Hong Kong, Hong Kong, China, [2]Dept. of Psychiatry, The Univ. of Hong Kong, Hong Kong, China, [3]Faculty of Med., Macau Univ. of Sci. and Technology, Macau, China

Abstract:

Congenital bile-duct dilatation (CDD), also known as choledochal cyst, is a complex congenital anomaly characterized by abnormal dilatation of the extrahepatic and/or the intrahepatic bile ducts. Type I with dilatation limited to the common bile duct represents the most common disease subtype, which accounts for 50%-80% of the CDD cases. Our previous genetic study suggested a genetic predisposition in the pathogenesis of the disease. To further elucidate the genetic causes of CDD, we performed trio-based whole exome and genome sequencing on type I CDD patients of Chinese ancestry. De novo analysis identified thirty-six damaging protein-altering variants, including five protein truncating variants and twenty-eight nonsynonymous variants. Truncating de novo mutations in evolutionarily constrained genes and missense changes with regional constraint were found to be enriched in cancer-related genes such as *PIK3CA*, *TLN1, CYLD*, *NK1*, and *MAP2K1*. Recessive and compound heterozygous damaging mutations were also identified in related pathways. Additionally, we detected significant enrichment of damaging mutations in cytoskeleton and microtubule-associated pathways. Spatiotemporal expression of these candidate gene was further evaluated using fetal single cell transcriptomic data. While CDD was reported to be associated with increased risk of developing biliary tract cancer, our findings of cancer gene predisposition may help stratification of patients with higher cancer risk to improve clinical outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5140T: Novel contributors to orofacial cleft risk identified at 7q33 and 10q24.32 through multi-ancestry meta-analysis

Authors:

J. Carlson[1], **X. Zhang**[1], Z. Erdogan-Yildirim[1,2], T. Beaty[3], A. Butali[4], C. Buxo[5], L. Gowans[6], J. Hecht[7], L. Moreno[4], J. Murray[4], I. Orioli[8,9], C. Padilla[10,11], A. Vieira[1], G. Wehby[4], E. Feingold[1,12], E. Leslie[13], S. Weinberg[1,2], M. Marazita[1,2], J. Shaffer[1,2]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, [3]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [4]Univ. of Iowa, Iowa City, IA, [5]Univ. of Puerto Rico Med. Sci. Campus, San Juan, PR, [6]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [7]Univ Texas McGovern Med Sch, Houston, TX, [8]Federal Univ. of Rio de Janeiro, Rio De Janeiro, Brazil, [9]INAGEMP (Natl. Inst. of Population Med. Genetics), Porto Alegre, Brazil, [10]Univ. of the Philippines Manila, Manila, Philippines, [11]Univ. of the Philippines System, Quezon, Philippines, [12]Oregon State Univ., Corvallis, OR, [13]Emory Univ., Atlanta, GA

Abstract:

Nonsyndromic cleft lip with or without cleft palate (CL/P) is a common birth defect caused by genetic and environmental risk factors. Despite the identification of many genetic loci harboring CL/P-risk variants, there are likely many more unknown genetic determinants of CL/P. Furthermore, while the process of embryonic facial development is well characterized, knowledge of the molecular mechanisms that underly this process and how their disruptions result in CL/P is lacking. Thus, we sought to identify novel CL/P-risk loci through a multi-ancestry study of CL/P. We performed genome-wide meta-analyses of CL/P from three cohorts totaling >14,000 individuals using METAL. In addition to replicating several known CL/P-risk loci, we also observed novel association in two regions: 1) the *CALD1* locus at 7q33 (lead variant rs17168118, p = 1.1e-10), and 2) a ~400kb region at 10q24.32 spanning *BTRC*, *DPCD*, *POLL*, and *FBXW4* (lead variant rs10883664, p = 8.9e-8). Next, we performed multi-tissue TWAS with s-MulTiXcan which identified several significant genes including *IRF6* and *VAX1* (two well-established CL/P-risk loci), *CALD1* (p = 2.6e-10), *BTRC* (p = 2.1e-9), *DPCD* (p = 1.6e-5), *POLL* (p = 4.6e-5), and *FBXW4* (p = 1.3e-6). These new loci have potential functional effects that inform the molecular mechanisms of CL/P. First, *CALD1* at the 7q33 locus plays an essential role in regulating smooth muscle and nonmuscle contraction. Preclinical models suggest Cald1 is involved in embryonic morphology through blood vessel endothelial cell migration and tube morphogenesis. *CALD1* is also highly expressed in human and murine craniofacial tissue across CS13-20/E8.5-10.5, during which the formation of the lip and palate occurs along with other organogenesis. The association signal in this region directly overlaps a known craniofacial super-enhancer affecting *CALD1* expression. The second locus at 10q24.32 harbors many genes that are all expressed in human craniofacial tissue during embryonic formation of the lip. This region comprises two topologically associated domains with an interaction region separating them, each overlapping known craniofacial super-enhancers. Further fine-mapping and functional studies are needed to determine the gene(s) affecting CL/P-risk at this locus. A promising candidate is *BTRC* which showed the strongest evidence of association in our TWAS and is involved in several relevant pathways, including Wnt and Hedgehog signaling, both previously implicated in CL/P. In summary, this study provides evidence of two novel loci associated with risk of CL/P and offers new insights into the molecule mechanisms underlying embryonic facial development.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5141T: Novel loci mapped for blood pressure trait trajectories in diverse adolescence and early adulthood.

Authors:

S. Okello[1], M. Graff[1], H-H. Chen[2], X. Zhang[1], J. B. McCormick[3], S. P. Fischer-Hoch[4], C. Kooperberg[5], U. Peters[5], M. Fornage[6], S. Buyske[7], A. E. Justice[8], G. Chittoor[8], E. Blanco[9,10], R. Burrows[10], C. Albala[10], J. L. Santos[11], B. Angel[10], S. Gahagan[9], S. S. Rich[12], L. A. Lange[13], E. Boerwinkle[14], R. Kaplan[15], M. Stanislowski[16], A. Mottl[17], H. Highland[1], J. Below[2], K. E. North[1]; [1]Dept. of Epidemiology, Gillings Sch. of Global Publ. Hlth., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Vanderbilt Genetics Inst. and Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Univ. of Texas Hlth.Sci. Ctr. at Houston, Sch. of Publ. Hlth., Brownsville, TX, [4]Univ. of Texas Hlth.Sci. Ctr. at Houston, Sch. of Publ. Hlth., Brownsville, TX, USA, Brownsville, TX, [5]Fred Hutchinson Cancer Ctr., Seattle, WA, [6]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [7]Dept. of Statistics, Sch. of Arts & Sci., Rutgers Univ., Piscataway, NJ, [8]Population Hlth.Sci., Geisinger, Danville, PA, [9]Div. of Academic Gen. Pediatrics, Child Dev. and Community Hlth.at the Ctr. for Community Hlth., Univ. of California at San Diego, San Diego, CA, [10]Dept. of Publ. Hlth., Sch. of Med., Pontificia Univ. Católica de Chile, Santiago, Chile, [11]Dept. of Nutrition, Diabetes and Metabolism, Sch. of Med., Pontificia Univ. Católica de Chile, Santiago, Chile, [12]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [13]Div. of BioMed. Informatics and Personalized Med., Sch. of Med., Univ. of Colorado Denver-Anschutz Med. Campus, Aurora, CO, [14]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [15]Dept. of Epidemiology and Population Hlth., Albert Einstein Coll. of Med., New York, NY, [16]Dept. of BioMed. Informatics, Sch. of Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [17]UNC Kidney Ctr., UNC Sch. of Med., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

IntroductionLife course trajectories of hypertension-related traits are associated with increased risk of multiple cardiovascular and cerebrovascular outcomes, with differences between those at low and high polygenic risk for hypertension emerging in the first years of life. Yet, little is known about the genetic etiology of blood pressure (BP) trajectories in adolescents and young adults. Previous genome-wide association studies (GWAS) have identified over 4000 loci associated with BP-related traits. However, most of these findings are from cross-sectional studies of Non-Hispanic white (NHW) populations. To address this gap, we examined the genetic architecture of change in diastolic BP (DBP), systolic BP (SBP), pulse pressure (PP), mean arterial pressure (MAP), and hypertension (SBP >=130 and/or DBP >=80 mmHg and/or antihypertensive medication use) in populations that self-reported diverse race and ethnic backgrounds from 5 cohorts: Population Architecture using Genomics and Epidemiology (PAGE) Study, Santiago Longitudinal Study (SLS), Cameron County Hispanic Cohort (CCHC), Vanderbilt University Medical Center Biobank (BIOVu), and Avon Longitudinal Study of Parents and Children (ALSPAC) cohort.

MethodsWe conducted a GWAS of trajectories of BP-related traits in non-pregnant adolescents and young adults aged 13 to 37 years (n = 29,629; study participants self-identified as 68.9% NHW, 20.3% Non-Hispanic black, 9.3% Hispanic/Latino, 0.9% East Asian, and 0.6% South Asian) with at least two BP readings, measured at least six months apart. First, we fit mixed-effects models and extracted residuals of the slope for each antihypertensive medication use-corrected BP trait adjusted for sex, body mass index, and chronological age at the first BP measurement. The slope residuals were later used as outcomes in the additive GWAS models. We performed fixed effects inverse variance weighted GWAS meta-analysis.ResultsWe identified forty-eight independent associations ($p < 5 \times 10{-}9$) with five BP-related traits, including five novel loci associated with hypertension (KRT18P50, CYCSP17, SLC35F3, LINC01505, and LINC02501), and two (HSDL2 and DAPK1) low-frequency potentially African American-specific loci associated with SBP. The novel independent signals for trajectories of BP traits are associated with body mass index, type 2 diabetes, and kidney and brain function in the GWAS catalog. ConclusionOur findings provide new insights into the genetic architecture of BP trait trajectories in adolescence and early adulthood and highlight the continued importance of conducting genetic studies in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5142T: Overlap between COPD genetic association results and transcriptional quantitative trait loci

Authors:

A. Saferali[1], R. Chase[2], C. Vollmers[3], M. Cho[2], P. Castaldi[2], C. Hersh[2]; [1]Brigham & Women's Hosp., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]UC Santa Cruz, Santa Cruz, CA

Abstract:

Background: Genome-wide association studies (GWAS) have identified genetic loci associated with chronic obstructive pulmonary disease (COPD). When integrated with GWAS results, expression quantitative trait locus (eQTL) studies can provide insight into biological mechanisms involved in disease by identifying single nucleotide polymorphisms (SNPs) that contribute to whole gene expression. However, there are multiple genetically driven regulatory and isoform-specific effects which cannot be detected in traditional eQTL analyses. Here, we identify SNPs that are associated with alternative splicing (sQTL) in addition to eQTLs to identify novel functions for COPD associated genetic variants.**Methods:** We performed RNA sequencing on whole blood from 3743 subjects in the COPDGene Study and lung tissue of 1241 subjects from the Lung Tissue Research Consortium (LTRC) and whole genome sequencing data on all subjects.

Associations between all SNPs within 1000 kb of a gene (cis-) and splice and gene expression quantifications were tested using tensorQTL. In blood, a total of 11,869,333 SNPs were tested for association with 58,318 splice clusters, and 8,792,206 SNPs were tested for association with 70,094 splice clusters in lung tissue.**Results**We identified 28,110 splice-sites corresponding to 3,889 unique genes that were significantly associated with genotype in COPDGene whole blood and 58,258 splice-sites corresponding to 10,307 unique genes associated with genotype in LTRC lung tissue. We found 7,576 sQTL splice-sites corresponding to 2,110 sQTL genes were shared between blood and lung, while 20,534 sQTL splice-sites in 3,518 genes were unique to blood and 50,682 splice-sites in 9,677 genes were unique to lung. To determine what proportion of COPD-associated SNPs were associated with transcriptional splicing, we performed colocalization analysis between COPD GWAS and sQTL data, and found that 33 COPD GWAS loci had evidence of colocalization between QTLs and COPD. The top five colocalizations between COPD and lung sQTLs include *NPNT*, *FBXO38*, *HHIP*, *NTN4* and *BTC.* Long read sequencing validated *NPNT, FBXO38* and *BTC* splice events at the isoform level, revealing that the *FBXO38* splice event likely leads to nonsense mediated decay, resulting in decreased overall *FBXO38* expression in the GWAS risk allele, while the GWAS risk allele in *BTC* results in increased expression of a longer isoform containing a transmembrane domain.**Conclusions:** A total of 33 COPD GWAS loci contain evidence of sQTLs, suggesting that analysis of sQTLs in whole blood and lung tissue can provide novel insights into disease mechanisms. *FBXO38* and *BTC* are of particular interest for follow-up.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5143T: Pathogenic variation in *SNCA* causes allelic biases in transcript use that are identified by long read RNA sequencing and have the potential to drive Parkinson's disease pathology.

Authors:

H. Macpherson[1,2,3], E. Gustavsson[4], J. Evans[5], K. Montgomery[6], S. Garcia-Ruiz[4], D. Murphy[6], N. Wood[7], H. Houlden[7], J. Hardy[6], S. Gandhi[6], M. Ryten[8]; [1]Genetics and Genomic Med., Great Ormond Street Inst. of Child Hlth., Univ. Coll. London, London, UK., London, United Kingdom, [2]Dept. of Neurodegenerative Diseases, Queen Square Inst. of Neurology, UCL, London WC1N 1PJ, UK., London, United Kingdom, [3]Aligning Sci. Across Parkinson's (ASAP) Collaborative Res. Network; Chevy Chase, MD, 20815, Maryland, MD, [4]Univ. Coll. London, London, United Kingdom, [5]The Francis Crick Inst., London, United Kingdom, [6]UCL, London, United Kingdom, [7]Inst Neurology, London, United Kingdom, [8]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom

Abstract:

It is often assumed that disease-associated genetic variants operate through a single primary effect, which is replicated in all tissues. The reality is likely more complex. Determining the impact of genetic variants on transcription based solely on gene structure can be challenging and likely misses important details. For example, alongside changes in protein sequences, missense variants can also affect splicing. Similarly, it is possible that copy number variants (CNVs) can operate through multiple mechanisms to cause disease. Misunderstanding or simplifying possible disease mechanisms in this way can result in incorrect therapeutic strategies. A good example of a well-defined gene with known missense and CNVs is *SNCA*, linked to Parkinson's Disease (PD). Current therapeutic strategies, using antisense oligonucleotides (ASOs), aim to downregulate expression of *SNCA*. However, it is possible that dosage is not the only mechanism by which CNVs and missense mutations could lead to PD. Long-read sequencing allows for better exploration of how CNVs and missense mutations affect splicing, by better defining the underlying gene and transcript structures. These techniques can be powerfully applied in combination with allele specific expression analyses to identify how transcription from an allele is affected by a particular variant. We performed targeted long read sequencing of *SNCA* to explore the impact of CNVs (*SNCA* triplication) and missense variants on *SNCA* transcription in iPSC-derived midbrain dopaminergic neurons. We have developed an allele specific expression pipeline and observe significant changes in allele specific transcript proportions in both variants compared to control cells. This raises important questions on the interpretation of variants and their pathogenesis. Similarly, this may impact the targeting strategy of RNA-based therapeutics such as ASOs. We anticipate that this effect is not unique to *SNCA* and suggest that there will be more examples of allele specific transcript proportion changes in other neurodegenerative disease loci in response to copy number or missense variants.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5144T: Pathway of Genetic Load Analysis Within a High-Risk Population in Nigeria to Analyze Genetic Risk for Developing Kernicterus

Authors:

K. Bussard-Serrano[1], Y. Salsabila[1], N. Kibiryeva[1], D. Bittel[1], J-B. Le Pichon[2], S. Riordan[2]; [1]Kansas City Univ., Kansas City, MO, [2]Children's Mercy, Kansas City, MO

Abstract:

Upwards of 85% of newborns worldwide are affected by jaundice each year. In most developed countries, accessible healthcare and preventative measures ensure this rarely leads to severe disease. However, in many developing countries (Sub-Saharan Africa, Southeast Asia, and parts of China), hyperbilirubinemia leading to kernicterus remains a major cause of life-long disability. When the newborn's body is unable to process bilirubin, blood levels can markedly increase. Though most unconjugated bilirubin in the body remains bound to albumen and unable to cross the blood brain barrier (BBB) some exist in a "free" unbound form capable of crossing the BBB. This lipid-soluble form of bilirubin is neurotoxic and can cause Acute Bilirubin Encephalopathy (ABE) and subsequent irreversible brain injury (kernicterus). Kernicterus is characterized clinically by dyskinetic cerebral palsy and sensorineural hearing impairment. While socioeconomic factors, inadequate healthcare access, and lack of preventative measures play a role, genetic predispositions also contribute to the higher prevalence in these regions. This research is based upon the hypothesis that genetic variations contribute to ABE and kernicterus susceptibility (Riordan et al, 2016). Previously collected samples were used for extraction of DNA and tracked medical records for an at-risk population of infants. DNA samples were collected from 82 newborns with severe hyperbilirubinemia from three different regions in Nigeria (Jos, Kano, and Zaria). Our analysis assessed gene variants associated with ABE and kernicterus susceptibility. We apply the Pathway of Genetic Load (PGL) Analysis previously described to examine two tiers of genes. The first tier includes *G6PD*, *UGT1A1*, among others. Variations in these genes may predispose neonates to increased hemolysis (*G6PD*) or impair the elimination process of unconjugated bilirubin (*UGT1A1*). Meanwhile, the second tier of genes includes those that may impact susceptibility to bilirubin. Total bilirubin (TB) and bilirubin-induced neurologic dysfunction (BIND) scores were used to categorize patients into groups with and without ABE. The infants were then followed clinically for 12 months for the development of kernicterus. Results from this research will expand the current understanding of how certain genetic variations predisposes individuals for hyperbilirubinemia and kernicterus.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5145T: Polygenic predictors of incident cognitive impairment and dementia-associated mortality in the REasons for Geographic and Racial Differences in Stroke (REGARDS) study

Authors:

D. Kim[1], K. A. Mcinerney[1], C. S. Dagli[2], N. C. Olson[3], N. D. Armstrong[2], M. Cushman[3], A. Bennet[2], T. Sofer[4], R. P. Sawyer[5], M. Crowe[2], S. Judd[2], R. Irvin[2], L. M. Raffield[1]; [1]UNC-Chapel Hill, Chapel Hill, NC, [2]Univ. of Alabama at Birmingham, Birmingham, AL, [3]Univ. of Vermont, Burlington,

VT, [4]Beth Israel Deaconess Med. Ctr. / Harvard Med. Sch., Boston, MA, [5]Univ. of Cincinnati, Cincinnati, OH

Abstract:

Late-life dementia has a strong heritable component. Individuals who develop late-life dementia may have vascular contributions to cognitive impairment and dementia (VCID), Alzheimer's disease (AD), or, more frequently, mixed pathology. Measurable AD- and VCID-related endpoints in cohort studies include incident cognitive impairment (ICI) and dementia as a contributing cause of death (DCCD). Cardiometabolic diseases (CMD) and their risk factors predict dementia in many cohorts, as well as ICI in the REasons for Geographic and Racial Differences in Stroke (REGARDS) study, but it is unclear how genetic predictors of CMD may influence risk of dementia-relevant endpoints. We studied the influence of genetic determinants for CMD and AD on ICI and DCCD in a sample with genetic data from REGARDS (83% self-reported Black participants). We used polygenic risk scores (PRS) for both CMD and AD as the exposures. We constructed PRS for five CMD-related traits, including blood pressure (BP; systolic BP, diastolic BP, and pulse pressure (PP) separately), venous thromboembolism, stroke, coronary artery disease, and type 2 diabetes, and AD and related dementias (ADRD) using the genome-wide significant SNPs and their estimated effects reported from the largest and latest available genome-wide association studies for each trait. ICI was defined using serial measures of learning, memory, and verbal fluency/executive function (362 cases from total eligible N: 5863). DCCD was from the National Death Index linkage (353 cases among total eligible N: 8820). We implemented logistic regression models for ICI and Cox proportional hazard models for DCCD. Covariates included age at baseline, sex, APOE2 and APOE4 genotypes, and 10 genetic principal components. The regression analyses were done separately by two self-report racial groups (non-Hispanic White and Black), as genotyping was stratified by race, and meta-analyzed. Adjusting for multiple testing, PRS-PP [HR (95% CI) 1.07 (1.02 - 1.13), p=0.006], PRS-ADRD [HR (95% CI): 1.37 (1.05 - 1.78), with nominal significance (p =0.019)], and the C allele of APOE4 defining variant rs429358 [HR (95% CI) 2.26 (1.88 - 2.72), $p=7.74 \times 10^{-18}$] were positively associated with DCCD. No heterogeneity in the associations was observed across self-report racial groups. No associations were observed for ICI. In conclusion, increased genetic risk for PP contributed to the risk of DCCD in REGARDS, concordant with known relationships between blood pressure and dementia. Known AD genetic risk factors also associated with DCCD. No associations were identified for ICI, suggesting heterogeneity in risk factors for these two dementia-relevant endpoints.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5146T: Predicting and communicating Alzheimer's Disease genetic risk

Authors:

E. Perez[1], A. Bolli[2], S. Klum[2], J. Kintzle[2], P. Di Domenico[3], G. Bottà[3], G. Busby[2]; [1]Brigham and Women's Hosp., Boston, MA, [2]Allelica, New York, NY, [3]Allelica, Rome, Italy

Abstract:

The contribution of genetics to Alzheimer's Disease (AD) risk has long focused on the presence of specific APOE alleles. However, large-scale Genome Wide Association Studies (GWAS) have demonstrated that variants spread across the rest of the genome also influence AD risk. These variants can be combined into polygenic risk scores (PRS) which have the potential to further refine genetic risk estimates. Despite debates surrounding the return of APOE genotype information to patients, the e4 allele, while neither necessary nor sufficient to cause Alzheimer's disease, remains of interest to individuals. To convey the most accurate AD risk, it is crucial to understand how PRS and APOE status together impact risk and to develop reporting strategies to communicate this risk effectively.

We developed a novel PRS for AD by combining 17 PRSs from the PGS catalog with a meta PRS approach, deriving PRS-specific weights from 1,278 cases and 135,375 controls from the UK Biobank (UKB). Within these PRSs, variants in the APOE region were removed to ensure that monogenic effect was not compounded. The performance of the metaPRS was assessed in an independent testing cohort of 2,808 cases and 320,098 controls from the UKB, adjusting for the number of e2 and e4 alleles, age, sex, and principal components of genetic ancestry. We found that monogenic and polygenic factors modulate AD risk both independently and in concert. The PRS had an OR per standard deviation of 1.37 (95%CI: 1.29-1.41) and we discover an interaction between e4 status and the PRS such that the effect of the PRS on AD risk increased by around 7% per SD for each e4 allele carried. We then calculated the RR of AD compared with the average population with risk for different combinations of APOE haplotype carriers and showed that risk in individuals homozygous for the e4 allele ranged from 3 to 25 times the population risk in the lowest and highest PRS percentiles respectively, whereas RR ranged from 1 to 6 times in e3/e4 carriers.

Providing personalized AD risk information has a range of applications, from proactive health and life management in motivated individuals to clinical trial enrollment. We show that combining an AD PRS with APOE status into genetic risk prediction offers a more individualized approach. Important reporting considerations include clinical communication of this risk in a concise and transparent way, acknowledging limitations in generalizability, providing a numerical risk range and addressing the dynamic nature of PRS. Effective translation of PRS plus

APOE models requires clear communication to mitigate misinterpretation and empower informed decision-making by patients and healthcare providers.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5147T: Predicting coronary artery disease using an exome-based estimate of KIV-2 repeat count

Authors:

N. Telis[1], H. Dai[1], D. Wyman[2], B. Khuder[1], F. Tanudjaja[1], S. White[1], N. Washington[3], W. Lee[1], E. Cirulli[1], S. Luo[4], K. Schiabor Barrett[2]; [1]Helix, San Mateo, CA, [2]Helix, San Diego, CA, [3]Helix, Albany, CA, [4]Helix, San Carlos, CA

Abstract:

Introduction: Lipoprotein a (Lp(a)) levels are an often unmeasured predictor of future risk of cardiovascular disease. Unlike LDL, HDL, and triglycerides, Lp(a) levels are believed to be genetically determined; they are not altered by lipid-lowering medications or dietary changes, and are consistent over a lifetime, making them a potential early-life predictor of adult morbidity. However, genetic predictors of Lp(a) are challenging to design. Repeats of the kringle IV subtype 2 domain (KIV-2) contribute significantly but are extremely challenging to quantify via next-generation sequencing; this variability is particularly impactful in non-European ancestries.
Objective: Develop a method to estimate individuals' total number of KIV-2 repeats from exome data and use this estimate to predict the incidence of coronary artery disease (CAD) in our all-comers cohort from health systems across the United States.
Methods: We analyzed ~110K individuals with paired clinical and exome sequencing data from six medical centers. We developed a coverage-based method to determine KIV-2 repeat counts. We validated that these counts associated with clinical Lp(a) measurements derived from the course of care (n = 2039 individuals). We calculated an expected Lp(a) in units of nmol/L using a genetic risk score (GRS) based on 43 variants in the Lp(a) gene. We identified instances of CAD via ICD9CM, ICD10CM and SNOMED terms present in the medical record.
Results: Adding the number of KIV-2 repeats to the GRS substantially improved prediction of clinical Lp(a) measurements, for example improving the r-squared for non-Europeans five-fold (from 0.05 to 0.22). Individuals with fewer than 15 estimated KIV-2 repeats and a high GRS-predicted Lp(a) (high risk group) have a mean measured Lp(a) of that is 10-fold that of those in the lowest category of GRS-predicted Lp(a) (low risk group). In the full cohort, most of whom had not been measured for Lp(a), we find that relative to the low risk group, the high risk

individuals are 1.6x more likely to have CAD.
Conclusion: Lp(a) is a potentially useful biomarker with consistency over a lifetime and implications for future cardiovascular morbidity. We developed a method for estimating the number of KIV-2 repeats, a challenging component of genetic prediction of Lp(a). This estimate can be used in combination with a GRS to predict Lp(a), especially in individuals with non-European ancestry where GRS-based estimates fall short. The final prediction is associated with increased risk of CAD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5148T: Prevalence of *RFC1* repeat expansions in an American idiopathic and diabetic neuropathy patient population

Authors:

K. Warcup[1], S. Ovunc[2], B. C. Callaghan[2,3], P. K. Todd[2,3,1]; [1]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI, [2]Dept. of Neurology, Univ. of Michigan, Ann Arbor, MI, [3]Ann Arbor Veterans Admin. Hlth.care, Ann Arbor, MI

Abstract:

Distal sensory predominant axonal polyneuropathy (DSP) is a common, debilitating, disorder that can result from a wide spectrum of causes, including diabetes mellitus. Many patients are diagnosed with DSP without a clear cause (i.e. idiopathic neuropathy). In 2019, a biallelic non-reference CCCTT/AAGGG repeat expansion in the second intron of *RFC1* was found as the genetic cause of CANVAS (cerebellar ataxia with neuropathy and vestibular areflexia syndrome), a sensory neuronopathy that also features cerebellar ataxia and vestibular dysfunction. Shortly after, a study of Italian patients with idiopathic axonal polyneuropathy revealed that *RFC1* expansions were common in this cohort, suggesting that this expansion may account for many cases of idiopathic neuropathy. The prevalence of *RFC1* expansions in American idiopathic and diabetic neuropathy populations remains unknown. To assess the prevalence of *RFC1* expansions and its influence on the development of diabetic neuropathy, we will investigate the allele frequency of *RFC1* repeat expansions within the Patient Neuropathy Research Registry, a well-characterized biorepository collection of American idiopathic and diabetic neuropathy patient samples. Using an established repeat-primed PCR genotyping method, we will genotype 1000 patient samples at the *RFC1* expansion locus. These data will be compared to existing short-read whole genome sequencing and conformational validation will be performed by long-read sequencing to confirm the size and composition of the repeat allele. We anticipate that this study will provide an accurate estimation of *RFC1* expansion allele

frequencies in an American neuropathy patient population while establishing a high-throughput genotyping workflow for the *RFC1* expansion allele.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5149T: Protection against motor deficits in a Parkinson Disease mouse model, achieved by CRISPR-mediated deletion of a neuronal α-synuclein (*Snca*) intronic enhancer.

Authors:

R. Boyd[1], A. Kho[1], S. Loftus[2], H. Ko[1], A. S. McCallion[1]; [1]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Abstract:

Parkinson Disease (PD) is a neurodegenerative disease that disrupts the quality of life of >8.5 million people worldwide. PD is characterized by the progressive onset of motor and cognitive impairments, and genome-wide association studies have implicated the involvement of >90 loci. Among these, α-synuclein (SNCA) is consistently identified as the greatest genetic risk factor for PD. Preferential degeneration of midbrain dopaminergic (mbDA) neurons and aggregates of SNCA surrounded by Lewy Bodies (LB) are pathognomonic for the disease, and variants that promote SNCA misfolding and overexpression, or events that result in gene amplification, have been linked to PD risk. Therefore, SNCA expression in PD is recognized as a likely therapeutic target. Previous work in mice has shown that Snca knockout is protective against LB formation, neuron loss, and motor impairments induced by intra-striatal injection of SNCA pre-formed fibrils (PFF). Here, we test whether cell-dependent titration of SNCA expression may provide a potential targeted therapeutic. Our group recently identified an intronic enhancer of SNCA and demonstrated that this sequence lies within an open chromatin region (OCR) within mbDA neurons. In transgenic reporter assays, this sequence directs gene expression in catecholaminergic neurons. Furthermore, sequencing this interval led to the discovery of two tightly linked variants (rs2737024 A>G; rs2583959 C>G) strongly associated with PD risk. To test the therapeutic potential of modulating SNCA expression by targeting this sequence, we developed a mouse model containing a 2.7kb deletion of the OCR encompassing these SNPs. Using smFISH, we show that this sequence is an enhancer of Snca; confirmed by a significant reduction in number of Snca copies per cell in mbDA neurons of heterozygous and homozygous enhancer deletion mice (2.98-3.52 fewer copies; p < 0.01) compared to their wild-type littermates. Next, to induce Parkinsonian phenotypes, we performed stereotaxic injections of SNCA PFF and saline (control) into the striatum of 10

mice/sex/genotype. After 6 months, mice undertook a battery of motor assays, including rotarod, grip strength, and pole descent, which revealed that mice lacking the enhancer were significantly protected against PD-like motor phenotypes. We will report these data along with molecular and histological assays to demonstrate the impact of diminishing neuronal Snca expression in this way on the onset and severity of PFF-induced histopathology, including LB formation and DA neuron loss. This work illuminates the therapeutic potential of this sequence in attenuating PD onset and progression.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5150T: PsychGenCOV19 : Mapping Molecular Signatures of COVID-19 Impact on Mental Health

Authors:

D. Davtian, G. Pathak, B. Cabrera-Mendoza, J. He, Y. Nunez, A. Lacobelle, J. Krystal, R. Pietrzak, J. Gelernter, R. Polimanti; Yale Univ., New Haven, CT

Abstract:

Since the emergence of the SARS-CoV-2 virus and its resulting disease, COVID-19, which led to a global pandemic, many studies have investigated mechanisms driving the infection and progression of the virus. To date, however, we still know little regarding the molecular signatures of COVID-19 and their long-term implications. Among health outcomes, COVID-19 has shown a complex relationship with psychiatric disorders and symptoms. In particular, COVID-19 has been associated with an increased risk of depression, anxiety and PTSD, while individuals with psychiatric disorders are more likely to be affected by more severe COVID-19 outcomes. Understanding the molecular mechanisms underlying the interplay between COVID-19 and psychiatric disorders can uncover key aspects of their shared pathogenesis, contributing to their prevention.
At Yale, the PsychGenCOV19 project aims to generate new genetic, epigenetic, and transcriptomic data in a cohort enriched for individuals presenting internalizing symptoms (e.g anxiety or depression). To date, more than 1,000 participants (~40% from underrepresented minorities) have been enrolled with a significant proportion of psychiatric disorders (~58% with depression, ~60% with anxiety and ~30% with PTSD). Leveraging initial blood-based transcriptomic data generated from 681 individuals (70% females), we conducted a preliminary differential gene expression analysis with respect to COVID-19 severity, including age, sex, and genetic principal components as covariates. A total of 118 differentially expressed genes were identified after multiple testing corrections (False Discovery Rate q<0.05). Among them, a strong

association was observed for the Tyrosine hydroxylase (TH) gene, which encodes the enzyme responsible for the conversion of tyrosine to dopamine and has been implicated in multiple psychiatric disorders. These differentially expressed loci were enriched for multiple gene ontology (GO) pathways, including dopamine biosynthetic process (GO:0006585), epinephrine biosynthetic process (GO:0042418), and circadian sleep/wake cycle (GO:0042745). Building on these preliminary results, we will integrate the upcoming genetic and epigenetic data to distinguish different mechanisms between COVID-19 and internalizing disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5151T: Rare protein-coding variants mapping on multi-organ imaging-derived phenotypes ⭐

Authors:

Y. Fan[1], J. Chen[2], Z. Fan[1], R. Wang[3], T. Li[2], H. Zhu[2], B. Zhao[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Regeneron Genetics Ctr., White Plains, NY

Abstract:

Human organ structures and functions serve as the endophenotypes for clinical outcomes. Previous genome-wide association studies (GWAS) have unraveled thousands of common variants associated with phenotypes derived from magnetic resonance imaging (MRI) of the brain and body. However, the role of rare coding variations was largely unknown. Here, we present an exome-wide association study encompassing 596 imaging traits from organs such as the brain, heart, liver, kidney, and lung, involving over 50,000 individuals in UK Biobank. We identified 107 variant-level associations and 224 gene-level associations (67/224 unique gene-trait pairs) across all organs and MRI modalities, such as *PTEN* with total brain volume, *TTN* with regional peak circumferential strain in the left ventricle, and *TNFRSF13B* with spleen volume. The singleton burden model and AlphaMissense annotations further uncovered 8 unique gene-trait pairs including the association between an approved drug target gene *KCNA5* and neuronal activity in the subcortical-cerebellum network. We observed genetic overlaps with other complex traits or diseases and provided coding-wise genetic evidence from the perspective of human organ imaging. More than half (53.4%) of our rare variant associations fell into the loci identified in previous GWAS, including all the associations from two genes (*VCAN*, *PLCE1*) contributing the most gene-trait pairs. The identified genes were enriched for drug targets and may aid drug discovery or elucidate potential side effects. In summary, we demonstrated how

rare variants advanced our understanding of the genetic effects on human organs and the underlying links to related traits or disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5152T: Refining cross-population polygenic risk scores to optimize trait prediction in diverse populations

Authors:

M. Patel[1], G. Chilton[1], H. Wheeler[2]; [1]Loyola Univ. Chicago, Chicago, IL, [2]Loyola Univ Chicago, Chicago, IL

Abstract:

Due to Eurocentricity of available genetic data, polygenic risk scores (PRS) have primarily been trained in populations of European descent. Differences in allele frequencies, effect sizes, and linkage disequilibrium patterns between populations can limit the transferability of PRS across different populations. Given that most data are from European genetic ancestries populations, we assessed how PRS trained in predominantly European populations with a small proportion of other ancestral populations like those in the Pan-UK Biobank performed in diverse populations from *All of Us* (AoU). PRS-CSx is a Bayesian PRS training method that combines genetic effects across populations via a shared continuous shrinkage prior to improve polygenic prediction in diverse genetic ancestries. We applied PRS-CSx to standing height GWAS summary statistics from five populations from the Pan-UK Biobank. These training sample sizes included 972 AMR (American), 419,596 EUR (European), 6556 AFR (African), 2397 EAS (East Asian), and 8657 SAS (South Asian) individuals. We performed a grid search of PRS-CSx phi parameters (1e-2, 1e-4, 1e-6) and three different seeds to assess training model variance. To determine how to weight the resulting PRS from each Pan-UK Biobank population, we validated the PRS-CSx models in 50% of each AoU population and tested the final weighted models in 50% of each AoU population. AoU includes clustering-defined populations of American (n = 39,241), European (n = 120,028), East Asian (n= 5268), African (n= 53,340), and South Asian (n= 2305) genetic ancestries. To evaluate the results, we compared the correlation between PRS-CSx-predicted height and observed height after adjusting for sex at birth, age, and 16 genotypic PCs. Focusing on the three largest AoU test populations, PRS-CSx models validated in the AoU European population performed the best, no matter the AoU test population (European R=0.215 ± 0.0002, American R=0.204 ± 0.0002, African R=0.158 ± 0.00006). However, when we validated in the AoU African population, performance was best in the African AoU test population (R=0.098 ± 0.0005), followed by American (R=0.087 ± 0.0006) and European (R=0.083 ± 0.0004). The

smaller correlation difference between European-validated and African-validated PRS-CSx model performance when we test in African individuals (0.06) compared to the correlation difference in American (0.12) and European (0.13) test populations indicates that increasing African genetic ancestry sample sizes at the training stage of PRS-CSx will improve performance the most compared to other genetic ancestries and should be the focus of future GWAS and PRS training.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5153T: Revealing the developmental-genetic basis for the human cranial vault

Authors:

J. Devine[1], S. Goovaerts[1], A. Pugner[1], N. Claessens[1], M. Yuan[1], J. Shaffer[2], S. Walsh[3], M. Shriver[4], S. Weinberg[2], J. Wysocka[5], P. Claes[1]; [1]KU Leuven, Leuven, Belgium, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Indiana Univ. Indianapolis, Indianapolis, IN, [4]Pennsylvania State Univ., University Park, PA, [5]Stanford Univ., Stanford, CA

Abstract:

The human cranial vault is a complex trait that protects the brain and closely interacts with the cranial base and face due to shared genetic, developmental, and biomechanical inputs. Morphological variation in the vault is both clinically and evolutionary significant, because it is associated with a range of disorders (e.g., craniosynostosis) and adaptations to selective pressures (e.g., childbirth constraints, increases in brain size). Yet the genetics of the vault are often overlooked at the expense of the brain, and even when they are not, these studies tend to be underpowered as radiation exposure and other imaging artifacts can limit sample sizes. Here, we acquire spatially dense, 3-D measures of vault shape from the UK Biobank (N=48,564) via data-driven phenotyping and perform a multivariate genome-wide association study. We recovered 1,510 genome-wide significant loci primarily enriched for processes related to embryonic and postnatal skeletal development, as well as embryonic and skeletal system morphogenesis. The identified loci implicate key craniofacial transcription factors, including known loci near *RUNX2*, *DLX5*, and *TBX15*, and novel loci near *TWIST1/2*, *MSX1/2*, *ALX4*, and *RUNX1*. We additionally identified LRP5, a major Wnt signaling receptor linked to bone density phenotypes and craniosynostosis, then also replicated several other loci linked to sagittal craniosynostosis, including *BMP7*. Our second most significant association was with *BNC2*, one of the most well-supported candidates for adaptive Neanderthal introgression. In total, we found 96 other genes in regions of Neanderthal introgression. This tight interplay

between vault and brain morphology was further confirmed by the fact that nearly 50% of the discovered vault loci were implicated in brain size and shape. Finally, we observed moderate genomic correlations between vault shape and schizophrenia as well as, to a lesser extent, other neurological disorders (e.g., clinical depression). Altogether these results provide the most statistically robust insights into the developmental-genetic basis for human cranial vault variation to date and further illustrate how such normal variation can reveal the biological pathways underlying craniofacial and neurological conditions.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5154T: Risk factors for coronary artery disease in individuals with high plasma Lp(a): A Mendelian randomization study

Authors:

N. Koran[1], L. Dufresne[2], E. MacKenzie[3], Y. Nagai[2], G. Thanassoulis[2,1], J. Engert[1,3,2]; [1]McGill Univ., Div. of Experimental Med., Montreal, QC, Canada, [2]McGill Univ. Hlth.Ctr. Res. Inst., Montreal, QC, Canada, [3]McGill Univ., Dept. of Human Genetics, Montreal, QC, Canada

Abstract:

Background: Coronary Artery Disease (CAD) is the leading cause of death worldwide, and thus an important area for research including causal factors that may identify novel treatments. One such causal factor is lipoprotein a (Lp(a)), which is involved in the development of atherosclerosis and constitutes an independent risk factor for CAD. Plasma Lp(a) levels are a strong independent risk factor for CAD, but variable event rates observed in people with high Lp(a) levels indicate there are other factors at play. Genome-wide association studies (GWAS) have previously identified genetic variants associated with plasma Lp(a) levels. However, Lp(a)'s mechanistic contribution to CAD development remains poorly understood, particularly in terms of its interaction with other risk factors. Mendelian randomization may help identify risk factors that have a stronger effect in individuals with high Lp(a). **Methods:** We conducted a GWAS for CAD in the UK Biobank in subjects with low plasma Lp(a) and again in subjects with high levels based on a threshold of 100 mg/dL. Summary statistics were gathered from non-UK Biobank GWAS for a variety of CAD risk-related traits, including LDL-C, HDL-C, Triglycerides (TG), ApoB, systolic blood pressure (SBP), diastolic blood pressure (DBP), BMI, type 2 diabetes (T2D), smoking (ever), alcoholic drinks per week, C-reactive protein (CRP), and interleukin-6 receptor (IL6R). These SNPs were harmonized with the CAD summary statistics and analyzed through inverse variance weighted and weighted median two-sample mendelian randomization. The results of these MR analyses were compared between high and low Lp(a) groups to identify

differences in risk. We also examined the direct effect of the CAD risk factors in the high and low Lp(a) groups. **Results:** We observed a significant interaction between the dichotomous Lp(a) variable and LDL-C, ApoB, and HDL-C, with an 0.05 increase in OR for LDL-C and ApoB and 0.03 decrease in OR for HDL-C (per standard deviation of sample). Two-sample Mendelian randomization analyses demonstrated differences in CAD risk across Lp(a) groups consistent with the significant results in the observational analyses. SBP, DBP, BMI, alcohol frequency, CRP, Smoking (ever) and T2D all demonstrated no risk difference between the high and low Lp(a) groups across analyses. **Conclusion:** The CAD risk factors LDL-C, ApoB, and HDL-C demonstrated stronger effects in individuals with elevated plasma Lp(a) levels. These results have the potential to better inform clinical guidelines and may also inform our understanding of the biochemical mechanism of Lp(a)'s contribution to CAD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5155T: Severe COVID-19 enriched in participants with G6PD deficiency in the All of Us Research Program dataset.

Authors:

E. Venner[1], H. Nguyen[1], S. Elsea[2], R. Gibbs[3]; [1]HGSC - Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Baylor Coll. Med., Houston, TX

Abstract:

Glucose-6-phosphate dehydrogenase (G6PD) deficiency affects over 400 million individuals worldwide, with over 10% of Black men in the US estimated to carry this genetic metabolic abnormality. G6PD is essential for preventing the depletion of cellular glutathione, which protects cells from oxidative damage caused by reactive oxygen radicals.The worldwide SARS-CoV-2 pandemic has led to many deaths. Patients with COVID-19 illness have a range of disease severity, driven by socioeconomic factors, healthcare utilization, ancestry, and comorbidities. As viral infection can cause the production of reactive oxygen species, leading to oxidative damage, previous studies investigated whether severe COVID-19 is more prevalent in individuals with biochemically assessed G6PD deficiency, finding that Black men under 65 years of age had a 1.5-fold increased likelihood of developing severe COVID-19, defined as requiring hospitalization, ICU admission, mechanical ventilation, or death.Using data from the *All of Us* Research Program, we identified a cohort of patients with severe COVID-19 (defined as having an emergency room, ventilation, or death encounter record associated with SARS-CoV-2 infection). We used the All of Us whole genome sequencing data (v7, n=245,388), and examined known to G6PD deficiency alleles. Of 11,734 participants with a SARS-CoV-2 infection, 347 show G6PD deficiency, and 1027

show variable deficiency. Severe COVID-19 was significantly enriched in the G6PD deficient group (261 / 347, 75% vs 6,361 / 10,360, 61%, z=5.21, p < 0.001).To understand how the relative factors contribute to severe COVID-19, we built a bootstrap aggregation model, using features such as age, BMI, ancestry, comorbidities (anemia, obesity, hypertension), G6DP deficiency, and socioeconomic factors (assisted income pct. in zip code, high school education pct. in zip code, median income in zip code, no health insurance pct. in zip code, vacant housing pct in zip code). Overall, the model performs well to predict severe COVID-19 (overall f1 score 0.98). In addition to BMI and age (effect size 0.27 and 0.18), the percentage with no health insurance in the participant's zip code had a large effect (0.22). Self-reported ancestry (effect sizes 0.006 - 0.01) and comorbidities (0.003-0.004) received lower weights.These analyses highlight the importance of jointly analyzing genetic information with socioeconomic and other environmental data and demonstrate the power of the All of Us data to enable analyses of this type. G6PD deficiency is enriched in patients who develop severe SARS-CoV-2 and may be an important consideration when evaluating risk and treatment approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5156T: Sex-dependent genetic architecture of tobacco-related biomarkers

Authors:

M. Chenoweth[1], C. Lerman[2], J. Knight[3], R. Tyndale[4]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]USC Norris Comprehensive Cancer Ctr., Los Angeles, CA, [3]Lancaster Univ., Bailrigg, United Kingdom, [4]Univ of Toronto, Toronto, ON, Canada

Abstract:

Tobacco use is highly heritable, differs by sex, and dose-dependently increases disease risk. Nicotine is metabolized to cotinine (COT) which is further metabolized to 3'hydroxycotinine (3HC). COT and COT+3HC are biomarkers which capture tobacco intake more accurately than self-reported measures (e.g. cigarettes/day). We conducted a genome-wide genotype-by-sex (GxS) interaction analysis of COT and COT+3HC in treatment-seeking smokers of genetically European ancestry (n=541 males, n=389 females) (NCT01314001). COT and COT+3HC were measured from blood samples collected at baseline. Linear regression models in PLINK2 included genotypes (coded additively), sex, GxS interaction terms, and covariates. For COT, five suggestive ($P<5x10^{-6}$) loci on chr 13, 4, 10, 6, and 12 were identified; the top variant was rs1813692 (3' of *TCB1D4*; beta=-0.24, se=0.05, GxS $P=1.29x10^{-6}$). Mean COT (in ng/ml) was 208, 230, and 253 in rs1813692 AA, AT, and TT females, compared to 276, 253, and 228 in males,

respectively. For COT+3HC, three suggestive loci on chr 18, 17, and 5 were identified; the top variant was rs4427876 (3' of *MIR924HG*; beta=0.30, se=0.06, GxS $P=7.39x10^{-7}$). Mean COT+3HC (in ng/ml) was 506, 378, and 318 in rs4427876 CC, CT, and TT females, compared to 262, 326, and 345 in males, respectively. Six of these eight suggestive loci were associated with tobacco-related (e.g. smoking initiation) and/or psychiatric (e.g. schizophrenia, depression, alcohol dependence) traits in the GWAS Catalog. A more granular understanding of the factors that differentially influence tobacco intake in women and men may identify risk factors for heavier use and sex-specific opportunities to promote cessation and mitigate disease risk.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5157T: Sex-Specific genetic insights into Alzheimer's Disease: integrative analysis of genomic, transcriptomic, and phenotypic Data

Authors:

Y. Jin[1], P. Paschou[2], A. Topaloudi[3], S. Shekhar[3]; [1]Purdue, West Lafayette, IN, [2]PURDUE Univ., West Lafayette, IN, [3]Purdue Univ., West Lafayette, IN

Abstract:

Despite the increased susceptibility of women to Alzheimer's disease (AD), there is a lack of sex-specific genetic studies and understanding of sex-specific backgrounds for the disorder. This study aims to address the heterogeneity of sex-specific AD by integrating seven different genomic and neuropathology datasets, including 2,660 female samples (2,011 cases, 649 controls) and 2,366 male samples (1,676 cases, 690 controls). We performed GWAS analyses and integrated the findings with transcriptomic, proteomic, and phenotypic data to identify genetic biomarkers for different sexes in AD progression.As expected, the top locus in all three AD neuropathology phenotypes studied (ncAD, Braak stage, and NP score) was 19q13.32, with top SNPs located in the APOE region for both sexes. No additional significant loci were identified in the male-specific GWAS for any phenotypes. However, three additional loci exceeded genome-wide significance in the female-specific GWAS: 2q14 (close to BIN1) and 4q27 (close to QRFPR) were significantly associated with ncAD, and rs17030228 (close to LOC102723854) was significant in the Braak stage GWAS. Gene-based analysis revealed that genes in the 19q13.32 region were significant for all three AD traits in both sexes, while the SGCZ gene was significant for female-specific ncAD. The sex heterogeneity test showed significant differences in the effects of rs4663105, rs17030228, and rs77285108 between males and females.Tissue enrichment analysis indicated a significant association between NP score for females and genes expressed in ovary tissue. To uncover proteins whose genetic expression correlates with neuropathological

aspects of AD, we conducted a Proteome-wide association study (PWAS). PWAS identified one protein, DGKE, significantly associated with ncAD for males. Notably, DGKE was a novel finding not previously implicated in AD-related studies. We continue to perform PheWAS to identify associations between genetic and phenotypic variation in European populations from the UK Biobank dataset. PheWAS analysis identified eight blood assay outcomes significantly associated with sex-specific PRS, with six for females and two for males. Additionally, we found three outcomes significantly associated with risk scores: celiac disease for ncAD females PRS, congenital anomalies of muscle, tendon, fascia, and connective tissue for Braak stage females PRS, and Alzheimer's disease for Braak stage males PRS.These findings highlight the importance of considering sex-specific genetic factors in AD research and provide new insights into the functional and biological mechanisms underlying the disease.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5158T: Sex-specific genetic underpinnings of Alzheimer's disease and body mass index

Authors:

A. Kulminski, D. Gerardo, F. Feng, I. Culminskaya, A. Nazarian, E. Loiko, Y. Loika; Duke Univ, Durham, NC

Abstract:

The Alzheimer's Association reports that more than two-thirds of the approximately 5 million Alzheimer's disease (AD) cases in the USA are women. Gaining insights into the sex-specific genetic architecture of AD and its risk factors could enhance the understanding of the mechanisms of sex disparities in health. We conducted genome-wide association studies (GWAS) for AD and its risk factor, body mass index (BMI), in men and women of European ancestry separately. The AD GWAS involved 168,962 men (including 6,229 cases) and 199,668 women (including 7,628 cases) from 10 cohorts. The BMI GWAS included 165,111 men and 194,074 women from six of the 10 AD cohorts. Both GWAS identified loci characterized by genetic associations reaching genome-wide significance ($p < 5\times10^{-8}$) in at least one sex, with six loci for AD and 196 for BMI, including overlapping APOE gene clusters. Sex-specific associations in these loci were determined by assessing the significance of differences in effects between sexes using the Wald chi-square test at a false discovery rate $q < 0.05$. Our analysis identified two sex-specific loci for AD and 125 for BMI. Notably, the AD sex-specific loci showed protective associations in men (beta=-0.17, 95% CI[-0.22,-0.11]) and no significant associations in women (beta=0.02, 95% CI[-0.02,0.07]) in the SERPINE2 gene cluster (e.g., rs6758638), and a

smaller AD risk for men (beta=1.16, 95% CI[1.10,1.22]) than for women (beta=1.35, 95% CI[1.30,1.41]) in the APOE gene cluster (e.g., the epsilon4-encoding rs429358 polymorphism). The identified sex-specific loci provide promising targets for gaining insights into sex disparities in AD and its risk factor, BMI.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5159T: Sex-Stratified Genome-Wide Association Study of Sulfonylurea Drug Response in Individuals with Type 2 Diabetes

Authors:

J. Breeyear[1], J. S. House[1], F. Akhtari[1], M. Kvale[2], H. S. Shah[3], M. Hedderson[4], K. M. Giacomini[2], J. Mychaleckyi[5], A. Doria[6], M. Wagner[7], J. B. Buse[8], D. M. Rotroff[9], A. A. Motsinger-Reif[10]; [1]Natl. Inst. of Environmental Hlth.Sci., Durham, NC, [2]Univ. of California San Francisco, San Francisco, CA, [3]Harvard Med. Sch., Cambridge, MA, [4]Kaiser Permanente Northern California, Oakland, CA, [5]Univ. of Virginia, Charlottesville, VA, [6]Harvard Med. Sch., Boston, MA, [7]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [8]Univ. of North Carolina Sch. of Med., Chapel Hill, NC, [9]Cleveland Clinic, Cleveland, OH, [10]Natl. Inst. for Environmental Hlth.Sci., Durham, NC

Abstract:

Sulfonylureas are effective antihyperglycemic agents that have been a cornerstone of Type 2 diabetes therapy for over 50 years. Pharmacogenomics has been applied to sulfonylureas to elucidate the genetic underpinnings of response variability, but so far has been limited to candidate gene studies focused primarily on either drug target genes, metabolizing enzymes, or diabetes risk genes. Furthermore, clinical studies have demonstrated sex differences in sulfonylurea response. Here we present the results of the first sex-stratified genome-wide association study (GWAS) of sulfonylurea response with data from the Action to Control Cardiovascular Risk in Diabetes (ACCORD) clinical trial and the Pharmacogenomics of Metformin (PMET1) study. We performed multi-ancestry genome-wide association studies of sulfonylurea response using common variants (MAF>0.01) in ACCORD, stratified by sex (male-only (n_{max} = 2,428); female-only (n_{max} = 1,957)) as well as combined (n_{max} = 4,385), followed by validation (same direction of effect, $p<0.05$) of the significant ($p<5x10^{-08}$) and nominal ($p<5x10^{-06}$) loci in PMET1. We performed gene-based analyses with MAGMA in FUMA. The combined, male-only, and women-only analyses identified five, eleven, and two loci associated with differential drug response at genome-wide significance, with additional loci reaching nominal significance. The significant loci
included *KAZN*, *KIF2B*, *SLC39A10*, *SPINK5*, *VSX2* (combined); *BCAS3*, *CRACR2A*, *KCNK2*, *TENM2*, *V*

SX2 (male-only); and *SLC30A8* (female-only). The significant loci *KAZN* and *BCAS3* as well as the nominal loci *PIK3R1*, *THEM7P*, *DLGAP1*, *TCEA3*, and *TMEM64* from the combined and male-only strata were validated in PMET1. From the gene-based test, we identified four nominal ($p<2.5x10^{-05}$) genes, including *CALCA*, *OR4F21*, *OR5K1*, and SPINK5. Previously reported associations of *POR* and *ABCC8* were replicated (p<0.01). Even with the modest sample size, we identified biologically relevant loci with known relevance to pathways involved in sulfonylureas' mechanism of action. As sulfonylureas stimulate insulin secretion through membrane depolarization by decreasing intracellular potassium leading to an influx of calcium, genes involved in potassium and calcium transport (*CALCA*, *CRACR2A*, *KCNK2*, *TMEM64*, *TENM2*) may play a role in the efficacy of sulfonylureas. Furthermore, recent work suggests a symbiotic relationship between microtubules pathways (*KIF2B*, *BCAS3*) and insulin secretion (*PIK3R1*, *SLC30A8*). Overall, these results demonstrate sex-biased results that identify potential alternate mechanisms impacting sulfonylurea response that merit further investigation.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5160T: Shared genetic architecture of brain shape and IBD is specifically enriched for inflammatory response genes

Authors:

S. Goovaerts, J. Devine, N. Claessens, M. Yuan, T. Kivisild, I. Cleynen, P. Claes; KU Leuven, Leuven, Belgium

Abstract:

The gut-brain axis (GBA), a bidirectional communication network that connects the central nervous system and the enteric nervous system of the gastrointestinal tract, is central to human health and disease. Its dysregulation plays a significant role in the pathophysiology of inflammatory bowel disease (IBD) as well as several psychiatric and neurodegenerative conditions. Therefore, a better genetic understanding of the GBA is key to understanding disease mechanisms and developing targeted therapies. Here, we conduct a genome-wide association study (GWAS) on multivariate cortical brain shape from 48,564 participants of the UK Biobank, yielding 1,215 independent genomic loci. Using the conjunctional false discovery rate (FDR), the resulting summary statistics were subsequently leveraged to investigate shared genetic signals with IBD based on previously published summary statistics on 25,042 cases and 34,915 controls. This yielded 217 shared genomic loci at 1% FDR, harboring genes involved in neuronal differentiation, GABAergic synaptic signaling, stress-response, and cell-cell adhesion. Additionally, many shared loci contain genes related to leukocyte activation; signaling involving

cytokines; inflammatory response; and immune response more broadly, including *IL23R*, *IL1R1/2*, *IL4/5/10/13/27*, *TNFRSF14*, *LTBR*, *ICOS*, *OSM*, *IRF2BP2*. These pathways were also specifically overrepresented among the shared loci versus all loci associated with brain shape, suggesting that immune-related processes specifically underlie the GBA. Together, these results align with existing knowledge on IBD pathophysiology and symptomology and suggest a shared genetic basis of the brain and gut conditions in the immune system. The specific enrichment of shared loci for inflammatory response genes may suggest that the genes that cause inflammation in the gut also cause inflammation in the brain with measurable effects from neuroimaging. This is especially relevant given the hypothesized relationship between neuroinflammation and psychiatric conditions. Altogether, these results create a foundation for further follow-up regarding the conditions affected by the GBA.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5161T: Single-cell profiling of proliferating T-cells reveals division-state-specific gene expression dynamics and links immune-disease variants to cell division states ⭐

Authors:

M. Gozzard[1], O. B. Bakker[1], X. Ibarra-Soria[2], G. Griffiths[1], T. S. Rupall[1], F. Lichou[1], N. Srinivasan[2], C. P. Jones[1], N. Nakic[2], A. Lorenc[1], G. Trynka[1]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]GlaxoSmithKline, Stevenage, United Kingdom

Abstract:

Immune disease variants point to the causal role of T-cell activation in disease pathogenesis. Activation of CD4+ T-cells triggers a cascade of dynamic changes, propelling cells towards cell division. Yet, T-cells exhibit distinct proliferation rates despite receiving homogenous *in vitro* stimulation. Activation is typically studied along a time course following stimulation rather than focusing on the extent of cellular response. To address this gap, we generated a high-resolution single-cell map of CD4+ T-cell proliferation.We activated three CD4+ T-cell subtypes: naive, memory, and regulatory cells. We profiled cells across five generations of division using single-cell transcriptomics and expression of 130 surface proteins. Our multimodal map encompasses 115,013 single cells, with 15,971 - 24,402 cells at individual generations.We identify 6701 division-dependent genes differentially expressed between generations. We detect established proliferation genes, affirming our trajectory while unveiling novel biological insights. We observed that increased cell divisions lead to changes in proinflammatory cytokine signalling across cell subtypes and major changes to metabolic pathways. At even greater

resolution, we identified division-dependent effects specific to cell subtypes, e.g. oxidative phosphorylation and MYC signalling were highly upregulated in memory and regulatory but not naive cells, highlighting strong metabolic differences between CD4+ T-cell subtypes throughout expansion.We observed that 72 genes with division-dependent changes were previously identified as T-cell activation eQTLs, 70 were subtype-specific. We previously showed 127 T-cell activation eQTLs colocalize with immune-diseases. Now, we demonstrate that 10 genes show cell division-dependent changes. For example, *IRF5* colocalizes with numerous immune-diseases, and its expression peaks at late activation time points in naive T-cells. Our proliferation data deconvolutes this signal further, *IRF5* expression peaks in 3rd-generation of divided cells and remains at high levels. This approach provides an additional temporal dimension resolving gene expression regulation dynamics.We provide a new level of granularity for understanding T-cell activation, highlighting subtype-specific and division-dependent gene expression changes. These insights provide a novel biological interpretation of the mechanisms through which immune disease genes regulate T-cell activation.The human biological samples were sourced ethically and their research use was in accord with the terms of the informed consents under an IRB/EC approved protocol.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5162T: Single-cell transcriptomics of nasal swabs identifies distinct viral signatures of pediatric respiratory syncytial virus infection

Authors:

T. Bradley, S. Khanal, C. Schwendinger-Schreck, D. Louiselle, R. McLennan, T. Pastinen, R. selvarangan, E. Grundberg; Children's Mercy Res. Inst., Kansas City, MO

Abstract:

Respiratory Syncytial Virus (RSV) is a leading cause of respiratory infections in young children, often resulting in severe clinical outcomes. Understanding the cellular and molecular mechanisms of RSV infection is critical for developing targeted therapies. Recent single-cell genomics studies, using tracheal aspirate and bronchoalveolar lavage, have elucidated pathologic immune responses, most notably by innate immune cells, that contribute to RSV disease severity. However, these approaches rely on invasive sampling typically restricted to critically ill and mechanically ventilated patients preventing insight into biological and immune response mechanisms underlying the complete spectrum of RSV infection in children. To this end, we piloted single-cell transcriptomics of epithelial and immune cells derived from nasal swabs routinely collected in emergency rooms or out-patient clinics for a suspected respiratory

viral infection. Specifically, within 12 hours of clinical testing, we obtained nasal swabs from pediatric patients under 5 years of age that were diagnosed with acute RSV infection and performed single-cell RNA sequencing to generate a comprehensive cellular atlas of the RSV infected nasal mucosa. We showed striking neutrophilia similar to what has been observed in lung samples from RSV-associated acute respiratory distress syndrome in children. Next, by performing combined mapping of the host and RSV transcriptome, we identified monocyte/macrophage populations as a primary cellular target of RSV. Infected macrophages exhibited a distinct proinflammatory phenotype, characterized by upregulated expression of cytokines and chemokines. Additionally, we observed alterations in gene expression related to antiviral responses and cellular metabolism, indicating a complex reprogramming of macrophage functions upon RSV infection. Using monocyte in vitro models, we also demonstrated that monocytes were permissive to RSV-infection, and RSV-infected cells exhibited increased inflammatory and antiviral responses. In conclusion, we provide novel insight into the cellular targets and molecular dynamics of RSV infection in the pediatric nasal mucosa. The identification of macrophages as key players in the proinflammatory response to RSV highlights potential therapeutic targets for mitigating the severe inflammatory responses associated with RSV infections. These findings underscore the utility of single-cell genomics in easily accessible clinical samples to unravel host-pathogen interactions at a granular resolution, paving the way for precision medicine approaches in pediatric respiratory diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5163T: Spatial transcriptome profiling of the pancreas uncovers novel targets and pathways in *Clec16a* knockout mice.

Authors:

M. Bakay[1], R. Pandey[1], B. Strenkowski[1], M. Matos[1], C. Kao[1], K. Nagaraju[2], H. Hakonarson[1,3]; [1]Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Sch. of Pharmacy and Pharmaceutical Sci., Binghamton Univ., Binghamton, NY, [3]Dept. of Pediatrics, the Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Multiple GWAS and functional studies have established CLEC16A as a significant gene in autoimmunity and as regulator of abnormal autoimmune responses. However, the precise mechanism of its actions remains largely unknown. To investigate the role of CLEC16A in autoimmunity we generated an inducible ubiquitous *Clec16a* KO mouse. The global *Clec16a* KO (*Clec16a*$^{\Delta UBC}$) exhibit a complex phenotype with marked immune dysfunction. Loss of CLEC16A

leads to an abnormal mitophagy, cell death and immune dysfunction. Recently, we reported striking pathological differences in the spleen, thymus, and pancreas of Clec16a$^{\Delta UBC}$ mice. Notably, the pancreas of Clec16a KO mice exhibited significant degeneration and immune cell infiltration, which were absent in the control group. Here, we employed spatial profiling (GeoMx) to further explore the relationship between inflammation and the mitophagy/autophagy process in the pathology of CLEC16A deficiency. We also aimed to identify new therapeutic targets to counteract the effects of *Clec16a* KO. Pancreas were harvested at five days post-KO-induction with tamoxifen. We used the morphology markers (CD45, pan-CK, nuclear stain) focusing on immune-rich regions to select the regions of interest (ROI). Total of 15 ROIs were identified and studied. Using GeoMx we achieved a technical success with >16,000 genes detected for RNA testing. We observed significant differences in cell deconvolution in control vs KO. Also, 1,605 genes were differentially expressed (DE) between KO and controls ($p<0.05$). Thus, DE genes accounted for 10% of all detected genes. We discovered a novel CXCL16/CXCR6 chemokine axis associated with Clec16a deficiency. This axis has been implicated in various pathological pancreatic conditions, including chronic pancreatitis, diabetes, and cancer. Under certain pathological conditions, CXCL16 is expressed in acinar cells where CXCL16 blockade can ameliorate damage in acute pancreatitis mouse model. In summary, 2D spatial transcriptome profiling has uncovered promising new insights in the Clec16a mouse model. Despite the small sample sizes in this pilot study, informative discoveries were made. Replication of results, additional time points and validation of functional significance are needed to confirm utility of the findings.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5164T: Targeted Resequencing of Otosclerosis Patients from Different Populations Replicates Results from a Previous Genome-Wide Association Study

Authors:

E. Fransen[1], L. Tavernier[1], T. Vanpoucke[1], I. Schrauwen[2], G. Van Camp[3]; [1]Univ. of Antwerp, Edegem, Belgium, [2]Columbia Univ., New York, NY, [3]Univ Antwerp, Antwerp, Belgium

Abstract:

Otosclerosis is one of the most common causes of hearing loss in young adults. It has a prevalence of 0.3-0.4% in the European population. Clinical symptoms usually occur between the second and fifth decade of life. Different studies have been performed to unravel the genetic architecture of the disease. Recently, a genome-wide association study (GWAS)

identified 15 novel risk loci and replicated the regions of three previously reported candidate genes. In this study, seven candidate genes from the GWAS were resequenced using single molecule molecular inversion probes (smMIPs). smMIPs were used to capture the exonic regions and the 3' and 5' untranslated regions (UTR). Discovered variants were tested for association with the disease using single variant and gene-based association analysis. The single variant results showed that 13 significant variants were associated with otosclerosis. Associated variants were found in five of the seven genes studied here, including *AHSG, LINC01482, MARK3, SUPT3H* and *RELN*. Conversely, burden testing did not show a major role of rare variants in the disease. In conclusion, this study was able to replicate five out of seven candidate genes reported in the previous GWAS. This association is likely mainly driven by common variants.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5165T: The Brain on Bilirubin - Genetic Analysis of Hyperbilirubinemia and Kernicterus Susceptibility in Nigerian Newborns

Authors:

Y. Salsabila[1], K. Bussard-Serrano[1], D. Bittel[1], N. Kibiryeva[1], J-B. Le Pichon[2], S. Riordan[2]; [1]Kansas City Univ., Kansas City, MO, [2]Children's Mercy Hosp., Kansas City, MO

Abstract:

Neonatal jaundice, commonly caused by elevated levels of unconjugated bilirubin, affects approximately 60-80% of newborns. Since the physiological systems involved in bilirubin processing takes time to mature, it is not uncommon to see temporary hyperbilirubinemia and jaundice in newborns. However, when this process becomes impaired and bilirubin is allowed to accumulate, neurological issues may arise. Most of the unconjugated bilirubin in the blood are bound to albumin, but some are not bound to any protein (“free bilirubin”). It is important for free bilirubin to be properly eliminated from the body as this is the toxic form that can cross the blood-brain barrier and damage the brain. Hyperbilirubinemia can lead to Acute Bilirubin Encephalopathy (ABE) which can progress to kernicterus if not promptly treated. The clinical presentation of kernicterus may involve movement disorders, cerebral palsy, hearing loss, among others. Certain brain regions appear to be more sensitive to bilirubin, especially those that regulate motor control and movement, auditory processing, and oculomotor function. While many factors play a role in the development of hyperbilirubinemia and kernicterus, studies have found an association with certain gene variants. For instance, gene variants and co-expression of G6PD and UGT1A1 are known contributing factors for hyperbilirubinemia and kernicterus susceptibility. This research will further examine the gene with differential

expression in the bilirubin susceptible versus resistant brain regions (derived from the Allen Brain Atlas). We hypothesize that genetic variation may contribute to kernicterus and brain regional susceptibility to bilirubin (Riordan et al, 2016). Here we present the analysis of DNA sequence data collected from 82 newborns from three different regions in Nigeria (Jos, Kano, and Zaria). Total bilirubin (TB) levels and the Bilirubin-induced Neurological Dysfunction (BIND) score were used to categorize infants into groups with and without ABE. The infants were then followed clinically for 12 months to monitor for the development of kernicterus. These analyses will expand the current knowledge of hyperbilirubinemia and kernicterus susceptibility, especially in an area of the world where socioeconomic factors and genetic predisposition play key roles in the increased occurrence of kernicterus.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5166T: The clinical benefits of genetically guided population health management ⭐

Authors:

J. O'Sullivan[1], M. Chuong[2], D. Thompson[2], R. Moore[2], M. Weale[2], G. McVean[3], F. Aguilera[2], P. Donnelly[2], E. Ashley[1], V. Plagnol[2], S. Harrison[2]; [1]Stanford Univ., Stanford, CA, [2]Genomics plc, London, United Kingdom, [3]Genomics Plc, London, United Kingdom

Abstract:

Background: As polygenic risk scores (PRS) have demonstrated their predictive validity across a range of cohorts and diseases, quantifying their clinical utility is necessary for broad acceptance. Because PRS are well suited to improve prevention and screening based on an individual's genetic risk profile, we decided to quantify the extent to which event and mortality rates could be reduced by adapting existing screening programmes based on PRS data.
Methods: Using UK Biobank incident data (n = 444,949), we explored the clinical benefits associated with measuring PRS across 21 diseases. We identified high genetic risk individuals (defined as PRS OR >2), and assessed all-cause mortality rates by genetic risk. For nine diseases with existing screening programmes/initiatives globally (melanoma, breast/prostate/colorectal cancer, abdominal aortic aneurysm, type 2 diabetes, hypertension, cardiovascular disease and atrial fibrillation) we estimated PRS-guided screening ages for high risk individuals, based on equivalence of risk with normal starting age. We then used published estimates of mortality benefit associated with these interventions to quantify the value of tailoring guidelines to genetic risk. Empirical event and mortality rates were measured in a held out set of UK Biobank to avoid overfitting.

Findings: 60.9% of the overall cohort was at high risk (odds ratio, OR >= 2), and 16.7% at very high risk (OR>= 3) for at least one of the 21 diseases. Across nine diseases with screening programmes, very high risk individuals reached the risk level associated with screening initiation on average 12.4 years earlier, and low risk individuals (OR < 0.5) reached that level 17.7 years later. Mortality was elevated for high genetic risk individuals: individuals in the top 20% of the aggregate genetic risk had an observed mortality risk 40% higher in women (HR: 1.42 (95%CI: 1.34 to 1.50), and 50% higher for males (HR: 1.51 (95%CI: 1.44 to 1.58) compared to individuals in the bottom 20% risk in the held out testing set. Across the same nine diseases, per 100,000 individuals, we estimate that tailoring screening programmes to genetic risk would present opportunities for early detection and/or prevention of 4,294 disease events that would otherwise occur before normal screening started, and save 120 disease-specific deaths.
Conclusion:While the benefit associated with any one disease is limited, knowledge of genetic risk, measured using PRS, has the potential to deliver substantial public health benefits when aggregated across conditions provided that existing screening/prevention programmes can be tailored accordingly.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5167T: The impact of donor and recipient polygenic burden on kidney transplant outcome ⭐

Authors:

K. Collins[1], E. Gilbert[1], V. Mauduit[2], K. Benson[1], E. Elhassan[3], C. O'Seaghda[3], C. Hill[4], A. McKnight[5], A. P. Maxwell[6], P. J. van der Most[7], M. H. De Borst[7], W. Guan[8], P. A. Jacobson[8], A. K. Israni[8], B. Keating[9], G. L. Lord[10], S. Markkinen[11], I. Helanterä[12], K. Hyvärinen[11], J. Partanen[11], S. F. Madden[1], S. Limou[2], G. L. Cavalleri[1], P. J. Conlon[3]; [1]Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [2]Nantes Univ., Nantes, France, [3]Beaumont Hosp., Dublin, Ireland, [4]Queen's Univ. Belfast, Belfast, Ireland, [5]Queen s Univ. of Belfast, Belfast, United Kingdom, [6]Queen's Univ. Belfast, Belfast, United Kingdom, [7]Univ. of Groningen, Univ. Med. Ctr. Groningen, Groningen, Netherlands, [8]Univ. of Minnesota, Minneapolis, MN, [9]Univ. of Pennsylvania, Philadelphia, PA, [10]Univ. of Manchester, Manchester, United Kingdom, [11]Finnish Red Cross Blood Service, Helsinki, Finland, [12]Helsinki Univ. Hosp., Helsinki, Finland

Abstract:

Background and hypothesis:
Many factors influence graft outcome, including donor and recipient age and sex, HLA mismatch, and donor type. Nevertheless, much of the variation in transplant outcomes is still

unexplained. We hypothesise that donor and recipient polygenic burden for traits related to kidney function, as well as a combination of donor and recipient polygenic burdens, may influence graft function post-transplant.

Methods:

We assembled a cohort of 6,060 living and deceased kidney donor-recipient pairs from seven transplant cohorts. Using PRSice2, we calculated polygenic risk scores (PRS) for kidney function related traits including hypertension, estimated glomerular filtration rate (eGFR), and intracranial aneurysm (IA) in both kidney donors and recipients. Donor and recipient PRS were discretized into high (top 10%), intermediate (middle 80%), and low (bottom 10%) burden. We investigated the association between donor PRS, recipient PRS, and combined donor and recipient PRS, and graft survival as well as recipient eGFR at 1 and 5 years post-transplant.

Results:

A standard deviation increase in donor PRS for hypertension, IA, and low eGFR was associated with a reduction in recipient eGFR at 1 year of 0.80, 0.75, and 1.29 mL/min/1.73m^2 (p=0.001, 0.002, 0.02 respectively). A standard deviation increase in recipient PRS for low eGFR was associated with a reduction in eGFR at 1 and 5 years post-transplant of 0.95 and 1.12 mL/min/1.73m^2 respectively (p=0.005, 0.001). Recipient eGFR at 1 year post-transplant was 14 mL/min/1.73m^2 lower when both donor and recipient had high PRS burden for low eGFR than when they both had low PRS burden. For a recipient with high PRS burden for low eGFR, their eGFR at 1 year post-transplant is 51 mL/min/1.73m^2 with a high burden donor and 57 mL/min/1.73m^2 with a low burden donor.

Conclusions:

The combination of donor and recipient PRSs has an impact on graft function. Much of the risk of having a high PRS burden recipient can be mitigated by having a low or intermediate PRS burden donor. These findings could play an important role in improving living donor transplant allocation decisions in the future.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5168T: The interactions of mosaic loss events with genetic risk are strongly associated with schizophrenia and bipolar disorders. ⭐

Authors:

S. Uchiyama[1], X. Liu[1], S. Takeo[2], Y. Kamatani[3], I. Masashi[4], C. Terao[1,5,6]; [1]RIKEN IMS, Yokohama, Japan, [2]Dept. of Psychiatry, Fujita Hlth.Univ. Sch. of Med., Toyoake, Aichi, Japan, [3]The Univ. of Tokyo, Minato-ku, Japan, [4]Dept. of Psychiatry, Nagoya Univ. Graduate Sch. of Med., Nagoya,

Aichi, Japan, [5]Clinical Res. Ctr., Shizuoka Gen. Hosp., Shizuoka, Japan, [6]Dept. of Applied Genetics, The Sch. of Pharmaceutical Sci., Univ. of Shizuoka, Shizuoka, Japan

Abstract:

Objectives: While the mosaic chromosomal alternations (mCAs) accumulate in the normal brain neurons during the developmental period, the contributions of mCAs to psychiatry disorders, including schizophrenia (SCZ) and bipolar disorders (BD), are not well unknown. **Methods:** mCAs in peripheral blood samples obtained from two Japanese cohorts (SCZ-dataset: 2,470 cases and 177,773 controls, BD-dataset: 3,732 cases and 177,773 controls) were detected by the Mosaic Chromosomal alterations (MoChA) pipeline. We evaluated the associations of mCAs with SCZ or BD using the logistic regression models adjusted by sex, age, age^2. We further assessed the associations of cell fraction (CF)>5% or >10% of mosaic events with SCZ or BD in the same manner as above to evaluate if the associations between these disease phenotypes and mCAs increased in the CF-dependent manner. We then stratified subjects into eight categories based on the genetic risk (low or high) and the mosaic loss status (positive, negative, CF<5%, or CF>5%) and evaluated the risks of SCZ or BD in each group. We separately assessed the following mosaic events: autosomal mCA, mosaic loss of X, and mosaic loss of Y, respectively. **Results:** We found autosomal mCAs, especially mosaic loss events, are significantly increased in SCZ [Autosomal mCAs; Odds Ratio (OR) = 1.72, P = 1.2×10^{-14}, loss; OR = 1.92, P = 2.7×10^{-7}] adjusted by sex, age, age^2. Of note, we found similar significant associations of autosomal mCAs and loss events with BD. These associations were highlighted in the young subjects in each dataset. Furthermore, we found the presence of CF>5% and 10% of loss events are positively associated with SCZ, showing the effect sizes increased in a CF-dependent manner, indicating the loss events of high CF contribute SCZ more than low-CF loss events (CF>5%; OR = 3.08, P = 4.2×10^{-21}, CF>10%; OR = 3.54, P = 2.9×10^{-18}). Intriguingly, similar significant associations between BD and CF>5% and 10% of loss events were also observed. Finally, interaction analysis showed that among low PRS groups, both mosaic loss (+) and mosaic loss-CF>5% subjects showed a significant slight increase in the risk of SCZ compared with low PRS and mosaic loss (-) subjects. Strikingly, among high PRS groups, the risk of SCZ was strongly and significantly increased in a mosaic loss CF-dependent manners (loss; OR = 4.08, P = 3.9×10^{-18}, loss-CF > 5%; OR = 6.61, P = 1.3×10^{-24}). The same positive interaction was also observed in BD subjects. **Conclusions:** mCAs, especially loss events, are strongly associated with SCZ and BD, modifying these disease genetic risks to accelerate the development of SCZ and BD.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5169T: The landscape of structural variation in bipolar disorder and schizophrenia across the frequency spectrum of diverse populations ⭐

Authors:

C. Liao[1], The Bipolar Exome Consortium (BipEx), The Schizophrenia Exome Sequencing Meta-analysis Consortium (SCHEMA), Genome Aggregation Database consortium (gnomAD), M. Daly[2], B. Neale[2], M. Talkowski[2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Bipolar disorder (BD) and schizophrenia (SCZ) are debilitating neuropsychiatric disorders that significantly impact cognitive and emotional functioning. Both disorders pose substantial challenges for diagnosis and treatment due to their overlapping symptomatology and shared genetic underpinnings. Recent advancements in genomic technologies have highlighted the role of structural variants (SVs) in neuropsychiatric disorders. Here, we present a comprehensive assessment of common and rare structural variation implicated in BD/SCZ across ancestrally-diverse populations with over 40% of participants being non-European genetic ancestry. We generated the largest dataset to date of rare protein-coding CNVs for 41,476 SCZ, 34,294 BD and 98,956 controls using GATK-gCNV from whole-exome sequencing data. We additionally called structural variation using GATK-SV for 7,035 SCZ, 2,004 BD and 20,854 controls from whole-genome sequencing (WGS). The strongest BD deletion association was BPNT2 (OR: 9.04, P= 2.67 x 10^{-09}), which is a direct target of the mood stabilizer lithium. For SCZ, the strongest association was NRXN1, a gene with a high CNV burden across neuropsychiatric disorders. We further identified a triplosensitive effect at the 15q11.2 reciprocal genomic disorder locus where 15q11.2 duplication was associated with risk to BD but the deletion is depleted in cases compared to control. We also found that the effect sizes for well established loci such as 22q11.2 are stable across diverse ancestries. Next, we assessed the noncoding regions from WGS and found a significant enrichment of deletions and duplications in highly constrained noncoding regions (OR_{DEL}=1.96, P < 1x10^{-33},OR_{DUP} = 2.02, P<1x10^{-39}). In contrast, SNVs had no significant enrichment in noncoding-constrained regions due to low power, highlighting the potential for SVs to prioritize noncoding regions. We additionally found enrichment of SVs disrupting enhancers and TADs that associate with high-confidence SCZ/BD genes. We further identify several 5'UTR and intronic deletions in high confidence SCZ and BD genes, demonstrating the potential of non-coding SVs for improving gene association signals. Finally, we developed an SNV-SV imputation panel to impute common SVs from BD and SCZ GWAS loci. We found that approximately 1-2% of GWAS associations may potentially be explained by SVs

that are in strong linkage disequilibrium with eQTLs from PsychENCODE and GTEx. Overall, we demonstrate that SVs can identify novel risk genes and loci, begin to unravel important non-coding regions associated with disease, and explain unresolved common variation associations in BD/SCZ.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5170T: Trans-ancestry Meta-analysis of HIV Acquisition Reveals Novel African American-specific Risk Loci

Authors:

K. Liao[1], **N. Gaddis**[2], D. M. Rosenblum[3], J. Marks[2], B. Tamraz[4], M-J. Milloy[5], K. Hayashi[6], K. Debeck[6], K. Xu[7], J. Goedert[3], D. Jacobson[8], B. Aouizerat[9], S. H. Weiss[3], E. Johnson[10]; [1]RTI Intl., Research Triangle Park, NC, [2]RTI Intl., Durham, NC, [3]Rutgers New Jersey Med. Sch., Newark, NJ, [4]UCSF, San Francisco, CA, [5]Univ. of British Columbia, Vancouver, BC, Canada, [6]British Columbia Ctr. on Substance Use, Vancouver, BC, Canada, [7]Yale Sch. Med., New Haven, CT, [8]Oak Ridge Natl. Lab., Oak Ridge, TN, [9]Univ California San Francisco, San Francisco, CA, [10]RTI Intl., Duluth, MN

Abstract:

HIV and AIDS are major public health concerns, disproportionally affecting African, African American, and Hispanic populations. While risk factors for HIV acquisition are multifaceted, genetic studies have suggested a heritable component of between 28-42%. To date, only signals in the *CCR5* and *HLA-B* genes have persisted, with most studies limited to individuals of European ancestry. A recent GWAS of HIV viral load in individuals of African ancestry identified an ancestry-specific signal for *CHD1L*, highlighting the value of diverse and stratified ancestry studies in understanding HIV etiology.

In this work, we combined six cohorts (Weiss, CFAR+COGA, ICGH, UHS, VPWIDS, WIHS) of people living with HIV and unaffected controls to form the largest ancestrally diverse cohort for HIV acquisition. Our study included individuals of European (9946 cases:14749 controls), African American (3676:5644), and Admixed American (630:667) ancestry, defined by genetic similarity to EUR, AFR, and AMR subsets of the 1000 Genomes Project (1kGP). We conducted ancestry-specific GWAS and the largest to-date (n=35,312) trans-ancestry GWAS meta-analysis of HIV acquisition. Significant risk loci were further analyzed using the FUMA GWAS interpretation platform.

The meta-analysis identified two significant signals: a replication in *HLA-B* (lead SNP rs41557415: Beta = -0.41, p=9.42e-9) and a novel locus on chr8:p24 (lead SNP rs76312508: Beta

= 0.82, p=1.45e-8). rs76312508 is present only in the four cohorts of African ancestry (MAFs between 1-2%}) and had consistent directions of effect across cohorts (Beta = {0.39, 0.46, 0.83, 1.03}). In gnomAD and 1kGP, this locus was confirmed to be present in African populations (MAF ~2%) and rare (MAF < 0.0001) in non-African populations. FUMA follow-up of the novel locus mapped variants to 56 genes (three by position and 53 by chromatin interaction). Two are distal protein coding genes (*FAM84B* and *MYC*), with *MYC* showing significant variation in gene expression across 44 tissues from the GTEx portal. *MYC*, a proto-oncogene, plays a role in regulating immune response and has been suggested to be important for HIV infection in activated T cells.
Our study is the largest trans-ancestry GWAS meta-analysis for HIV acquisition to date. We find the first robust GWAS signal outside of HLA-B, driven by SNPs observed in individuals of African ancestry. While follow-up on the mechanism of the chr8:p24 locus is needed, increased risk for carriers of the lead SNP may contribute to increased rates of HIV acquisition in individuals of African ancestry. This finding illustrates the need for larger stratified genetic studies of HIV in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5171T: Transcriptome prediction and BMI polygenic risk score performance varies across African genetic ancestries populations from four diverse environments

Authors:

I. Gregga[1], C. Choo-Kang[1], A. Luke[1], P. Bovet[2], E. V. Lambert[3], D. E. Rae[3], K. Bedu-Addo[4], J. Plange Rhule[4], T. Forrester[5], L. R. Dugas[1], H. E. Wheeler[6]; [1]Loyola Univ. Chicago, Chicago, IL, [2]Unisanté Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, [3]Univ. of Cape Town, Cape Town, South Africa, [4]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [5]Univ. of West Indies, Kingston, Jamaica, [6]Loyola Univ Chicago, Chicago, IL

Abstract:

Given genetic and environmental diversity, genetic prediction based on African-ancestries individuals in the US or UK may not transfer well to other African diaspora regions. The Modeling the Epidemiological Transition Study (METS) is a longitudinal study of cardiometabolic health in five African-ancestry populations. We performed genome-wide genotyping and imputation in 735 METS participants from the US, Jamaica, South Africa, and Ghana. Using principal component analysis, we determined that US and Jamaican participants are genetically more closely related to Ghanians than to South African participants, which was expected given

historical migration patterns. We aimed to explore how genetic models for the transcriptome and body mass index (BMI) built using US-based African-ancestries populations perform in these four diverse populations.
A subset of METS participants, 38 in Ghana and 38 in the US, also had whole blood RNA-Seq data. For these, we predicted the expression from genotypes of 6,727 genes using published African-ancestries transcriptome models, which were built in the US-based TOPMed MESA study. The mean Spearman correlation for predicted vs. observed expression was significantly greater in the US (0.162) than in Ghana (0.149, p=5.7E-4).
Next, we used the three largest populations in All of Us (European n=105,791, African n= 46,784, American-admixed n=34,717) to build cross-validated polygenic risk score (PRS) models for BMI with PRS-CSx, a Bayesian method that combines genetic effects across populations via a continuous shrinkage prior. We then tested our PRS models in 735 METS participants. Spearman correlations between African-ancestry-validated PRS and measured BMI were US=0.221; Jamaica=0.142; South Africa=0.152; Ghana=0.021. This performance pattern roughly follows the UN human development index ranking of these four sites, with the US performing the best and Ghana performing the worst.
PRS performance varies between groups across the African diaspora, suggesting the influence of environmental differences in genetic prediction of BMI. Transcriptome prediction performance also varies significantly between the US and Ghana. Overall this work supports the importance of global representation in genetics research for more accurate healthcare.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5172T: Transcriptome-wide Association Study Identifies Novel Risk LociAssociated with Age-Related Macular Degeneration

Authors:

M. Grunin[1,2], Y. Mustafa[2,3], M. Gorski[4], I. M. Heid[4], W. S. Bush[1,5], J. L. Haines[1,6,5], International Age-related Macular Degeneration Genomics Consortium; [1]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [2]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [3]System Biology and Bioinformatics, Dept. of Nutrition, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [4]Dept. of Genetic Epidemiology, Univ. of Regensburg, Regensburg, Germany, [5]System Biology and Bioinformatics, Dept. of Nutrition, Case Western Reserve Univ. Sch. of Med.,, Cleveland, OH, [6]Cleveland Inst. For Computational Biology, Case Western Reserve Univ., Cleveland, OH

Abstract:

Age-related macular degeneration (AMD) is a complex degenerative blindness disorder affecting 200 million individuals yearly, with a global cost of care of $343 billion. Previous GWAS implicated 34 main loci associated with AMD (International Age-related Macular Degeneration Genomics Consortium [IAMDGC]) but identifying effective therapeutic targets remains difficult. The innate immune system, particularly monocytes/macrophages, are suspected to play a crucial role in AMD pathogenesis, although underlying mechanisms remain poorly understood. We integrated monocyte eQTL and AMD GWAS data to identify genetic risk genes for AMD. We utilized IAMDGC genomic chip data for 16,108 cases/18,038 controls (15,616/16,723 EUR, 207/322 ASN, 50/357 AFR, and 235/636 OTH) and imputed to TOPMed v2 imputation panel, build 38. Principal components were calculated and ancestry-specific GWAS analyses were conducted using a REGENIE Firth test and meta-analyzed with METAL. 51,162,741 variants (MAF>0.1, INFO>0.8) were analyzed across ancestries. We combined genetic data from controls with European ancestry (n=432) with monocyte expression data in several states of stimulation (naive, LPS 2h. LPS 24h, and IFN 24h; Fairfax et al. 2014) to develop elastic-net regression models of genetically-regulated expression. After filtering for models with significant association to gene expression, we performed a transcriptome-wide association study (TWAS) of monocyte expression on AMD in each stimulation state using the MetaXcan software suite. A second complementary modeling strategy using multivariate adaptive shrinkage (MASH), which borrows information across treatments, was also implemented for further TWAS analysis.
16 TWAS significant associations after Bonferroni-correction were identified across the genome, 14 in known AMD loci and 2 novel loci. The top result in LPS stimulated monocytes was the gene PLEKHA1 (P=6.13E-107), and the top result in naive monocytes was the gene HTRA1 (P=6.90E-78), both in the known HTRA1/ARMS2 AMD-associated locus. BUB3 (P=7.24E-10) and NUDT21 (P=4.96E-02) were newly associated with AMD. BUB3 is associated with choroidal thickness in a European population. Our second MASH pipeline identified multiple significant novel loci across the genome.
These TWAS results represent genes strongly regulated by variants known to be associated with AMD, as well as novel loci. Our results support the role of genetic regulation of monocyte inflammatory processes in AMD, and highlight novel biology underpinning pathogenesis. Validation of these candidate genes using gene expression data may elucidate novel risk genes and therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5173T: Transcriptomic Effects of Mood Stabilizing Drugs on the Gene Regulatory Network in Cell Villages of Human iPSC-derived Cortical Neurons

Authors:

L. Lacbawan[1], S. D. Detera-wadleigh[2], N. Akula[1], F. J. McMahon[3]; [1]Natl. Inst. of Mental Hlth., NIH, Bethesda, MD, [2]NIMH, NIH, Bethesda, MD, [3]NIH/NIMH, Bethesda, MD

Abstract:

Lithium (LiCl) and valproic acid (VPA) are first-line treatments for bipolar disorder (BD). Extensive research has been done on the effects of these medications but data on the molecular mechanisms of action that mediate mood regulation remain elusive. Historically, pharmacological studies used animal models, established cell lines, or postmortem tissues. Our study used renewable human iPSC-derived neuronal cells and a “cell village” culture approach to determine the impact of mood stabilizers on gene expression and transcriptional regulation. iPSCs from four healthy, unrelated individuals were differentiated into neural progenitor cells and then pooled together into a “cell village,” and subsequently differentiated into cortical neurons. Neurons grown post differentiation up to six weeks were divided into three treatment groups: a) untreated, b) treated with 1 mM LiCl for two weeks, and c) treated with 1 mM VPA for 72 hours. Accessible chromatin was analyzed using ATAC-seq and analysis is ongoing. Transcriptomic analysis of untreated versus lithium treated neurons showed ~70 genes that were significantly differentially expressed (p adj<0.05). VPA-treatment revealed ~5900 genes that were significantly differentially expressed (p adj<0.05) in untreated versus VPA-treated neurons, and g:Profiler analysis of these genes showed enrichment in synapse organization, axon guidance, and nervous system development. VPA treatment also revealed downregulation of several genes previously implicated in neuropsychiatric disorders including *CACNA1C*, which lies within a well-known BD and schizophrenia (SCZ) GWAS locus and encodes a voltage-dependent calcium channel; *AKAP11,* a risk gene for BD and SCZ; and *SETD1A* where loss of function variants has been associated with SCZ. VPA treatment showed upregulation of several *HDAC* genes, which are involved in chromatin remodeling. This study shows the strong effect of VPA on gene regulation in living neurons that may be relevant to its mood stabilizing action. The “cell village” strategy applied to human iPSC-derived neurons provides a potentially effective model to investigate drug mechanisms in the central nervous system. Additionally, differentially expressed genes and enriched pathways are potential targets for future investigations into novel therapeutics.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5174T: Transcriptomic profiling analysis during tolerance induction in dendritic cells from women with Systemic Lupus Erythematosus

Authors:

A. Hernández-Ledesma[1], S. Salazar-Magaña[1], E. Coss-Navarrete[1], D. Ramírez-Espinosa[1], L. Tinajero-Nieto[2], E. Torres-Valdez[3], A. Peña-Ayala[2,4], G. Félix-Rodríguez[5], G. Frontana-Vázquez[6], F. Rosetti[7], S. Fernández-Valverde[8], M. Gutierrez-Arcelus[9], D. Alpízar-Rodríguez[10], A. Medina-Rivera[1]; [1]Laboratorio Internacional de Investigación sobre el Genoma Humano, Univ. Natl. Autónoma de México, Querétaro, Mexico, [2]Hosp. Gen. Regional No. 1, Inst. Mexicano del Seguro Social, Querétaro, Mexico, [3]Hosp. Gen. Regional No. 2, Inst. Mexicano del Seguro Social, El marqués, Mexico, [4]Inst. Natl. de Rehabilitación “Luis Guillermo Ibarra Ibarra”, Ciudad de México, Mexico, [5]Hosp. Star Médica, Querétaro, Mexico, [6]Hosp. Gen. Regional No. 2, Inst. Mexicano del Seguro Social, Querétaro, Mexico, [7]Departamento de Inmunología y Reumatología, Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Ciudad de México, Mexico, [8]Univ. of New South Wales, Sidney, Australia, [9]Brigham and Women's Hosp., Harvard Med. Sch.. Broad Inst.. Boston Children's Hosp., Harvard Med. Sch.., Boston, MA, [10]Res. Unit, Colegio Mexicano de Reumatología, Ciudad de México, Mexico

Abstract:

Tolerance mechanisms prevent the immune cells to recognise and react against self-antigens. Due to its anti-inflammatory properties, tolerogenic dendritic cells (tolDCs) are a promising resource towards tolerance induction in autoimmune diseases. The contribution of genetic factors to Systemic Lupus Erythematosus (SLE) has been widely reported, still, there is a need to deepen into the genetic variants and regulatory mechanisms involved, especially in populations underrepresented in research. To address this, we explored the transcriptomic profiles of dendritic cells from Mexican women with SLE during tolerance induction. Monocytes were isolated using Lymphoprep and a magnetic beads cell-enrichment kit. Monocytes from each volunteer underwent two differentiation processes: for monocyte-derived dendritic cells (moDCs) they were stimulated with GM-CSF and IL-4; whereas for tolDCs, GM-CSF, IL-4 and IL-10 were administered. On the 7th day, cells were split into two conditions: one was kept supplemented as previously described, and the other was additionally supplemented with imiquimod (IMQ) to evaluate maturation and activation. RNA from each cell type was obtained and sequenced. Transcriptomic profiles of 23 women with SLE and 10 controls were recovered, and differential gene expression analyses were performed. First, we explored the genes that were differentially expressed upon differentiation of monocytes into dendritic cell phenotypes.

Among them, we identified 443 genes that had different expression directions between SLE and controls. Notably, IL-36, a pro-inflammatory cytokine previously associated with SLE activity, was upregulated in SLE samples. We also identified 944 genes with differing expression patterns during differentiation of monocytes into either moDCs or tolDCs. Genes that were key for the tolDC phenotype were associated with lipid metabolism, response to stimuli, and inhibition of dendritic cell maturation and activation, such as IL-21. We then explored the differences between moDCs and tolDCs in SLE, we found a set of genes that enables clustering of these cell types. Several of these genes had been associated with SLE before, such as IRF7 and DNASE1L3. We also observed that disease activity and the use of corticoids contribute to transcriptomic profiles. Overall, our findings uncover the complex gene expression differences between people with SLE and controls, as well as among cell types during differentiation towards a tolerant phenotype. Analyses of ancestry, genetic variants and allele specific expression will be performed to deepen the regulatory mechanisms potentially involved in these differences.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5175T: Uncovering novel genetic associations for lung, liver, and renal-system fibrosis through multi-trait analysis.

Authors:

G. Parcesepe[1,2], E. Joof[1,2], R. J. Allen[1,2], J. K. Quint[3], I. D. Stewart[4], G. Aithal[5,6], C. Scotton[7], H. Longhurst[8], G. Lassi[9], R. G. Jenkins[4], L. V. Wain[1,2], DEMISTIFI consortium; [1]Dept. of Population Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom, [2]NIHR Leicester BioMed. Res. Ctr., Leicester, United Kingdom, [3]Sch. of Publ. Hlth., Imperial Coll. London, London, United Kingdom, [4]Margaret Turner Warwick Ctr. for Fibrosing Lung Disease, Natl. Heart and Lung Inst., Imperial Coll. London, London, United Kingdom, [5]Faculty of Med. & Hlth.Sci., Univ. of Nottingham, Nottingham, United Kingdom, [6]NIHR Nottingham BioMed. Res. Ctr., Nottingham, United Kingdom, [7]Med. Sch., Univ. of Exeter, Exeter, United Kingdom, [8]Dyskeratosis Congenita (DC) Action, UK, United Kingdom, [9]Translational Sci. and Experimental Med., Res. and Early development, Respiratory and Immunology, R&D BioPharmaceuticals, AstraZeneca, Cambridge BioMed. Campus, Cambridge, United Kingdom

Abstract:

Fibrosis, the excessive accumulation of scar tissue, is a common pathological process in various organs in the body. While it serves as a natural response to injury or inflammation, chronic or unresolved fibrosis can lead to organ dysfunction and failure. The underlying molecular mechanisms driving fibrosis share common pathways involving inflammation and dysregulated

tissue repair processes, and understanding these shared mechanisms and identifying organ-specific differences can facilitate the development of targeted therapies to mitigate fibrosis and preserve organ function.Genome-wide association studies (GWAS) were performed in UK Biobank separately for lung, liver and renal fibrosis where cases were defined as individuals with any fibrotic disease in that organ. These were then meta-analysed with publicly available GWAS results for the relative single fibrotic diseases from non-UK Biobank cohorts. Multi-Trait Analysis of GWAS (MTAG) was conducted on the three traits summary statistics, to increase the power to detect signals. Novel signals were defined as those associated with a given trait at a significance of $P<5x10^{-8}$ in the MTAG, but not significant at this threshold in the GWAS. Novel signals were mapped to a gene using; distance from gene, eQTL, splice QTL and functional annotation to propose the most probable mapping of variants to genes. We explored whether these novel signals had been previously reported for individual fibrotic diseases utilising the GWAS catalog.Two novel signals for liver fibrosis and three novel signals for renal-system fibrosis were identified. The variants for three of the novel signals, near *PLEKHA5* (liver)*, DHX36* (renal-system)*,* and *MYO19* (renal-system) had not previously been associated with any fibrotic conditions. The variants near *APOE* (liver) and *SH2B3* (renal-system) had previously been associated with non-alcoholic fatty liver disease and inflammatory bowel disease, respectively. By analysing fibrotic disease at the organ level and employing a methodology that leverages the power of correlated traits, we identified new signals for renal-system and liver fibrosis. These included signals previously reported for specific fibrotic diseases of other organs, as well as novel signals not previously reported for any disease.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5176T: Understanding the impact of genetic variants on Alzheimer's disease with mass spectrometry proteogenomics

Authors:

H. Guturu[1], B. Lacar[1], A. Alavi[1], S. Ferdosi[1], A. Stukalov[1], G. Venkataraman[1], M. de Geus[2], H. Dodge[2], C-Y. Wu[2], P. Kivisakk[2], S. Das[2], B. Hyman[2], S. Batzoglou[1], S. Arnold[2], A. Siddiqui[1]; [1]Seer, Inc., Redwood City, CA, [2]Massachusetts Gen. Hosp. (MGH), Charlestown, MA

Abstract:

Proteogenomics is an emerging area of research that may provide novel insights into Alzheimer's pathophysiology and therapeutic targets. While protein quantitative trait loci (pQTL) analysis has recently been performed with cerebrospinal fluid and brain samples, identification of pQTLs in easily accessible sample types like plasma could enable access to larger scale studies

and provide insight to systemic factors influencing Alzheimer's disease (AD) progression. Here, we identify pQTLs in an AD cohort using mass spectroscopy (MS)-based proteomics from plasma samples. With this study, we aim to examine the impact of genetic variation on protein abundance and reveal the potential functional implications in AD. Blood samples were collected from 1,005 participants in the Massachusetts Alzheimer's Disease Research Center. Participant metadata included demographics, cognitive scores, clinical diagnoses and the plasma biomarker pTau-217 which has been shown to have strong diagnostic accuracy. Proteomics data were obtained by processing plasma samples with Proteograph™ XT (Seer, Inc.) before analysis on an Orbitrap™ Exploris™ 480 MS. Participant genotypes were obtained with low pass sequencing by Gencove. MS data were searched by DIA-NN. The low pass sequencing data was mapped and imputed using GLIMPSE. We identified pQTLs and differential protein abundances based on AD status and rate of progression. We then applied Mendelian randomization methods to identify putative causal protein markers of AD and cognitive decline. To reveal candidate proteins associated with AD, we performed differential abundance analysis and identified 138 protein groups that were different between healthy and AD groups. Additionally, eight protein groups were associated with dementia progression using Cox regression analysis. For pQTL identification, we performed a preliminary study with a pilot subset of the samples (N = 161) with array genotypes. We performed pQTL analysis using peptide abundances for each nanoparticle as well as protein abundance inferred by MaxLFQ rollup due to increased sensitivity of peptide level measurements. With peptide level quantification, we identified 55 Bonferroni-corrected pQTLs (16 cis and 39 trans) while protein quantification revealed 10 Bonferroni-corrected pQTLs (5 cis and 5 trans). We will evaluate the reproducibility of the results from both quantification levels in the full data for defining our final set of pQTLs generated across all 1,005 participants. Low pass sequencing and deep plasma MS proteomics data enable identification of cohort specific biomarkers that may be invisible to more targeted approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5177T: Unraveling the human brain cell-cell interactome: Shared mechanisms across prevalent neuropsychiatric and neurodegenerative diseases

Authors:

C. Porras[1], L. Sheu[1], P. B. Chandrashekar[2], R. Kosoy[3], M. Pjanic[1], J. Bendl[4], P. N.M.[1], K. Therrien[1], D. Mathur[1], S. Kleopoulos[1], Z. Shao[1], M. Alvia[1], C. Casey[1], A. Hong[1], S. Argyriou[1], D. Bennett[5], G. Voloudakis[1], V. Haroutunian[1], G. Hoffman[1], J. Fullard[1], K. Girdhar[6], D. Wang[7], D. Lee[1], P.

Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Univ. of Wisconsin-Madison, Madison, WI, [3]Mount Sinai Sch. of Med., Bronx, NY, [4]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [5]Rush Univ. Med. Ctr., Chicago, IL, [6]Mount Sinai, New York, NY, [7]Univ. of Wisconsin - Madison, Madison, WI

Abstract:

Cell-cell Interactions (CCIs) mediate transcriptomic dysregulation by responding to and altering the microenvironment thereby contributing to the progression of multiple neuropsychiatric (NPD) and neurodegenerative (NDD) diseases, including Alzheimer's disease (AD), schizophrenia, and bipolar disorder. Recent studies suggest a significant portion of cellular and molecular mechanisms are shared across psychiatric and neurological disorders, despite distinct phenotypic features. Firstly, we focused on generating reproducible CCI networks from a large-scale human brain disease atlas. We inferred over 38 million CCIs between 27 neuronal, glial, and immune cell-types using the PsychAD single-nucleus RNA-seq cohort. The dataset covers transcriptomic profiles of over 6 million individual nuclei in the prefrontal cortex collected from 1,494 donors affected by various NPDs and NDDs. We validate these CCIs by ensuring inferred ligand-receptor pairs are colocalized with Xenium and Visium spatial transcriptomic assays. Secondly, we prioritize 3,063 differentially regulated CCIs mediating eight NPDs and NDDs after accounting for biological and technical variations with linear mixed-effects regression models. We show that during the pathogenesis and progression of NPDs and NDDs, a significant portion of genetic and epigenetic subprocesses are shared and driven by CCIs. We show that about 9% of disease-associated CCIs are shared between NDDs and NPDs, with predominantly immune-glial CCIs differentially dysregulated in NDDs. In contrast, neuronal-glial CCIs are predominantly differentially dysregulated in NPDs. Next, we focus on identifying CCIs mediating AD progression using 3 independent measures of neuropathology and severity of clinical dementia. We show that brain immune cell receptor-mediated CCI genes are significantly enriched for AD GWAS variants and drive disease progression more than ligand-mediated CCIs. Finally, we combine the identified CCIs and whole-genome sequencing into an interpretable deep-learning model to enable personalized NPD and NDD risk prediction from individual genotypes. Our results move beyond independent single-gene analyses, incorporating ligand-receptor gene-pair and cell type-specific interactions to elucidate the higher-order molecular mechanisms underlying NPDs and NDDs, identifying targets for pharmacological intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5178T: Unravelling the genetic architecture of inflammatory bowel disease multiplex families with rare and common variant polygenic risk scores

Authors:

D. Jans[1], J. Depovere[1], H-S. Lee[2], Y. Abakkouy[1], S. Becelaere[1], J. Sabino[1,3], S. Vermeire[1,3], I. Cleynen[1]; [1]KU Leuven, Leuven, Belgium, [2]Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [3]Univ. Hosp. Leuven, Leuven, Belgium

Abstract:

Crohn's disease (CD) and ulcerative colitis (UC), the two main forms of inflammatory bowel disease (IBD), result from an inappropriate response of the mucosal immune system against the gut microbiome and leads to chronic inflammation in the intestine. The genetic architecture of IBD is diverse: a minority of patients have a monogenic form, while the majority have a polygenic susceptibility. In some families, a striking number of relatives are affected by IBD, with currently unknown cause(s). Here, we aim to determine the role of known common and rare IBD risk variants in multiplex families.

We calculated polygenic risk scores (PRS) for relatives of 54 CD and CD/UC families (133 CD, 13 UC, 101 unaffected first-degree relatives) with at least three affected first-degree relatives. A sporadic dataset of 1,200 CD and 598 unrelated controls was used for comparison. A common variant PRS (cvPRS) was calculated based on the CD effect sizes from de Lange et al (2017), with a p-value threshold ≤ 1e-5 and MAF ≥ 0.01. A rare variant PRS (rvPRS) was calculated based on Sazonovs et al (2022) with a p-value threshold ≤ 0.05 and a MAF ≤ 0.01, hereby avoiding overlap with the cvPRS.

As a group, affected relatives had a higher cvPRS than unaffected relatives (p = 4.71e-6, beta(se) = 0.68(0.15)). In contrast, the rvPRS was similar between affected and unaffected relatives. However, there was a large heterogeneity between families for both cvPRS and rvPRS. Six (11%) families had a mean cvPRS lower than that of unrelated controls ('low cvPRS'), and two (4%) families had a lower rvPRS than that of controls ('low rvPRS'). On the other hand, 23 (43%)and 14 (26%) families had a cvPRS or rvPRS higher than the mean cvPRS or rvPRS of sporadic cases, respectively ('high cv/rvPRS'). Notably, no family had both a low cvPRS and low rvPRS, and only four families had a high PRS for both. Interestingly, three families with a low cvPRS had a high rvPRS. Furthermore, no correlation was found between cvPRS and rvPRS on an individual level, or by mean PRS per family. Finally, when dividing the families in low, medium or high cvPRS groups, the rvPRS did not differ between these groups.

Our results indicate that in multiplex families common and rare variant burden are independent, and highlights the heterogeneity in genetic architecture between families.

Although families might have other (genetic) risk variants than sporadic cases which were not tested here, many (61%) families have an elevated genetic risk either through a high common or rare variant burden. This has important implications for further research and the clinic. For example, families with low cvPRS and rvPRS should be prioritized to be sequenced for pathogenic variants.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5179T: Unveiling Gene Regulatory Programs Linked to Autism Spectrum Disorder Through Multimodal Sequencing of Human iPSC-Derived Neurons

Authors:

J. Liu, W. DeGroat, P. Matteson, J. Millonig, A. Kreimer; Rutgers Univ., Piscataway, NJ

Abstract:

Autism Spectrum Disorder (ASD) is a group of pervasive neurodevelopmental disorders characterized by impaired social and communication skills and repetitive and restrictive behavior. Accumulating evidence suggests that ASD-associated variants are enriched in noncoding regions of the genome, many of them in gene-distal enhancers that mediate their effects by changing the expression of target genes.
In this study, we analyzed multi-omic datasets: RNA-seq (transcriptomic data), ATAC-seq (open chromatin), and ChIP-seq for H3K27ac (enhancer mark) from three disease-relevant cell types: human induced pluripotent stem cells (iPSCs), iPSC-derived neural progenitor cells (NPCs), and iPSC-derived neurons (iNs) sampled from healthy individuals and healthy siblings of ASD patients. Our results show different transcriptomic and chromosomal accessibility profiles among the different cell types, highlighting cell-type-specific regulatory elements. RNA-seq data analysis reveals different gene expression patterns between healthy individuals and ASD siblings within the same cell types, suggesting the regulatory effect of ASD genetic background.
Gene set enrichment analysis finds the glycosylation pathway, known to be altered in various neurological disorders, to be enriched in NPCs of ASD siblings compared to healthy controls, and neurodevelopmental-associated pathways exhibit less significance in iPSCs of ASD siblings compared to healthy controls. Finally, a reconstruction of cell-type specific enhancer-promoter interaction networks combined with their sub-structure analysis and ASD variant mapping provides a comprehensive framework for studying ASD associate variants in the context of gene regulatory networks. In summary, the results of this work enhance our understanding of non-

coding genetic variants' effects and suggest hypotheses regarding their mechanism of transcriptional network disruption that may lead to disease phenotype.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5180T: Validation of Multi-ancestry Polygenic Scores for Lipid Levels in 3,119 Participants from Samoa and American Samoa

Authors:

T-A. Yapp[1], M. Krishnan[2], S. Liu[1], S. Manna[3], H. Cheng[4], T. Naseri[5,6], M. Reupena[7], S. Viali[8], J. Tuitele[9], R. Deka[4], N. Hawley[10], S. T. McGarvey[11,12,6], D. Weeks[1,13], R. L. Minster[1], J. Carlson[1,13]; [1]Dept. of Human Genetics, Univ. of Pittsburgh, Pittsburgh, PA, [2]Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univ. of Pittsburgh Sch. of Dental Med., Pittsburgh, PA, [4]Dept. of Environmental Hlth., Coll. of Med., Univ. of Cincinnati, Cincinnati, OH, [5]Naseri & Associates Publ. Hlth.Consultancy Firm and Family Hlth.Clinic, Apia, Samoa, [6]Intl. Hlth.Inst., Sch. of Publ. Hlth., Brown Univ., Providence, RI, [7]Lutia i Puava ‘ae Mapu i Fagalele, Apia, Samoa, [8]Sch. of Med., Natl. Univ. of Samoa, Apia, Samoa, [9]Dept. of Publ. Hlth., Government of American Samoa, Pago Pago, American Samoa, [10]Dept. of Chronic Disease Epidemiology, Yale Sch. of Publ. Hlth., New Haven, CT, [11]Dept. of Epidemiology, Sch. of Publ. Hlth., Brown Univ., Providence, RI, [12]Dept. of Anthropology, Brown Univ., Providence, RI, [13]Dept. of Biostatistics, Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Introduction: Dyslipidemia is a significant risk factor for cardiovascular disease and a major contributor to the global burden of disease. Cardiovascular disease is the leading cause of death in Samoa, accounting for 34% of deaths. Exploring the genetic determinants of blood lipid traits in Polynesian individuals could help address research inequities and has the potential to provide additional insight about the biological foundations of such traits in other populations. A 2021 study by Graham et al. conducted a multi-ancestry genome-wide genetic discovery meta-analysis of lipid levels and generated polygenic scores (PGS) from ~1.65 million individuals from East Asian, admixed African American, Hispanic, and South Asian populations. In the current study, we applied these PGS for four traits—LDL-C, HDL-C, triglycerides (TG), and total cholesterol (TC)—in a sample of 3,119 participants across three time-separated cohorts recruited from Samoa and American Samoa in 1990-1991, 2002-2003, and 2010.
Methods: We calculated PGS in the Samoan cohorts using the variants and weights derived from the 2021 study. First, the PGS variants were lifted over to hg38 and then harmonized with genome-wide imputed variants in the Samoan cohorts. We calculated individual-level PGS by

summing the products of genotype dosage and variant weights for each PRS variant. We assessed performance of the PGSs in each cohort with partial r^2 and bootstrapped confidence intervals from linear regression models for each trait adjusting for age and sex.
Results: The PGS for LDL-C had r^2 = ~8% across the 3 Samoan cohorts, which was marginally lower in performance to East Asian and South Asian populations (r^2 = ~8-10%) and much lower than African American and Hispanic populations (r^2 = ~10-16%) in the 2021 study. The PGS for HDL-C had r^2 = ~10% for the discovery and 2002 cohort but had r^2 = ~5% in the 1990 cohort. TC had a PGS r^2 = ~10% across the three cohorts. The PGS for TG had r^2 = ~5-7% across the three cohorts.
Discussion: Our findings suggest that PGS derived from multi-ethnic ancestry populations have reduced predictive power when applied to the Samoan population. Additionally, the differences in r^2 values between traits could provide evidence that certain traits are more influenced by the environment than others. This highlights the need to build PGS in a variety of environmental contexts and ancestries to improve the accuracy and transferability of genetic risk prediction across diverse populations. Further research is needed to refine and validate PGS in the Samoan population and to assess their potential clinical utility in risk stratification and targeted interventions for blood lipid traits.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5181T: Variability in predictive and stratification performance of epilepsy polygenic risk scores in Taiwanese population biobank and clinical cohorts

Authors:

Y-S. Ke[1], L-H. Wang[1], C-R. Ho[2], M-H. Tsai[2], Y-C. Feng[1]; [1]Natl. Taiwan Univ., Taipei, Taiwan, [2]Kaohsiung Chang Gung Mem. Hosp., Kaohsiung, Taiwan

Abstract:

Epilepsy affects 1% of the global population and presents a variety of types and electroclinical syndromes, rendering diagnosis a major challenge. With >20 loci identified from GWAS for common epilepsies, polygenic risk scores (PRS) have merged as a potential diagnostic aid. While European (EUR) studies have demonstrated polygenic enrichment in epilepsy, evaluations in Asian populations remain limited. Here, we explored the predictive and stratification performance of epilepsy PRS in the Taiwanese population using data from a longitudinal population biobank and a curated clinical cohort. Leveraging the Taiwan Biobank (TWB) linked to the 20-year National Health Insurance Research Database (NHIRD), we analyzed data from

104,680 individuals, identifying 1,331 epilepsy cases via ICD codes and 722 via ICD plus antiseizure medications. Additionally, we examined a clinical epilepsy cohort of 886 people (572 cases) with deep, curated phenotypes from the Epi25 consortium Taiwanese cohort. PRS of all-epilepsy (EPI), generalized (GGE) and focal epilepsy (FE) were computed using the latest ILAE GWAS results (92% EUR, 3% African, 5% Asian). Predictive performance was evaluated through time-to-event and lifetime risk models. Findings in the TWB-NHIRD cohort indicated a moderate effect of EPI and FE PRS, with a 6-14% elevated risk per unit increase in PRS. The top 10% PRS was 1.27 times more likely to develop epilepsy than the lower 90% (p = 0.004). No significant PRS effect was observed for GGE despite high heritability. Notably, FE PRS was enriched in patients with initial unclassified seizures (HR = 1.10, p = 0.03), but no difference was seen between those progressing to a specific epilepsy diagnosis and those who did not. Effect sizes were similar between the two-tier definitions and between EUR and multi-ancestry PRS. In contrast, the Epi25-TWN cohort revealed a much stronger polygenic burden, particularly from multi-ancestry PRS (OR = 2.47, p = 6e-23; AUC ~0.75) than EUR PRS (OR = 1.16, p = 0.046; AUC ~ 0.54). Per-unit FE PRS predicted a 3.18-fold higher FE risk (p = 2e-29), and EPI PRS increased the risk of GGE by 2.08-fold (p = 2e-4). Individuals in the top 10% PRS were linked to a 6.87 times greater risk of epilepsy (p = 7e-7). Together, these results suggest nontrivial variability in PRS utility for target cohorts with varied phenotyping procedures and highlight challenges in defining heterogeneous disorders such as epilepsy using insurance claims data, which cannot accurately ascertain clinical symptoms. Our Epi25-TWN findings also underscore the importance of larger, diverse epilepsy GWAS to further enhance the portability and applicability of PRS for precision epilepsy.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5182T: Variants in *CHCHD2* on European and Amerindian ancestral background in Hispanic PD patients

Authors:

K. Nuytemans, L. Infante, E. Gu, A. Martinez, C. Kosanovic, C. Luca, J. Margolesky, H. Moore, C. Singer, A. Griswold, I. Haq, J. Vance; Univ. of Miami, Miami, FL

Abstract:

Background: Several reports in Japanese and Chinese families have linked CHCHD2 to Parkinson's Disease (PD), with additional few reports in Western European and Caucasian populations. Little to no information is available whether variants in this gene are seen in Hispanic PD patients. To reduce this gap in knowledge, we screened a cohort of Hispanic PD

patients, enriched for patients from the Caribbean. Method: Our current cohort consists of 150 Hispanic PD patients ascertained through collaboration with the Movement Disorder Clinic at the University of Miami Hospital and outreach event throughout South Florida. We performed Sanger sequencing for all four exons of the CHCHD2 gene. Local ancestry surrounding variants of interest was determined using genotyping data (Illumina's Global Diversity Array with Neurobooster).Results: We identified missense variant p.Pro34Leu, reported with uncertain significance) in one patient from Puerto Rico on European local ancestry (as previously reported). Additionally, we observed stoploss variant (p.Ter169LeufsTer3 = rs35957514) on Amerindian haplotype background in six individuals, confirming observed ancestral specificity in gnomAD data. Lastly, we identified common variants in 5'UTR (rs816407, rs10043) and 3'UTR (rs7406) at slightly lower frequencies than reported in gnomAD (72-75% versus 80% respectively).Conclusion: Genetic analyses of PD in underrepresented population groups are necessary as information from research in other continental populations is not always generalizable across race/ethnicity. Additionally, as admixed populations (such as Hispanic populations) have several ancestral contributions, it is important to assess frequency of variants on all relevant ancestries in these populations. The biomedical characterization of PD across race/ethnicity will help the understanding of disease mechanisms in all patients ultimately preventing further health disparities.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5183T: Whole exome sequencing analysis to identify rare variation in polycystic ovary syndrome (PCOS)

Authors:

C. Parker[1], R. Bauer[1], G. Stokes[1], C. Welt[2], R. Legro[3], M. Hayes[1], M. Urbanek[1]; [1]Northwestern Univ., Chicago, IL, [2]Univ. of Utah, Salt Lake City, UT, [3]MS Hershey Med. Ctr, Hershey, PA

Abstract:

Background: Polycystic ovary syndrome (PCOS) is a common multisystem endocrine disorder affecting 6-15% of women of reproductive age and is a major cause of anovulatory infertility. PCOS is a heterogeneous disorder with multiple sub-phenotypes modulated by both environmental and genetic factors. PCOS is diagnosed by the occurrence of hyperandrogenemia, oligo- or anovulatory, and/or polycystic ovaries. A study conducted in twins concluded that PCOS is highly heritable (h^2=0.72); additionally, first-degree relatives of individuals with PCOS have an increased risk of several sub-phenotypes of PCOS. **Objective:**Our goal is to identify the genes and pathways that directly impact the etiology of PCOS with the intent to further our

understanding of the underlying pathology associated with PCOS phenotypic heterogeneity and expedite the development of targeted treatment. **Methodology:**We performed whole exome sequencing (WES) on 718 phenotyped individuals with PCOS and 300 age-matched phenotyped controls. Our cohort comprises individuals of European and African ancestry from across America. We generated a comprehensive catalog of protein-coding variants identified in our cohort and will use *in silico* prediction tools to prioritize likely-to-be pathogenic rare (MAF<1%) and/or low-frequency (MAF<5%) variants. Additionally, we will perform gene-based burden testing using (1) our 300 phenotyped controls and (2) public database controls. Lastly, we will perform pathway analyses on genes with an enrichment of the prioritized variants. We expect the variants identified by WES to map to multiple genes and pathways including novel genes and pathways as well as genes and pathways previously predicted to play a major role in the etiology of PCOS. **Results:**We have completed quality control, annotation, and filtration of 718 PCOS cases and 300 controls and identified 160,431 rare/low-frequency protein-coding variants. Within these variants, we have replicated our lab's previous findings of variants in genes associated with insulin resistance, the gonadotropin signaling pathway, and the Anti-Müllerian Hormone (AMH) signaling pathway. **Conclusion:**In our WES dataset of 718 PCOS cases, we have replicated the identification of rare protein-altering genetic variants in pathways predicted to be impaired in individuals with PCOS supporting a critical role for genetics in the etiology of PCOS. We are confident that our gene-based burden testing and pathway analyses will allow us to continue replicating known genes and pathways and identify novel genes and pathways implicated in PCOS pathology, leading to improved personalized phenotyping and treatment of PCOS.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5184T: Whole genome sequencing analysis reveals rare variant contribution to coronary artery disease in the Japanese population

Authors:

H. Ieki[1,2,3,4], K. Ito[1], S. Zhang[5], S. Koyama[6,7], H. Yoshida[1,4], N. Enzan[1,6], R. Kurosawa[1], K. Miyazawa[1], K. Ozaki[8], Y. Onouchi[9], Y. Kamatani[10], C. Terao[1], M. Snyder[3,2], I. Komuro[4]; [1]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [2]Stanford Cardiovascular Inst., Stanford, CA, [3]Stanford Ctr. for Genomics and Personalized Med., Stanford, CA, [4]Dept. of Cardiovascular Med., Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [5]Univ. of Florida, Gainesville, FL, [6]Broad Inst., Cambridge, MA, [7]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [8]The Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [9]Chiba Univ, Chiba, Japan, [10]The Univ. of Tokyo, Minato-ku, Japan

Abstract:

Coronary artery disease (CAD) remains the leading cause of death worldwide. Recent genome-wide association studies (GWASs) have significantly enhanced our understanding of the genetic architecture of CAD. However, these methods primarily focus on relatively common variants detected by array genotyping and imputation. The contribution of rare variants has not been fully characterized. Therefore, investigating rare variants using genome sequencing is necessary to reveal the entire genetic architecture of CAD. Here, we sequenced the whole genome of a total of 1,952 Japanese early-onset myocardial infarction (MI) patients and 3,843 controls from the Biobank Japan (BBJ) and Nagahama Study. We then annotated each variant with ANNOVAR and compared the allele count of rare damaging variants between cases and controls. We observed a significant excess of putative loss-of-function (pLOF) variants in *LDLR*, a causal gene for familial hypercholesterolemia (allele count: 22 in cases and 1 in controls). Subsequently, we performed aggregated rare variant association analyses using SAIGE-GENE+, where just *LDLR* showed suggestive significance ($p = 2.3*10^{-5}$). Thus, to improve the power to detect rare variants, we developed a machine learning model to predict disease status from a cumulative effect of rare nonsynonymous variants and to develop a genetic risk score based on rare variants (Rare variant GRS). We compared the characteristics and the predictive performance of the GRS with the GWAS-based polygenic risk score (PRS). Both the rare variant GRS and PRS significantly discriminated CAD cases from controls (area under the curve [AUC] = 0.58; $p = 1.02*10^{-4}$ and 0.61; $p = 1.09 * 10^{-6}$, respectively) and integrative risk estimation using both scores further improved disease prediction (AUC = 0.66; $p = 3.2*10^{-14}$). We compared the relationship between clinical parameters and these genetic risk scores. As a result, rare variant GRS was significantly associated with LDL cholesterol level ($r = 0.21$; $p = 4.5*10^{-5}$), while PRS showed significant correlation with uric acid and triglycerides level ($r = 0.14$; $p = 1.2*10^{-3}$ and $r = 0.11$; $p = 8.2 * 10^{-5}$, respectively), suggesting that these risk scores have different characteristics. These results support the fact that rare variants also have distinctive features in CAD development.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5185T: Whole genome sequencing association testing in 520,000 individuals identifies rare variants and non-coding regulatory elements associated with HbA1c and Type 2 diabetes

Authors:

G. Hawkes[1], H. Wright[1], R. Beaumont[1], C. Armirola-Ricaurte[1], N. Owens[1], K. Chundru[1], A. Wood[1], K. Patel[1], A. Murray[1], I. jackson[1], T. Frayling[1], C. Wright[1], I. Barroso[1], A. Manning[2], J. Ferrer[3], M. Weedon[1]; [1]Univ. of Exeter, Exeter, United Kingdom, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Ctr. for Genomic Regulation, Barcelona, Spain

Abstract:

Glycated hemoglobin (HbA1c) is the standard biomarker used for diagnosing diabetes, for which large-scale GWAS have identified hundreds of common variants. Exome sequencing has also identified rare coding variants and genes that impact HbA1c, however the majority of common variants lie in non-coding regions of the genome, remaining unresolved by exome sequencing. Here, we use whole genome sequencing (WGS) to discover and replicate rare non-coding single variant and aggregates associated with HbA1c.

We performed WGS association testing for HbA1c in 441,647 individuals of genetically-inferred European, African and South-Asian ancestry in the UK Biobank. We tested single variants (minor-allele-count >= 5) and rare (minor-allele-frequency < 0.1%) non-coding aggregates in regions which showed evidence of conservation and/or constraint. We then replicated our results in 80,166 individuals of diverse ancestries from All of Us with measured HbA1c.

We identified 72 rare and low frequency (MAF<1%) variants associated with HbA1c, 17 of which were non-coding. This included a novel rare (MAF=0.2%) single variant associated with decreased HbA1c (beta =-0.99mmol/mol; P < 2.95-10), 1kbp upstream of SCRT1, which showed evidence of replication (beta = -1.10mmol/mol, P = 0.01). This cis-regulatory variant was also associated with decreased glucose levels (beta = -0.161mmol/mol, P = 7.42e-9) and reduced risk of type 2 diabetes (OR = 0.75, P = 6.28e-4). We show that the associated variant occurs within a region of open chromatin, and exhibits active enhancer histone modifications in human islets. The region is bound by islet transcription factor MAFB, and the variant disrupts the MAFB cognate binding motif. We subsequently identified a second independently associated variant in the same MAFB binding site that substantially reduced HbA1c levels (beta = -6.31mmol/mol; P = 7.04e-3). We also identified 16 rare non-coding aggregates independently associated with HbA1c. For example, we identified a non-coding region upstream of PIEZO1 (P = 9.74e-48), a gene where coding variation has been previously been associated with HbA1c, with replication in All of Us (P = 2.20e-3). This aggregate showed no evidence of association with glucose or T2D (P > 0.05), suggesting that it acts via a non-glycaemic pathway.

By leveraging population-scale WGS we identified a novel cis-regulatory locus important in beta-cell function. We further demonstrate that our rare-variant aggregation approach in the non-coding genome can identify new biology. Overall, our findings demonstrate the importance for rare non-coding variant discovery for understanding the genetic aetiology of diabetes.

Session Title: Complex Traits and Polygenic Disorders Poster Thursday Session

Board 5186T: Whole-genome identification of short tandem repeats underlying idiopathic peripheral neuropathy using short-read and long-read sequencing

Authors:

Z. Tang[1], D. E. Miller[2,3], S. Thomas[4], H. R. Reynolds[5], J. Milbrandt[1], A. Hoke[4], S. Jin[1]; [1]Washington Univ. Sch. of Med., Saint Louis, MO, [2]Univ. of Washington, Seattle, WA, [3]Seattle Children's Hosp., Seattle, WA, [4]Johns Hopkins Univ., Baltimore, MD, [5]Indiana Univ., Bloomington, IN

Abstract:

Peripheral neuropathies (PN) result from peripheral nerve damage and affect over 20 million individuals in the U.S. PN can be categorized into subtypes based on the underlying disease mechanisms. However, 25-30% of PN cases are idiopathic (IPN) with unknown causes, indicating there is a limited understanding of the mechanisms underlying this condition. This study explores complex genetic variation, particularly short tandem repeat (STR) expansions, to uncover novel pathogenic mechanisms in IPN. To investigate the incidence and impact of STRs in IPN we performed short-read whole-genome sequencing (srWGS) on 799 individuals with IPN from the Peripheral Neuropathy Patient Registry who tested negative for known PN-related genes.

Our STR-calling pipeline, employing ExpansionHunterDenovo and ExpansionHunter, was benchmarked by comparing repeat motifs and repeat lengths for a known pathogenic *RFC1* repeat expansion with those reported by the 1000 Genomes ONT Sequencing Consortium. Our pipeline successfully identified all five reported *RFC1* expansion carriers and a novel disease-associated ACAAG repeat motif in the same gene using srWGS data. The significant underestimation of repeat lengths in our results compared to the 1000 Genomes long-read data highlights the limitation of short-read sequencing in accurately estimating repeat lengths. Using our calibrated STR-calling pipeline, we analyzed srWGS data for 526 of 799 IPN patients and 887 age- and ethnicity-matched controls, revealing that 46 individuals (8.7%) carried an STR expansion in the *RFC1* gene, with 27 individuals (5.1%) heterozygous for the pathogenic AAGGG motif and 8 individuals (1.5%) homozygous for the same motif. The novel ACAAG motif was detected in 4 individuals (0.8%) in the IPN cohort and was absent in controls. We are actively working to complete analysis of the remaining 273 srWGS samples, perform case-control burden tests, and functional annotation to prioritize novel putative pathogenic STR expansions. Targeted long-read sequencing will be performed to validate the top candidate STR expansions, followed by genotype-phenotype correlation analysis.

Overall, our findings suggest that STR expansions, including but not limited to those in

the *RFC1* gene, may contribute to IPN. The data from this study will provide new insights into the underlying mechanisms of IPN, paving the way for the development of more accurate diagnostic testing and novel therapeutic options.

Session Title: Epigenetics Poster Thursday Session

Board 1001T: A flavonoid associated with slower epigenetic aging

Authors:

R. Saad[1,2], R. Costeira[1], P. Matias Garcia[3], C. Dias[2,4], C. Menni[1], M. Waldenberger[3], J. Bell[1]; [1]Dept. of Twin Res., King's Coll. London, London, United Kingdom, [2]Great Ormond Street Hosp. for Children NHS Fndn. Trust, London, United Kingdom, [3]Res. Unit Molecular Epidemiology, Inst. of Epidemiology, Helmholtz Zentrum München, German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [4]Dept. of Med. & Molecular Genetics, King's Coll. London, London, United Kingdom

Abstract:

Background

Dietary flavonoids and their metabolite byproducts are compounds that have been investigated for their roles in health and ageing.

3,7-dimethylxanthine, or Theobromine (TB), is a commonly consumed flavonoid derived from cocoa (*Theobroma cacao*) and produced as a byproduct of caffeine metabolism. TB has been linked to extended lifespan in *C. elegans* and affects genome-wide DNA methylation in mammals.

We aimed to study the association between TB, epigenetic markers of ageing, and genome-wide DNA methylation.

Methods

Six candidate xenobiotic flavonoids were measured in 509 blood samples from the TwinsUK cohort using mass spectrometry (Metabolon Inc.). We examined associations between the flavonoids and DNA methylation profiles, generated from the Infinium Human Methylation 450K BeadChip, accounting for biological and technical covariates.

We first investigated associations with epigenetic markers of biological ageing: GrimAgeAcceleration, DNAmTL (a surrogate marker of telomere length) and DunedinPACE. We then extended analysis to 430,768 probes genome-wide.

Targeted replication was pursued in 1160 samples from the population-based KORA cohort.

Results

Plasma TB levels were significantly associated with reduced epigenetic age acceleration, as

measured by GrimAge (p<2e-7) and DNAmTL (p<0.001).
Sensitivity analyses, including covariates of other cocoa and coffee metabolites, suggest that the effect is specific to TB. This association was most pronounced in smokers (p<2e-16).
GrimAge and DNAmTL associations with TB were replicated in the KORA cohort with the same direction of effect (p = 7.2e-08 and p = 0.007, respectively).
Epigenome-wide association analysis of TB revealed one methylation signal at FDR 5%: cg06126421 (6p21.33; β=0.135, p = 5.6e-8), previously associated with smoking, cardiovascular disease and all-cause mortality.
Two other top-ranked signals included cg25463779 (*FAM101A/RFLNA*); β=0.076, p = 3.9e-7) and cg04865110 (*JAKMIP1*; β=0.11, p =2.9e-6), previously associated with obesity and addiction, respectively.
Discussion
Our results indicate that TB has beneficial effects on health and ageing, as captured by both GrimAgeAcceleration and DNAmTL.
Epigenome-wide signals annotate to loci of biological importance; suggesting potential mechanisms for the effects of TB.

Session Title: Epigenetics Poster Thursday Session

Board 1002T: A Modified Atkins Diet is Feasible, Tolerable, and Shows Early Signs of Efficacy in Kabuki Syndrome Type 1

Authors:

L. Meier[1], A. Kalinousky[2], J. Johnson[1], J. Harris[1]; [1]Kennedy Krieger Inst., Baltimore, MD, [2]Johns Hopkins Univ., Baltimore, MD

Abstract:

Kabuki syndrome (KS) is a Mendelian disorder of the epigenetic machinery characterized by cognitive deficits, distinct facial features, multiple organ system anomalies, and growth deficiency. Individuals with KS have a distinct cognitive profile with deficiencies in visuospatial skills. The most common form of KS is KS type 1 (KS1) caused by pathogenic variants in KMT2D (75%), which encodes a methyltransferase that targets H3K4. Mouse models of KS1 have shown reversal of this cognitive phenotype with epigenetic therapies including histone deacetylase inhibitors (HDACi). One study showed KS1 mice placed on a ketogenic diet have improvement of hippocampal memory defects and b-hydroxybutyrate (BHB) which is produced when the body is in ketosis is known to have HDACi activity. Based on this, we hypothesized that a dietary therapy and elevated BHB levels may improve visuospatial cognition in individuals with KS1. 10 adults

with KS1 underwent a modified Atkins diet (MAD) for 12 weeks. Blood BHB levels were measured every 3 weeks. A KS-specific visuospatial cognitive battery consisting of 6 subtests was performed before and after diet.
Results: Individuals with KS1 showed excellent compliance with the diet with minimal side effects and 0 had to stop early. All individuals were above the minimum ketotic level BHB of 0.03 mmol/L on at least two measurements. 6/10 participants had at least one level greater than 1.0 mmol/L. 2/10 participants had at least one level greater than 2.0 mmol/L. One participant had a level greater than 2.0 mmol/L on all 3 BHB levels. Average BHB level at the 12 week point was 0.9 mmol/L (0.1-2.06). 4/6 subtests showed no significant difference pre to post diet. On the Beery visual motor integration test (VMI), when performing a paired t-test between the pre and post diet scores, there is a significant difference ($p = 0.0039$), with the post diet scores higher than the pre-diet scores. On the Benton judgement of line orientation (JLO), when performing a paired t-test between the pre and post diet scores, there is a trend towards improvement ($p = 0.0781$). Interestingly, the individual who consistently had the highest levels of BHB, displayed drastic score improvement on the JLO, increasing from 1 to 15. Adults with KS1 showed excellent compliance with MAD with only mild side effects and all individuals achieved some level of ketosis. While the trial was not designed for efficacy, cognitive results showing some improvement with diet are promising. Intriguingly, the only participant to have all BHB levels above 2.0 mmol/L showed drastic improvement on one subtest. Future studies should consider efficacy-focused trials.

Session Title: Epigenetics Poster Thursday Session

Board 1003T: A probabilistic model for RNA-binding QTL discovery

Authors:

D. Meyer[1,2], M. Schertzer[2], E. Nguyen[3], E. Mazzoni[3], J. Humphrey[4], D. Knowles[2,1]; [1]Columbia Univ., New York, NY, [2]New York Genome Ctr., New York, NY, [3]NYU Grossman Sch. of Med., New York, NY, [4]Mount Sinai Sch. of Med., New York, NY

Abstract:

Background: Genetic associations with RNA splicing (splicing QTLs) are enriched in GWAS risk alleles, implicating splicing in pathways of disease heritability. However, the underlying molecular mechanisms by which these variants impact splicing is unclear. One hypothesis is that they alter splicing factor binding, but no public data is presently available to support this hypothesis. Our work focuses on addressing this gap by discovering variants which alter the binding of RNA-binding proteins (including splicing factors), which we refer to as RNA-binding

QTLs (rbQTLs). **Methods**: We employ a combination of experimental and computational methods for discovering rbQTLs. Experimentally, we generated new formaldehyde RNA immunoprecipitation and sequencing (fRIP-seq) datasets to identify RNA sites transcriptome-wide bound by a specific RBP. This method has two specific advantages for rbQTL discovery over CLIP methods, including the capture of both direct and indirect RNA binding and generating larger RNA fragments that are more likely to contain a heterozygous SNP imperative for downstream allelic analysis. Computationally, we developed a probabilistic model that learns latent variables for the true proportion of alternate to reference alleles for heterozygous SNPs in both baseline RNA expression and in the fRIP-seq data given observations from sequencing in each condition. From these two learned variables, we compute the size and significance of the effects of both allele-specific expression and allele-specific binding for each variant with sufficient coverage. **Results**: To initially develop and test our rbQTL pipeline, we performed fRIP-seq on the neurodegenerative disease-associated RBP TDP-43 in iPSC-derived motor neurons. We compared our probabilistic model to two simpler approaches, a binomial test and a chi-square test. Importantly, we find that our model and the binomial test both detect a known ALS-associated splicing QTL in the *UNC13A* gene, rs12608932, as significant rbQTLs. We find that our model has fewer false negatives than the chi-squared model and fewer false positives than the binomial model. We are currently working to validate significant rbQTLs from these models by comparing predictions to binding peaks, known TDP-43 motifs, and known splicing QTLs. **Conclusions**: Our framework finds biologically meaningful rbQTLs, outperforming basic statistical tests. Going forward, we hope to generate new RNA-IP data in biologically relevant cell types across different RBPs and genetic backgrounds, and to apply our computational framework to find relevant rbQTLs in these contexts.

Session Title: Epigenetics Poster Thursday Session

Board 1004T: Advantages in modeling neuropsychiatric disorders in iPSC derived cell types to functionally evaluate genetic risk factors ⭐

Authors:

M. Madden[1], W. Du[1], H. Ramcharan[1], Y. Sami[1], X. Luna[1], K. Hazel[2], M. Gallagher[1], O. Corradin[3,1]; [1]Whitehead Inst., Cambridge, MA, [2]Whitehead Inst., Cambirdge, MA, [3]MIT, Cambridge, MA

Abstract:

The vast majority of genetic variants linked to disease are found in non-protein coding regions of the genome and impact the function of cell type-specific regulatory elements. This

complicates the translation of disease association to mechanistic insights into disease biology. Functional studies that investigate the impact of non-coding variants must be performed in the appropriate cellular context. This is particularly challenging for the study of neuropsychiatric traits as standard cell line models for neurons and glia are limited in capturing the specific neuronal and glial subtypes. We sought to investigate the extent to which specific neuron and glial populations derived from induced pluripotent stem cells (iPSCs) can be used to model genetic risk loci associated with neuropsychiatric disorders. We differentiated iPSC cell lines to medium spiny neurons, forebrain neurons, neural progenitor cells, microglia, microglia progenitors and astrocytes and performed ChIP-seq for H3K27ac. H3K27ac is a marker of active enhancers and promoters and can be utilized to evaluate cell type specificity in gene regulation. We next sought to compare these results to ChIP-seq of the corresponding in vivo tissues. We performed ChIP-seq on neurons sorted from the nucleus accumbens and prefrontal cortex and leveraged published datasets from microglia and astrocytes sorted from the brain. After standard quality control thresholding, we identified on average 64,138 ChIP-seq peaks in each iPSC or tissue sample. Globally, ChIP-seq data correlated robustly with the corresponding iPSC cell state. We found on average 47% of H3K27ac peaks active in the in vivo sorted cells to also be active in our iPSC-derived models. When compared to results from GWAS studies, we found iPSC-derived medium spiny neurons to explain a substantial proportion of genetic risk to substance use disorders and insomnia, whereas H3K27ac peaks in the forebrain neurons were enriched for schizophrenia and depression. This suggests that iPSC-derived models can be used to evaluate the impact of GWAS variants in specific neuron and glia subtypes. This approach demonstrates how functional studies in iPSCs within regions of comparable genomic landscape can be a guiding step in parsing genomic risk factors for neuropsychiatric diseases.

Session Title: Epigenetics Poster Thursday Session

Board 1005T: Association of BMI with Whole Blood DNA Methylation: An Epigenome-Wide Study of Bangladeshi Adults

Authors:

Z. Liao[1], J. Li[1], N. Jain[2], K. Demanelis[3], F. Jasmin[1], M. Kibriya[4], H. Ahsan[1], B. Pierce[1]; [1]Univ. of Chicago, Chicago, IL, [2]The Univ. of Chicago, Chicago, IL, [3]Univ. of Pittsburgh, Pittsburgh, PA, [4]Univ Chicago, Chicago, IL

Abstract:

Health conditions related to both high BMI and low BMI impact a sizable percentage of the world's population. Increased BMI is associated with an elevated risk of various chronic

diseases, including diabetes, cardiovascular diseases, and cancer; while low BMI may be a sign of undernourishment and is associated with anemia, osteoporosis, cardiovascular abnormalities, and immune function. One potential mechanism by which BMI influences health outcomes is through alterations in DNA methylation (DNAm).
We performed an epigenome-wide association study (EWAS) to examine the relationship between BMI and DNA methylation (DNAm) in a population with a high prevalence of underweight individuals. For 1,186 adults participating in the Health Effects of Arsenic Longitudinal Study in Bangladesh (17% with BMI < 17 kg/m2), DNA methylation (DNAm) was measured at ~850,000 CpGs in leukocyte DNA using the Illumina EPIC array. We evaluated the association between BMI and DNAm using linear models adjusting for covariates and surrogate variables representing cell type composition. After identifying BMI-associated CpGs (FDR 0.05), we evaluated their consistency with prior studies using the MRC-IEU EWAS catalog. Additionally, we conducted pathway analyses of BMI-associated CpGs using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Hallmark gene sets.
In total, we identified 93 CpGs associated with BMI using a Bonferroni significance threshold, 82 of which (88.2%) were novel (not previously reported in the EWAS Catalog). We identified 1,061 CpGs at an FDR of 0.05, 1012 of which (95.4%) were novel. BMI-associated CpGs were enriched in non-CpG islands and shores and depleted in islands. The top two CpGs annotated within the *CPT1A* gene (cg00574958, cg05325763). Eight BMI-associated CpGs annotated within the *IFITM1* gene. Among the BMI-associated CpGs (FDR<0.05), we observed significant enrichment of genes annotating to the estrogen response pathway, immune response, and cytokine-cytokine receptor interaction (p;<0.05).
This unique study of DNAm correlates of BMI in a low-BMI population provides evidence that exposure to BMI associates with DNAm at many CpGs in the epigenome. Epigenetic alterations of leukocyte DNAm could serve as potential mediators of BMI-related health outcomes and as biomarkers for exposure and effects in impacted populations.

Session Title: Epigenetics Poster Thursday Session

Board 1006T: Associations between depressive symptoms and DNA methylation among Black women in the InterGEN study

Authors:

B. Taylor[1], Y. Zhao[1], N. Perez[2], R. M. Masterson Creber[1], J. Y. Taylor[1], K. D. Cato[3]; [1]Columbia Univ. Sch. of Nursing, New York, NY, [2]New York Univ. Rory Meyers Coll. of Nursing, New York, NY, [3]Univ. of Pennsylvania Sch. of Nursing, Philadelphia, PA

Abstract:

Background: Depression is one of the most common mental health disorders and affects an estimated 20% of adults in the United States. Women are twice as likely to be diagnosed with depression compared to men. Black women are at a higher risk of depression due to socioeconomic factors. It can be difficult to diagnose depression due to varied presentation and unwillingness of sufferers to report their symptoms. DNA methylation (DNAm) has been associated with depression and has also been linked to psychosocial stressors. Machine learning (ML) is useful for analyzing complex data, such as DNAm data, and can have high predictive accuracy. The objective of this study was to examine associations between depressive symptoms and DNAm and compare ML methods for predicting depressive symptoms. **Methods**: This study, guided by the University of Illinois at Chicago Model for Genetic and Epigenetic Research, was a secondary analysis of data from the Intergenerational Impact of Genetic and Psychological Factors on Blood Pressure (InterGEN) study. This was a longitudinal study of Black mothers and a child aged 3-5 years. Interviews were conducted at four time points six months apart and included demographic characteristics and social determinants of health. DNAm data was obtained from saliva samples and processed with the Illumina Infinium Methylation EPIC (850K) BeadChip. Differential methylation analysis was performed at the site and region levels using the *minfi* R package, and ML models were implemented using the *caret* R package. **Results**: Among the 235 women with usable DNAm data and complete depressive symptom scores, approximately one-third reported depressive symptoms higher than normal. The dmpFinder function identified six differentially methylated positions (DMPs) with a q-value less than 0.05; all were located in noncoding regions. The bumphunter function identified 25 differentially methylated regions (DMRs). Twelve of these regions annotated to genes associated with neurological diseases or disorders; five regions annotated to genes involved in neurological pathways; two regions annotated to genes with a neurologically related function. Random forest had the highest predictive accuracy (0.74) followed by extreme gradient boost (0.65). In the top 25 DMPs identified by these models, six were selected by both models, indicating some consistency and agreement. There was no overlap between these DMPs and those identified by dmpFinder. **Conclusions**: It is possible that these findings did not replicate the findings of prior studies because prior studies were not performed on Black women. Future research should include more diverse samples of participants.

Session Title: Epigenetics Poster Thursday Session

Board 1007T: Autism spectrum disorder specific longitudinal patterns of DNA methylation from birth to age 5

Authors:

Z. He[1], J. A. Bidinger[1], A. Y. Song[1], J. L. Ames[2], L. A. Croen[2], J. N. Constantino[3,4], M. Fallin[5], C. Ladd-Acosta[1]; [1]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]Kaiser Permanente Northern California, Pleasanton, CA, [3]Emory Univ. Sch. of Med., Atlanta, GA, [4]Children's Hlth.care of Atlanta, Atlanta, GA, [5]Emory Univ. Rollins Sch. of Publ. Hlth., Atlanta, GA

Abstract:

Background: Autism spectrum disorder (ASD) is a DSM-V neurodevelopmental condition that begins *in utero* and can be reliably diagnosed in childhood. DNA methylation (DNAm) is linked to ASD and may serve as an early-life biomarker that can aid in screening, diagnosis, and intervention and/or inform etiology. However, prior studies were limited by post-diagnosis DNAm collection in childhood and did not consider DNAm changes over time. Given ASD's early life origins and dynamic DNAm changes during early life, characterizing ASD-specific DNAm changes from birth to childhood in longitudinal studies is warranted. **Objectives:** 1) Identify DNAm changes at birth associated with childhood ASD diagnosis. 2) Characterize DNAm trajectories from birth to age 5 and their association with ASD. **Methods:** Epigenome-wide DNAm profiling was performed using the Illumina EPIC array on dry blood spot samples collected at birth from 179 participants in the Study to Explore Early Development. 152 participants passed rigorous quality control, including 73 cases and 79 controls. Extant DNAm of peripheral blood samples collected at age 5 was available for a subset of participants (24 cases and 26 controls), measured in a prior batch using the Illumina 450K array. Batch effects were adjusted using an empirical Bayes method, leveraging 9 duplicates included in both batches. We used linear models to assess associations of DNAm at birth with childhood ASD diagnosis. We used linear models with generalized least squares to assess longitudinal associations of changes in DNAm from birth to age 5 with ASD, allowing for within-individual correlations of DNAm over time. Associations were adjusted for sex, maternal race, and estimated cell type compositions. **Results:** For DNAm at birth, we found 6 suggestive CpG sites associated with ASD status (M-value mean difference range -0.39 - 0.32, $P < 10^{-5}$), which were mapped to genes related to ASD, intellectual disability (ID), neurodegeneration and behavioral symptoms, including *DEAF1*, *ADRA1A*, *CACNA1C*, *CAPN2*, and *NDST1*. Longitudinal analysis revealed 3 CpG sites showing DNAm changes from birth to age 5 associated with ASD status (difference of M-value longitudinal changes range 0.27 - 0.50, $P < 10^{-5}$), which were mapped to genes related to ASD and ID, including *HIPK4* and *CCDC87*. For all 3 CpG sites, average DNAm levels were lower at birth and increased from birth to age 5 in cases, while they were higher at birth and remained stable or decreased from birth to age 5 in controls. **Conclusion:** Blood DNAm at birth and DNAm trajectories from birth to age 5 are associated with childhood ASD status. This suggests ASD-specific patterns of DNAm may exist in early postnatal life stages.

Session Title: Epigenetics Poster Thursday Session

Board 1008T: Clonal Hematopoiesis Paradox Resolved by Profiling Allele-Specific Methylation and Hydroxy-Methylation

Authors:

J. Van Amburg[1], F. Puddu[2], J. Scotcher[2], A. Bick[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Biomodal, Inc., Cambridge, United Kingdom, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Background: Clonal Hematopoiesis of Indeterminate Potential (CHIP) is a common, age related, blood disorder characterized by the acquisition of somatic mutations in blood stem cells resulting in a clonal outgrowth. ~70% of patients with CHIP have mutations in just two genes, *DNMT3A* and *TET2*. These genes have a discordant effect on DNA methylation. *DNMT3A* leads to the addition of methyl groups to cytosine (eg 5-methylcytosine, mC). *TET2* catalyzes the first step in removing methyl groups by converting 5-methylcytosine to 5-hydroxymethyl cytosine (5hmC). Here we seek to understand how mutations in genes with discordant effects on methylation can converge on the same pro-proliferative phenotype. Methods: We modeled CHIP in human primary hematopoietic stem cells by introducing loss of function (LoF) mutations in DNMT3A, TET2 and control targeting guides using CRIPSR-Cas9 (N=4 individuals) and expanded the edited cells in culture for 7 days. CRISPR editing efficiencies were ~60-80% for DNMT3A and TET2. We then leveraged the Biomodal duet evoC library preparation chemistry to obtain haplotype resolved, single base-resolution of DNA sequence, mC and 5hmC in a single workflow, enabling characterization of allele-specific methylation. Libraries were captured with the Twist Human Methylome targeted panel and then sequenced to >50x depth on the Illumina NovaSeq platform. Results: *DNMT3A LoF* induces a genome-wide hypomethylated state whereas *TET2 LoF* cells cause a genome-wide hypermethylated state. DNMT3A and TET2 LoF mutations resulted in largely non-overlapping differential methylation in cis-regulatory elements although both converged on genes implicated in proliferation and inflammation. For example, *DNMT3A led to* hypomethylation of CpG sites in cis-regulatory elements (CREs) close to both *JAK2* (0.579-fold reduction, q value %lt; 4.1x10^-6) and *RHOA* (0.364-fold reduction, q &lt: 2.3x10^-7). *TET2* LoF caused increased methylation of CREs close to IL1R1(1.57-fold increase, q < 3.1 x 10^-7) and *CREBBP* (1.14-fold increase, q %lt; 4.2 x 10^-7). In the 5hmC, there was no significant difference between DNMT3A LoF and control samples. However, *TET2 LoF resulted in* significant increases in 5hmC at an allele-specific level compared to matched controls. For example, both *RUNX3* (.28-fold increase, q < 1.7 E-8) and *HDAC4* (0.21-fold increase, q < 3.78x10^-3) had allele-specific increases in 5hmC abundance. Conclusion: DNMT3A and TET2 loss of function leads to differential

methylation and hydroxy-methylation of non-overlapping set of genes that converge on common pathways including HSC proliferation and inflammation.

Session Title: Epigenetics Poster Thursday Session

Board 1009T: Comparative DNA methylation profiling in PBMCs of cirrhosis with and without hepatocellular carcinoma

Authors:

H. Lim, C. I. Amos, J. Byun, H. El-Serag, A. P. Thrift; Baylor Coll. of Med., Houston, TX

Abstract:

Background: Hepatocellular carcinoma (HCC) is a leading cause of death worldwide. In the United States, HCC incidence and mortality have increased rapidly over the past two decades. Despite advancements in treatment, cure remains only possible for fewer than 10% of HCC patients. Most HCC patients are diagnosed with advanced disease and 5-year survival remains <20%. Cirrhosis is the precursor to more than 80% of HCCs diagnosed in the United States. Aberrant DNA methylation occurs in an early phase of carcinogenesis and can be detected in peripheral blood mononuclear cells (PBMCs), which are thought to detect the epigenetic reprogramming of the host immune system during the development of malignant tumors. However, the difference of DNA methylation patterns in PBMC between cirrhosis and HCC is still unclear. **Methods:** Twenty-two cirrhosis subjects who later developed HCC and gender/age-matched with 22 cirrhosis subjects who remained cancer-free during an average 4-year follow-up were retrospectively identified, and PBMC samples at baseline were retrieved. All the samples were analyzed by Infinium MethylationEPIC microarrays covering~850K CpG sites. We used the Champ package to identify differentially methylated CpGs sites (DMPs) and genes that represent specific biological pathways between cirrhosis patients with and without HCC. eFORGE was used to identify the significantly enriched CpGs for DNase I hypersensitive site, 15 chromatin states, and 5 histone marks in specific tissues and cells. **Results:** In total, 675 DMPs were identified by comparing cancer and cirrhosis, with 177 hypomethylated and 498 hypermethylated (p-value < 0.05 and | delta Beta value| > 0.05). Copy number alteration (CNA) regions with gains and loss were frequent in chromosomes 6 (Chr6) in HCC group. The significant CpGs were enriched for repressed polycomb or hetetochromatin states in liver tissues (FDR q <0.01). We characterized the DMP location across the chr6 footprint profile in liver. Two DMPs (cg00269725 and cg00658413) were covered by *PAX6* and *MZF1* motifs. After excluding 6 HCV active patients, we found 556 DMPs (Hypo:332, Hyper:224). Four chr 6 enriched DMPs (cg03197101, cg06053307, cg12078157, and cg21457401) were covered

by *E2F1, ZFP206,* and *SZF11* motifs. **Conclusions:** We identified the methylation landscape of cirrhosis and HCC and provided new methylation markers. Comprehensive epidemiologic and epigenetics studies are necessary to validate our findings in larger sample sizes and to understand the underlying mechanism leading to the transition from cirrhosis to HCC.

Session Title: Epigenetics Poster Thursday Session

Board 1010T: Comprehensive analysis of epigenome and transcriptome-based correlation network analysis in severe asthma risk prediction

Authors:

T. Mersha, E. Y. Dessie, Y. Gautam; Cincinnati Children S Hosp. Med. Ctr., Cincinnati, OH

Abstract:

Asthma is a chronic inflammatory disease of the airways associated with epigenetic and genetic risk factors. Co-expressed and co-methylated gene modules-based approaches can identify gene sets that are specifically asthma-associated biological process beyond known candidate genes. However, limited studies reported predictive models associated with asthma severity and lung function based on genome-wide weighted correlation network analysis (WGCNA) and machine learning methods. We performed epigenome and transcriptome analyses of asthmatic and control subjects (n =142) using WGCNA to identify significantly co-methylated and co-expressed modules associated with asthma severity and lung function. We used machine learning approaches to select differentially methylated and expressed genes followed by methylation risk score (MRS) and transcriptomic risk score (TRS) based risk predication. There were 18 methylated CpG sites and 28 differentially expressed genes significantly associated with asthma severity. The risk prediction model based on asthma status revealed MRS (AUC = 0.91), TRS (AUC = 0.97) and the jointed analysis (MRS +TRS) with AUC = 0.98. In addition, MRS and TRS were strongly associated with asthma severity and reduced absolute forced expiratory volume in one second (FEV1) ($P < 0.001$). An independent validation showed that TRS (AUC=0.86) was better than MRS (AUC= 0.66) based asthma prediction model. Moreover, integrated analysis of asthma-severity associated co-methylation and co-expression modules showed high overlap with enriched biological pathways including Th1 and Th2 cell differentiation and notch signaling. Overall, the finding suggests the utility of multi-omics based integrated model for asthma risk prediction and could serve as reliable method for asthma diagnosis.

Session Title: Epigenetics Poster Thursday Session

Board 1011T: DeepFace: Deep learning-based framework to contextualize orofacial cleft-relatedvariants during human embryonic craniofacial development

Authors:

Y. Dai[1], T. Itai[2], G. Pei[2], Z. Zhao[3]; [1]Univ. of Texas Hlth. Sci. center at Houston, Houston, TX, [2]UTHlth., Houston, TX, [3]Univ Texas HSC Houston, Houston, TX

Abstract:

Orofacial clefts (OFCs) are common congenital birth defects, yet the causal genetic variants remain largely unknown. Previous multiethnic studies have identified dozens of associated loci for both cleft lip with or without cleft palate (CL/P) and cleft palate alone (CP). Although several nearby genes have been highlighted, the 'casual' variants are largely unknown. To fill in this gap, we developed DeepFace, a convolutional neural network, model to evaluate the functional impact of OFC-associated genetic variants using SNP activity difference (SAD) scores during human craniofacial development. The DeepFace model is trained with 204 epigenomic assays from crucial human embryonic craniofacial developmental stages of post-conception week (pcw) 4 to pcw 10. DeepFace achieved a Pearson Correlation Coefficient ranging from 0.50 to 0.83 for 12 epigenetic features. The narrow peak features like H3K4me3 and H3K79me2 obtained the highest prediction accuracy, while the two broad repressive marks, H3K27me3 and H3K9me3, have the lowest medium performance. We systematically assess curated CP and CL/P risk variants from GWAS Catalog and latest GWAS summary statistics to form the OFC-associated SNPs sets. Similarly, we curated two control sets using two irrelevant trait variant collections Alzheimer's disease (AD) and schizophrenia (SCZ). DeepFace predicts the epigenomic 'activity' probability for each variant using a 1024-bp subsequence from a 131-kb sequence. We identified that OFC-associated SNPs exhibit higher SAD scores across various variant categories compared to less-related groups, indicating a context-specific impact during craniofacial development. Comparing control variants from unrelated traits, such as AD and SCZ, showed that OFC-significant variants consistently presented higher median absolute SAD scores, particularly in intronic, ncRNA_exonic, and upstream/downstream categories. Six SNPs were identified with significant linear relationships to SAD scores, suggesting temporal regulatory roles. Cell-type specificity analysis pinpointed trophoblast cells as having the highest enrichment of risk signals associated with OFCs, followed by endothelial and stromal cells. Overall, DeepFace harnesses distal regulatory signals from extensive epigenomic assays, offering new perspectives for prioritizing OFC variants using contextualized functional genomic features. This model has the potential to be extended to study other complex diseases or traits by providing

high-resolution predictions of the regulatory impacts of genetic variants, advancing our understanding of the molecular mechanisms underlying OFCs.

Session Title: Epigenetics Poster Thursday Session

Board 1012T: Differential DNA Methylation Enables Accurate Prediction for *C9orf72* Pathogenic Repeat Expansion

Authors:

N. Ramesh[1], L. M. Reus[2,1], R. Ophoff[1]; [1]Univ California Los Angeles, Los Angeles, CA, [2]Amsterdam UMC, Amsterdam, Netherlands

Abstract:

The hexanucleotide (G4C2) repeat expansion in the promoter region of C9orf72 is the most common genetic cause of Frontotemporal Dementia (FTD) and Amyotrophic Lateral Sclerosis (ALS). Previous studies have identified DNA hypermethylation in the C9orf72 promoter region among subjects with high repeat expansions. We performed a genome-wide DNA methylation (DNAm) analysis to identify differentially methylated probes associated with C9orf72 expansion status and applied these results to build a DNAm predictor to identify repeat expansion carriers. DNAm data was obtained from 27 individuals with a known pathogenic G4C2 repeat expansion (mean age = 62.2 years, SD 6.7) and 250 individuals without the expansion (mean age = 63.3 years, SD 8.4) from an FTD case/control cohort recruited in the Netherlands. DNAm data was collected using the Illumina EPICv2 array, which assayed the DNAm status of 937,055 CpG sites. DNAm values for each CpG probe were regressed against repeat expansion status corrected for age, sex, cellular fractions, smoking status, and experimental batch. To predict C9orf72 expansion status, LASSO regression was applied to 23 CpG probes at the C9orf72 locus. The training was performed on 70% of the data and accuracy was tested on the remaining 30%. We repeated the process 100 times to account for prediction variability.
We identified 9 CpG probes with genome-wide significant evidence of association with repeat expansion status, with 8 of these probes located near the repeat locus. The LASSO model (identifying nine informative CpGs) was highly predictive of repeat expansion status with an average accuracy of 98.6%. Restricting the model to include only CpGs present on the EPICv1 array did not change the accuracy. However, when limited to CpGs present on the Methyl450K DNAm array, accuracy diminished significantly (average Type I error rate of 63.4%). Applying the DNAm predictor to another EPICv2 data set with 2,600 subjects without known ALS/FTD cases identified four C9orf72 repeat expansion carriers (prevalence of 0.15%). The latter results are being validated by PCR and clinical assessment of subjects.

Our results revealed significant methylation differences within the C9orf72 locus associated with the hexanucleotide repeat expansion. We developed a highly accurate DNAm predictor as biomarker for identifying C9orf72 expansions in clinical settings and in large DNAm cohort studies. Our findings highlight the potential use of DNAm biomarkers for the detection of pathogenic repeat expansions and other genomic variants.

Session Title: Epigenetics Poster Thursday Session

Board 1013T: Distance-dependent synergy among neighboring genomic regulatory elements

Authors:

M. Maurano, R. Ordoñez, A. Ribeiro-dos-Santos, G. Ellis, H. J. Ashe, N. Berastegui, M. Majewski, B. Camellato, J. D. Boeke; NYU Inst. for Systems Genetics, New York, NY

Abstract:

Noncoding disease and trait-associated genetic variation frequently overlaps genomic regulatory elements such as DNaseI hypersensitive sites (DHSs). While regulatory elements are typically considered individually, they are embedded in a rich landscape of functional elements, and the intervening distances play a critical role in modulating their function and activity. We have developed a high-throughput multiplexed delivery pipeline for our recently developed Big-IN technology for scarless genome engineering, which we use to analyze regulatory function in a constant chromosomal context replacing the *Sox2* Locus Control Region (LCR). Applying this pipeline to dissect complex enhancer clusters, we have analyzed >100 engineered payloads in 520 mouse embryonic stem cell (mESC) clones. We characterize heterotypic DHS combinations from different key mESC regulator loci and uncover pervasive context dependencies whereby DHS activity depends on its partner DHS. We generate quantitative profiles delineating the functional impact of intervening distance between neighboring DHSs, and show that this synergy decays as a characteristic function of distance with an influence extending upwards of several kilobases. We generate similar distance profiles dissecting transcription factor position on the 10-500 bp scale. These spacing and position profiles will serve as key functional fingerprints for mammalian genomic architectures and provide critical context for the interpretation of the impact of disease- and trait-associated variants in non-coding regulatory regions.

Session Title: Epigenetics Poster Thursday Session

Board 1014T: DNA methylation and type 2 diabetes in West Africans: A Mendelian randomization analysis

Authors:

K. Meeks[1], E. L. van der Linden[2], A. R. Bentley[1], T. L. Assimes[3], N. Franceschini[4], P. Henneman[2], F. Chilunga, P[2], A. P. Doumatey[1], C. Agyemang[2], A. A. Adeyemo[1], C. Rotimi[1]; [1]NIH, Bethesda, MD, [2]Amsterdam UMC, Amsterdam, Netherlands, [3]Stanford Univ. Sch. of Med., Palo Alto, CA, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Epigenome-wide association studies (EWAS) have identified DNA methylation sites (CpGs) associated with type 2 diabetes (T2D) and related glycemic traits in several global populations. However, the cross-sectional nature of these studies does not allow for inferring causality of identified CpGs. In addition, epigenetics studies in West African populations are scarce despite the high and rising burden of T2D in these populations. Hence, we aimed to identify CpGs causally associated with T2D among West Africans. We used the Illumina EPIC DNA-methylation array to profile DNA methylation on DNA from white blood cells collected from 879 Ghanaians (RODAM study) and 333 Nigerians (AADM study) who were not on glucose-lowering medication. Meta-analyses of EWAS revealed 30 CpGs associated with T2D, 26 with HbA1c, and 3 with homeostatic model assessment estimates of insulin sensitivity (HOMA-S). These 59 CpGs were carried forward as exposures to two-sample Mendelian randomization (MR) analysis for causal inference with T2D as the outcome. We identified methylation QTLs (meQTLs) for 43 of these 59 CpGs using imputed genotyping data from the AADM study. LD clumping was performed to select independent meQTLs that were used as instrumental variables (SNP-exposure associations). Genome-wide association analyses for T2D were performed on a non-overlapping set of 4,120 Nigerians and Ghanaians from the AADM study to derive the SNP-outcome associations. Using the inverse variance weighted (IVW) MR method with both cis and trans meQTLs as instruments, 30 CpGs (70%) showed significant causal associations with T2D. We performed multiple sensitivity analyses and used MR Egger regression to explore potential pleiotropic bias. Out of the 30, we found that CpGs cg14688343 (*LAMTOR1*), cg10778880 (*FBN1*), cg26591384 (*FAM107B*), and cg02645285 (*DLG2*) were robustly causally associated with T2D as they were also associated using only cis-meQTLs as instrumental variables (P-values < 5×10-6) and showed nominally significant MR Egger estimates as well (P-values < 0.1). *FBN1* encodes a preprotein processed to asprosin, which has been shown to regulate glucose homeostasis. The FAM107B protein has been reported to be differentially increased in T2D compared with non-T2D individuals. Our study provides evidence for a causal association between DNA methylation and T2D in West African populations near genes relevant to glycemia. In addition to better understanding of disease mechanisms, these four CpGs with

evidence of causal associations could be prioritized as potential biomarkers for early detection of disease or as drug development targets.

Session Title: Epigenetics Poster Thursday Session

Board 1015T: Elucidating DNMT3A instability mutation in clonal hematopoiesis

Authors:

G. Ryu[1], Y. Koh[2], S-S. Yoon[1,2]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract:

In recent years, there has been growing interest in understanding the role and function of clonal hematopoiesis (CH). *DNMT3A*, which encodes an enzyme that carries out de novo DNA methylation, has emerged as the most commonly observed gene associated with CH. A mutation in *DNMT3A* could cause hematopoietic stem cell proliferation and differentiation disorders, leading to the development of various hematologic malignancies. Unlike the high prevalence of R882 missense mutations associated with acute myeloid leukemia, the set of *DNMT3A* mutations observed in CH is mostly non-R882 missense mutations. **Furthermore, Certain sets of non-R882 missense mutations lead to protein instability, impacting its function in ways distinct from other mutations.** This highlights the need for further investigation into the specific consequences of these mutations on *DNMT3A* function and their role in CH. Using TOPMed samples having both WGS and Illumina EPIC methylation array data, we compared beta values and calculated local-pairwise methylation discordance (LPMD) to measure local methylation randomness.

The *DNMT3A* R882 mutation is well-known for its function of hypomethylation, a unique characteristic that differentiates it from other mutations in *DNMT3A*. In contrast, *DNMT3A* instability inducing mutation (IIM) does not change overall methylation levels, similar with other *DNMT3A* mutations. **But interestingly, *DNMT3A* IIM differentiates itself from other *DNMT3A* mutations with regards to local methylation randomness.** We compared the cumulative mean value of LPMD with a pairwise Wilcoxon-test. *DNMT3A* IIM shows a significant difference in local methylation randomness compared to *DNMT3A* R882($p=0.006$) or other *DNMT3A* mutations($p=0.004$). **Furthermore, *DNMT3A* IIM induces local methylation randomness, not uniformly, but in specific regions.** Among the various chromosome cytobands, 9q21.33 is the most significant region whose local methylation randomness changed (adj $p<0.001$). This suggests that, in CH, unstable IIM-

impacted *DNMT3A* proteins operate in a specific region that have the potential to impact DNA methylation dynamics. **Additionally, differential methylation region analysis revealed that individuals with *DNMT3A* IIM exhibited significant enrichment of the NF-kB target gene set compared to healthy controls (adj p<0.001).** This suggests a potential role of *DNMT3A* IIM in modulating immune response and inflammation-related pathways. Altogether, our work provides important insights into CH by contrasting the impacts of instability inducing mutations and missense mutations in *DNMT3A*, with specific investigation of altered DNA methylation dynamics.

Session Title: Epigenetics Poster Thursday Session

Board 1016T: Epigenetic Age in Veterans with Gulf War Illness in the Gulf War Era Cohort and Biorepository (GWECB)

Authors:

X. Qin[1,2], S. H. Boyle[1], A. Thompson[1], T. Redding IV1[1], A. Stone[3], W. C. Carter[3], L. M. Griffin[3], Y. Shi[4], S. Pyarajan[4], K. Markianos[4], J. Vahey[1,2], M. G. Johnson[1], A. Press[1], J. Upchurch[1], S. Petry[1,2], K. Sims[1], E. J. Gifford[1,2], C. Williams[1], E. Hauser[1,2]; [1]Durham VA Hlth.Care System, Durham, NC, [2]Duke Univ., Durham, NC, [3]Central Arkansas Veterans Hlth.care System, Little Rock, AR, [4]VA Boston Hlth.Care System, Jamaica Plain, MA

Abstract:

Many Veterans who were deployed to the 1990-1991 Gulf War report a constellation of pervasive chronic symptoms such as alterations in mood and cognition, fatigue, and musculoskeletal problems known as Gulf War Illness (GWI). The etiology and effect of GWI on risk of dying from diseases are unknown. Our previous research used survey data and the VA Corporate Data Warehouse to show GWI was associated with an accelerated time for reaching a multimorbidity threshold and was not moderated by age. The goal of this work was to explore the effect of GWI on epigenetic measures of aging. GWECB consists of DNA and survey data from 1275 individuals who served in the US Armed Forces in 1990-1991. Self-reported symptoms and their severity captured by the GWECB survey determined Centers for Disease Control GWI symptom domains and GWI case status (No GWI, Mild GWI, Severe GWI). Methylation data were generated from the Illumina EPIC array with Illumina quality control (QC) performed at the Pharmacogenomics Analysis Laboratory. The methylation analysis pipeline developed for the Million Veteran Program was used for additional sample, probe and batch QC and quantile-normalization. Epigenetic age (EA), age acceleration and smoking status measures were calculated using the R packages methylclock, DunedinPACE. and EpiSmokEr. Association

testing of EA with GWI case status, symptom domains, deployment to the Gulf War, and Military Unit Component was performed using linear regression unadjusted and adjusted for chronological age, gender, race, smoking status. The totals meeting QC thresholds were 1036 participants (181 no GWI, 629 mild GWI and 226 severe GWI) and 679491 (CpGs). The correlation of age measures with chronological age are 0.83 (Horvath), 0.84 (Hannam), 0.76 (Levine), 0.91 (SkinHorvath), 0.94 (BLUP) and 0.93 (EN). Using these measures, no association of EA estimates with GWI Severe was observed. Overall, the mean (SD) DunedinPACE score was 1.26 (0.13). In adjusted models, relative to no GWI, the effect size for Mild GWI was 0.015 (p=0.14) and for Severe GWI was 0.050 (p=4.73E-05). Similar effects were observed with the individual GWI severe symptom domains: mood 0.031 (p=2.25E-04), fatigue 0.043 (p=7.71E-05), musculoskeletal 0.042 (p=7.91E-07). There was no additional association with deployment to the Gulf War 0.023 (p=0.44) and Military Unit Component 0.026 (p=0.39).In summary, EA measures designed to detect associations with age, mortality and morbidity did not show a statistically significant association with GWI Severe case status. The DunedinPACE measure, designed to measure the pace of aging, showed a statistically significant association with GWI.

Session Title: Epigenetics Poster Thursday Session

Board 1017T: Epigenetic and genetic diagnostic innovations for differentially methylated regions in cell-free DNA from cancer patients and high-throughput urine samples, respectively

Authors:

R. Sen[1], C. Michelon[1], B. Egan[1], B. Delatte[1], M. Evans-Holm[2], N. Lumbres[2], P. Montilla-Perez[1], M. A. Paradise[1], S. E. Mayorga[1], P. Lentz[2]; [1]Active Motif, Carlsbad, CA, [2]Convergent Genomics, South San Francisco, CA

Abstract:

Here we present novel modifications to improve an epigenetic and a genetic method using low input, minimally invasive liquid biopsy samples for diagnostic and therapeutic applications. Epigenetic modifications of DNA regulate transcription, development and disease. The most abundant epigenetic DNA modification in eukaryotes is 5-Methylcytosine (5-mC) which alters gene expression and genome metabolism. Deregulation of 5-mC is associated with many diseases including cancers. Study of mutations or epimutations (e.g. 5-mC variations) offers a high diagnostic potential to assess the presence, stage and outcome of several cancers. Recent focus is on 5-mC in cell-free DNA (cfDNA) because cfDNA is a sample derived from liquid biopsy using minimally invasive procedures, hence less traumatic for patients. cfDNA is often released

from apoptotic cells, and methylation patterns of cfDNA reflects that of actual tumors. However, detection of 5-mC in cfDNA is challenging because input material is often quite scarce. Hence, we present a modified method using Methyl-CpG-binding domain protein 2 (MBD2) to capture and detect highly methylated DNA genome-wide. Our method has high sensitivity which works with as low as 5 nanograms of input cfDNA, with specificity over 100-fold mC:C detection ratio in cfDNA. Our results show altered methylation patterns in cfDNA from healthy versus patients of colorectal and breast cancers with striking differences in methylation of key tumor-associated genes. The method detects epigenetic modifications using cfDNA which is a low input sample, obtained using minimally invasive liquid biopsy procedures. Another liquid biopsy sample collected non-invasively is urine which contains low input genetic material for disease research. We further present a workflow where high-throughput urine samples are processed in a consistent, quick and economical manner using pixelated-ultrasound technology. The technology significantly improves sample preparation workflow for subsequent next-generation sequencing (NGS) analysis to diagnose mutations associated with cancers and other diseases. The technology offers a reliable platform to obtain consistent high-throughput samples which pass critical quality control to proceed with NGS. Together, our results present novel modifications to two methods - one epigenetic and one genetic - using minimally invasive liquid biopsy-derived samples like cfDNA and urine, both of which are sources of low input genetic material, with diagnostic and therapeutic applications.

Session Title: Epigenetics Poster Thursday Session

Board 1018T: Epigenetic Evidence that Heavy Alcohol Consumption is a Common, Large Risk Factor for Mortality in Patients with Acute Coronary Syndrome.

Authors:

R. Philibert[1], A. Andersen[2], S. R. Beach[3], J. Mills[4], E. Milefchik[1], P. Emma[1], K. Dawes[1], M. Joanna[1], G. Weeks[1], E. Froehlich[1], K. deBlois[1], J. Long[4], F. Ahmad[1]; [1]Univ Iowa, Iowa City, IA, [2]Univ Iowa Hosp. & Clinics, Iowa City, IA, [3]Univ. of Georgia, Athens, GA, [4]Univ. of Iowa, Iowa City, IA

Abstract:

Background: Heavy alcohol consumption (HAC) is a risk factor for the development of coronary heart disease (CHD). In part, because of the unreliability of self-reported intake of alcohol, the relationship of HAC to mortality in those with CHD, and whether the relationship of HAC to CHD is moderated by other clinical risk factors is not well understood. The use of objective biomarkers of alcohol consumption could help bridge this knowledge gap. **Methods:** In order to

improve our understanding, we examined the relationship of Alcohol T-Scores (ATS), an epigenetic biomarker of chronic HAC, with all-cause mortality in 217 subjects admitted for CHD-related acute coronary syndrome (ACS). Then, in an attempt to set the basis for understanding how alcohol could lead to CHD, we examined the relationship of ATS values to clinical risk factors for CHD in a separate prospectively ascertained cohort of 304 subjects. **Results:** We found that 65% of the subjects admitted for ACS had ATS values indicative of chronic HAC. ATS values were significantly elevated ($p<0.02$) in those who died ($n=28$) with a Cox Proportional Hazards model showing a slightly larger effect of ATS levels than age on all-cause mortality survival (overall model, $p<0.003$). Subjects in the highest decile of ATS scores had a 2.4-fold increase in the risk for mortality as compared to those in the lowest decile. Although men and women had similar ATS values, 24% of the men, but only 4% of women, had a current or past diagnosis of an alcohol use disorders (AUD). Only two subjects were referred for substance use evaluation during hospitalization. Regression analysis of ATS values to clinical risk factors for CHD in a separate prospectively collected cohort showed nominally significant positive correlations of ATS values with systolic blood pressure, triglycerides and smoking intensity, and inversely with body mass index. **Conclusions:** These findings suggest that HAC is an underappreciated driver of CHD related morbidity and mortality. They also suggest that HAC is particularly under-recognized in women. These findings are alarming because alcohol use disorders are inherently treatable and suggest that by systematically assessing those admitted for CHD for alcohol use disorders, many deaths could be averted. Further studies to confirm these findings and to understand the mechanisms through which alcohol exerts its deleterious effects in those with CHD are indicated.

Session Title: Epigenetics Poster Thursday Session

Board 1019T: Epigenetic scores of plasma proteins as biomarkers of Amyotrophic lateral sclerosis

Authors:

X. Qian[1], A. Al-Chalabi[2], J. Veldink[3], F. Garton[4], A. McRae[1]; [1]Inst. for Molecular BioSci., Brisbane, Australia, [2]Inst Psychiatry, London, United Kingdom, [3]Univ Med Ctr Utrecht, Utrecht, Netherlands, [4]Univ. of Queensland, Brisbane, Australia

Abstract:

Amyotrophic lateral sclerosis (ALS) is a complex neurodegenerative disease with mechanisms that are still being understood. Changes in circulating proteins can provide insights into these mechanisms and support biomarker development, but large-scale quantification is often

prohibitively expensive. Here, we explore EpiScores, a cost-effective alternative using protein epigenetic scores derived from DNA methylation data (DNAm). These EpiScores served as proxies for 109 circulating proteins in ALS cases and controls across three cohorts: the Australian ALS cohort (1,395 individuals: 782 cases and 613 controls), the King's College London (KCL) ALS cohort (1,348 UK individuals: 1,078 cases and 270 controls), and the Netherlands ALS cohort (1,816 individuals: 1,175 cases and 641 controls). Association models corrected for age, sex, and predicted cell type proportions in the DNA methylation measurement were used. A meta-analysis identified 14 significant proteins: BMP-1, E-selectin, Lymphotoxin abeta, Aminoacylase-1, CCL25 C-C, PAPPA, GHR, FcRL2, B2-microglobulin, EZR, Complement C9, GZMA, CCL11, and VCAM1. Notably, GHR and PAPPA, both critical for IGF-1 generation and delivery, are part of a pathway that has had substantial interest in ALS. These findings, including previously unidentified proteins in ALS, offer new insights into the disease's underlying mechanisms and potential avenues for biomarker discovery.

Session Title: Epigenetics Poster Thursday Session

Board 1020T: Epigenome-wide association study in multi-ethnic Asian population identifies novel markers for skin physiology measures and atopic dermatitis

Authors:

M. Loh[1,2,3,4], H. Ng[1], D. Tay[1], Y. Yew[4,1]; [1]Nanyang Technological Univ., Singapore, Singapore, [2]Genome Inst. of Singapore, Singapore, Singapore, [3]Imperial Coll. London, London, United Kingdom, [4]Natl. Skin Ctr., Singapore, Singapore

Abstract:

IntroductionSkin diseases are a major public health problem. While mortality rate is low, individuals with skin diseases often suffer from impaired quality of life plagued by itch and pain. Epidermal barrier plays a critical function in skin health. Recent studies have suggested the involvement of DNA methylation in affecting epidermal differentiation, barrier dysfunction and the innate immune system. However, these studies have been predominantly carried out in small number of individuals on a targeted basis. In this study, we aim to identify variations in DNA methylation associated with skin physiology measures and atopic dermatitis (AD) among a large multi-ethnic Asian population.
MethodsWe performed an epigenome-wide association study (EWAS) against skin physiology measures (transepidermal water loss (TEWL), skin surface moisture and skin pH), using peripheral blood samples from East Asians, Malays and South Asians from Singapore (n=2,147).

We additionally performed an EWAS against AD in East Asians only (n=140 cases; 1,630 controls), in view of the low case numbers in Malays and South Asians. DNA methylation was measured on the Illumina EPIC array, with epigenome-wide significance set at P<8.62E-8, as determined via permutation testing.
ResultsOne CpG site reached epigenome-wide significance for skin pH (cg23222772; *CHSY1*; P=1.83E-08). CHSY1 is expressed in multiple cell types throughout the epidermis and dermis, and has been proposed to contribute to increased chondroitin sulfate, a glycosaminoglycan responsible for keeping the skin hydrated. At a nominal p-value threshold of P<1E-05, we further identified a total of 9, 20 and 21 CpG sites respectively for TEWL, moisture and pH in the meta-analyses, and 7 CpGs for AD in Chinese. Initial annotation and expression quantitative trait methylation (eQTM) analysis suggest that the identified CpG sites are likely to be play functionally important roles. *ALX4*, annotated to cg07696033 for TEWL (P=1.39E-06) and to cg07736627 for skin moisture (P=1.18E-06), is a homeobox gene that has been reported to play a role in skin differentiation and hair follicle development, and is involved in skin layer formation. Further downstream analyses such as Mendelian Randomization and colocalization, accompanied by pathway analyses are in progress.
ConclusionTo the best of our knowledge, this is the first epigenome-wide analysis to date conducted on skin physiology measures and AD. The identification of these methylation markers could provide new insights into pathways underlying AD and other skin diseases, and inform the basis of future experiments and potential therapeutic agents.

Session Title: Epigenetics Poster Thursday Session

Board 1021T: Evaluating Methylation Analysis on Microbial Community Standards to Advance the Study of the Human Microbiome

Authors:

X. Yang[1], R. Yancey[1], A. De Silva[1], J. Bhasin[2], K. Booher[1]; [1]Zymo Res. Corp., Irvine, CA, [2]Zymo research, Irvine, CA

Abstract:

The human microbiome, comprising all microorganisms living in and on our bodies, significantly impacts health by influencing immune responses, skin health, respiratory function, and mental well-being. Disruptions in the microbiome are linked to health issues such as allergies, autoimmune diseases, skin disorders, and neurological conditions, and the microbiome is now considered our “last organ.” Epigenetic modifications regulate the microbiome by altering gene expression without changing the DNA sequence, affecting microbial behavior and interactions.

Understanding these modifications is crucial for developing therapeutic strategies and improving industrial applications involving microorganisms. Previously, the lack of tools to analyze methylated residues in bacterial genomes hindered our understanding of their functions. However, new DNA sequencing technologies, including single-molecule, real-time sequencing, and nanopore-based sequencing, now allow for the systematic detection of all forms of methylated DNA at a genome-wide scale, offering unprecedented opportunities to understand bacterial epigenomes. Despite these advancements, existing methods and modification prediction algorithms have limitations that can lead to inconsistent results, especially with varying epigenetic modification abundances. To address this, we benchmarked epigenetic modifications (m4C, m6A and m5C) in ZymoBIOMICS Microbial Community Standards, a well-defined mock community of Gram-negative and Gram-positive bacteria and yeast with varying sizes and cell wall compositions. We detected the methylation landscape in Microbial Community Standards and identified diverse methylation motifs and differences in genomic elements. We also compared and validated different tools for future de novo discovery. This study paves the way for integrating methylation signals in sub-strain analysis and provides community standards for benchmarking and assessing the performance of metagenomic workflows.

Session Title: Epigenetics Poster Thursday Session

Board 1022T: Extensive methylation array variant-under-the probe (VUP) binding disruption: evidence for probe binding disruption for VUPs up to 50 bp from the CpG

Authors:

A. Annis[1], J. Lefaive[1], T. Blackwell[1], G. R. Abecasis[1], E. K. Silverman[2], D. L. Demeo[2], A. V. Smith[1], L. J. Scott[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Brigham & Womens Hosp., Boston, MA

Abstract:

DNA methylation is essential for physiological development and is influenced by genetic and environmental factors. Probe-based array data provide estimated methylation levels; these can be distorted by genomic Variations Under the Probe (VUPs) that disrupt probe binding. If undetected, VUPs can elevate the contribution of background fluorescent signal and bias the observed percent methylation, increasing noise (decreasing power) or causing incorrect inference in QTL and causal inference analyses.
Using TOPMed Lung Tissue Research Consortium (n=1,248) whole genome sequence and blood-based MethylationEpic Beadchip array data, we quantitated the proportion of individuals and

CpG probes (n=719,320) with VUPs and identified the extent to which VUPs at different positions under the probe (1-50 bp) disrupt probe binding. 66.1% of probes had ≥1 individual with a VUP, 19.4% had >10, and 9.7% had >100. European-ancestry individuals had a mean of 0.37±0.01 VUPs per probe (n=1,141). African-ancestry individuals (n=51) averaged 17.9% more VUPs per probe (mean = 0.43±0.01).
We estimated probe binding disruption as the mean homozygous alternate minus mean homozygous reference $\log_{10}$ total fluorescent intensity (alt-ref difference): mean of -0.19±0.22 for 21,678 type II single VUP probes with ≥1 homozygous alternate allele individual. Negative values indicate VUP alternative allele disrupted probe binding.
To characterize the extent of binding disruption, we compared the VUP-based alt-ref differences to probes with: 1) no binding disruption (no VUP and a single non-probe side variant 1 to 50 bp from the CpG; n=24,659) and 2) complete binding disruption (only background signal detected for alternate variants at the guanine of the CpG, VUP position 1; n=1,224). 88.7% of the VUP alt-ref differences fell between the means of the alt-ref differences for fully bound (-0.01±0.09) and completely disrupted (-1.06±0.32) probes.
To assess probe binding disruption by VUP distance from the CpG, we calculated mean alt-ref differences for probes 2-3, 4-10, 11-20, 21-30, 31-40, and 41-50 bp from the CpG; we found means of -0.78±0.31, -0.32±0.25, -0.19±0.17, -0.18±0.17, -0.15±0.16, and -0.06±0.11 (n=606, 2,636, 4,600, 4,489, 4,818, and 4,529), all significantly different from the fully bound probe mean ($p<2\times10^{-16}$, one-sided t-test). We found that VUP-caused binding disruption can occur at any point along the probe, including modest disruptions at 41-50 bp from the CpG. This suggests the need for probe-specific approaches to control false positive associations and retain power in analyses, particularly to retain information for samples with African ancestry individuals.

Session Title: Epigenetics Poster Thursday Session

Board 1023T: Genome Organization with CUT and Tag (GO-CaT) identifies substructure of lamina-associated domains in developing human brain neurons

Authors:

C. Zhang, E. Gil, S. Hamid, C. Wang, L. Wang, J. Li, A. Kriegstein, D. Lim; Univ. of California, San Francisco, San Francisco, CA

Abstract:

Approximately 30-40% of the human genome is attached to the lamina through lamina-associated domains (LADs) of varying sizes (10 kb-10 Mb). Previous studies in cultured murine and human cell lines have revealed that LADs possess a substructure with diverse characteristics. However, the substructure of LADs in cells from developing human tissues has yet to be examined. In this study, we developed Genome Organization with CUT and Tag (GO-CaT) to map LAD substructure in neurons of the developing human brain. While T1- LADs have the transcriptional and epigenomic characteristics of cell type-invariant LADs, T2- LADs had a distinct epigenomic state that include a strong enrichment of promoter-enhancer DNA interactions despite a lowly- expressed transcriptional level. In the developing neurons, T2-LADs define genomic regions that later play key roles in neuronal differentiation and function. Furthermore, T2- LADs are enriched for neurological GWAS phenotypes including those of human cognitive disorders. These studies highlight the utility of GO-CaT for LAD subdomain mapping in human tissues and the importance of nuclear compartment associated genome organization.

Session Title: Epigenetics Poster Thursday Session

Board 1024T: Genome-wide dysregulation of R-loops in Ataxia Telangiectasia neurological pathogenesis

Authors:

K. Westover, Y. Hou, Y. Li, Y. Wang, B. Yao; Emory Univ., Atlanta, GA

Abstract:

Ataxia Telangiectasia (AT), a neurodegenerative disease characterized by cerebellar degeneration of Purkinje cells that control balance and movement, affects up to 1 in 40,000 to 100,000 people worldwide. A recessive early childhood onset disorder, AT is caused by mutations within the ataxia telangiectasia mutated (ATM) threonine/serine kinase which plays crucial roles within the DNA damage response (DDR). However, the precise molecular mechanisms underlying AT pathogenesis and how ATM loss-of-function leads to deficient DDR remain elusive. R-loops, three stranded RNA-DNA structures composed of an DNA-RNA hybrid and a non-template DNA strand, have emerged as key components of double strand break (DSB)-induced DDR. Mounting evidence has documented critical roles of R-loops in both causing and responding to DSBs. As DSBs and the failure of their repair play major roles in the pathology of AT, R-loop dysregulation is likely to contribute to AT pathogenesis. We have successfully reprogrammed AT patient-derived fibroblasts and lymphoblastoid cells lines into induced pluripotent stem cells (iPSCs) and differentiated them into neuronal progenitor cells (NPCs). Our

preliminary data has demonstrated that AT patient-derived NPCs contain higher baseline levels of DSB markers yH2A.X and 53BP1, as well as higher levels of R-loop accumulation than age-matched control NPCs. Induction of DNA damage, in the form of DSBs, by irradiation results in increased levels of both R-loops and damage markers compared to untreated control cells after 6 hours of recovery time. We are currently utilizing DNA-RNA Immunoprecipitation and sequencing (DRIP-seq) to systematically assess global and loci-specific alterations in R-loop distribution in AT-derived NPCs as well as the aberrant response of R-loops to irradiation. We are also utilizing RNA-seq to explore how this dysregulation may affect gene expression, potentially leading to disease progression. We hypothesize that in AT, the lack of target phosphorylation by ATM could globally dysregulate R-loop formation in patient-derived cells and underlie AT progression. Understanding the molecular mechanism underlying R-loop dysregulation in AT and its downstream impact on gene expression will allow us to identify new potential therapeutic approaches.

Session Title: Epigenetics Poster Thursday Session

Board 1025T: Genomic architecture of lowly expressed bidirectionally transcribed enhancers and their links to complex traits

Authors:

M. Koido[1,2], K. Tomizuka[2], K. Takeuchi[3,4], R. Son[3,4], A. Oguchi[3,4], H. Sano[4], Y. Ishikawa[2], S. Bhagat[3], T. Kasahara[3], X. Liu[2], N. Tanaka[2], T. Hirai[4], K. Hamanaka[3], K. Toritani[4], Y. Murakawa[3,4], C. Terao[2,5,6]; [1]Lab. of Complex Trait Genomics, Dept. of Computational Biology and Med. Sci., Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Tokyo, Japan, [2]Lab. for Statistical and Translational Genetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [3]Inst. for the Advanced Study of Human Biology, Kyoto Univ., Kyoto, Japan, [4]RIKEN-IFOM Joint Lab. for Cancer Genomics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [5]Clinical Res. Ctr., Shizuoka Gen. Hosp., Shizuoka, Japan, [6]The Dept. of Applied Genetics, The Sch. of Pharmaceutical Sci., Univ. of Shizuoka, Shizuoka, Japan

Abstract:

Active enhancers define cell identity and are closely linked to the basics of complex traits. Bidirectionally transcribed candidate enhancers (btcEnhs) can delineate active enhancer regions at high resolution and facilitate fine-mapping of genetic associations in complex traits. However, their quite low expression levels and technical noises hamper the comprehensive identification of these enhancers. Here, we developed a nearly random pool index method to sensitively detect cross-donor differences and deeply sequenced the 5’-end of RNAs from 432

lymphoblastoid cell lines. From the genuine 5'-end reads in 87 million uniquely mapped reads per sample, we identified 200,000 btcEnhs, of which 129,000 btcEnhs (low-btcEnhs) were detectable in less than half of the samples. Notably, low-btcEnhs well marked known regulatory elements: 75% and 45% of low-btcEnhs overlapped with open chromatin regions and distal enhancer-like signatures in ENCODE3, respectively. We found that, across expression levels, btcEnhs were conserved and associated with polygenic signals for immunological diseases (for example, 51-fold heritability enrichment of low-btcEnhs for multiple sclerosis). Finally, we identified 13,382 btcEnhs genetically associated with 1.4 million variants (false discovery rate < 5%; termed eEnhancer), 19% of which were low-btcEnhs. The eEnhancers were enriched in constrained regions and putative causal variants of especially leucocyte-related traits, particularly when sharing genetic regulations with nearby eEnhancers. Our results provide the comprehensive genomic architecture of btcEnhs in the B-cell lineage and their genetic associations with complex traits.

Session Title: Epigenetics Poster Thursday Session

Board 1026T: Identification of a shared transcriptomic signature in sex chromosome aneuploidies: A study on 47,XYY and 47,XXY iPSCs ⭐

Authors:

A. Adamo, V. Astro, K. Cardona-Londoño, L. Cortés-Medina, R. Alghamdi, G. Ramírez-Calderón, F. Kefalas; King Abdullah Univ. of Sci. and Technology, Thuwal, Saudi Arabia

Abstract:

Male sex chromosome aneuploidies are frequent genetic aberrations in humans characterized by additional X or Y complements in the contest of a 46,XY karyotype. Among male aneuploidies, Jacob (47,XYY) and Klinefelter (47,XXY) syndromes share several clinical features, including sterility, hormonal deficits, neurocognitive delay, and skeletal-muscle defects. Despite the high incidence, a comprehensive understanding of the molecular mechanisms underlying the onset of the clinical features of sex aneuploid patients remains elusive. In this study, we generated and characterized the inaugural cohort of 47,XYY human induced pluripotent stem cells (iPSCs). We conducted a comprehensive transcriptional analysis encompassing JS primary fibroblasts, iPSCs, and neural stem cells (NSCs) juxtaposed with KS cells. Our study unraveled the existence of a transcriptional feedback mechanism modulating non-PAR X chromosome genes (NPX) homologs in Y supernumerary cells, a phenomenon not detected in X aneuploid iPSCs. Furthermore, our analyses identified a shared transcriptomic signature between JS and KS, discernible already at the iPSC stage, with a notable enrichment for processes related to

neurological functions. This transcriptomic convergence underscores potential commonalities in the molecular pathways underpinning the pathophysiology of male sex chromosome aneuploidies. Importantly, we found that only minimal DNA methylation differences exist between 47,XYY and 46,XY cells. On the other hand, in supernumerary X chromosome cells, the well-established phenomenon of DNA methylation spreading on the supernumerary X chromosome in KS has been linked to a global transcriptional impact in primary cells. Therefore, the identification of highly similar patterns of differentially expressed genes in JS and KS could indicate that, in X aneuploidies, the DNA methylation of the supernumerary X could play a marginal role in regulating the global transcriptome at the pluripotent stage. Our work expands the knowledge on Y and X chromosome cross-modulation and their impact on human gene expression and suggests that disease modeling of sex aneuploidy through iPSCs may provide an ideal platform for understanding the global effects of aneuploidies in disease-relevant cell types.

Session Title: Epigenetics Poster Thursday Session

Board 1027T: Integrated analysis of genetic variation, gene expression and methylation changes in temporal lobe epilepsy

Authors:

J. Kalami[1], **G. E. Kuodza**[2], M. N. Nyamari[3], M. K. Goita[4], S. Osata[5], O. H. Ifeoluwa[6], A. Olaitan[7]; [1]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [2]Univ. of California, Davis, CA, [3]Pwani Univ., Kilifi, Kenya, [4]Univ. of Sci. Techniques, and Technologies of Bamako (USTTB), Bamako, Mali, [5]Univ. of Nairobi, Nairobi, Kenya, [6]Obafemi Awolowo Univ. Teaching Hosp., Ile-Ife, Nigeria, [7]African Society for Bioinformatics and Computational Biology, Cape Town, South Africa

Abstract:

Background: Epilepsy is a neurological disorder characterized by recurrent seizures due to abnormal brain activity. Temporal lobe epilepsy (TLE) is a common form, often resistant to treatment. Understanding the molecular mechanisms underlying TLE can inform better diagnosis and therapeutic strategies. **Methods:** We analyzed public datasets from Gene Expression Omnibus (GEO) to identify differentially expressed genes (DEGs) and differentially methylated regions (DMRs) in hippocampal and neocortex tissues of mesial temporal lobe epilepsy (mTLE) patients, as well as high-frequency (HFS) and low-frequency seizure (LFS) patients. Variant calling and gene ontology analyses were conducted to elucidate functional pathways. **Results:** Our analysis identified 2139 DEGs between hippocampus and neocortex tissues, significantly enriching gliogenesis, immune response, ion transport, and metabolic

processes. Additionally, 2,195 DMRs distinguishing TLE with and without hippocampal sclerosis were linked to neuron projection development, synapse assembly, and signal transduction pathways. Overlapping genes between DEGs, DMRs, and a specific epilepsy-associated gene set included *CELSR1*, *RBFOX3*, *SYN2*, *CDH13*, and others, emphasizing their potential as biomarkers for TLE. **Conclusion:** This study highlights the complex interplay between gene expression and methylation in epilepsy. Identifying biomarkers such as *CELSR1*, *RBFOX3*, *SYN2*, *CDH13*, and others offers promising avenues for early diagnosis and personalized treatment of TLE. Future research should validate these findings in more extensive, diverse cohorts to develop effective, targeted therapies for improving patient outcomes.

Session Title: Epigenetics Poster Thursday Session

Board 1028T: Integrative analysis of RNA-Seq, long-read DNA sequencing, and clinicopathological profiles of LINE-1 retroelement in non-small cell lung cancer

Authors:

Y. Wang; Texas A&M Inst. of BioSci.s and Technology, Houston, TX

Abstract:

Aberrant expression of LINE-1 retrotransposons and their encoded ribonucleoprotein LINE-1 ORF1p and reverse transcriptase ORF2p has been associated with non-small cell lung cancer (NSCLC) tumorigenesis and poor clinical outcomes. In the present study, we generated a comprehensive profile of LINE-1 retroelement expression in NSCLC patient populations by examining the LINE-1 status of 121 NSCLC cell lines using in silico analyses of RNA-seq databases and integrating these metadata with long-read DNA sequencing, measurements of ORF1p in tumor and non-tumor patient samples, and clinicopathological profiles including sex, race, age, and immune cell infiltration. Unbiased analysis of these resources revealed that LINE-1 transcript reads along with 3’ run-on and antisense reads were higher in lung squamous cell carcinomas (LSQCC) compared to lung adenocarcinomas (LUAD), adenosquamous carcinomas, and large cell carcinomas. While no significant differences in LINE-1 transcripts were noted between males and females, multivariant analyses showed elevated transcript levels in tumor cell lines derived from African Americans compared to Asians or Caucasian Americans. The expression of LINE-1 transcripts was positively correlated with age, particularly for LINE-1 run-on transcripts. Immunohistochemical analyses of LINE-1 ORF1p in tissue sections of LSQCC and LUAD, along with their matched, adjacent non-tumor counterparts, revealed significant increases in ORF1p expression in NSCLCs compared to non-tumor tissues, with higher levels in

LSQCC compared to LUAD. As seen with NSCLC cell lines, male and female tissue samples showed similar profiles irrespective of tumor type, with remarkably higher levels of ORF1p seen in African Americans compared to Caucasian Americans, and higher levels in older males (>75) compared to younger males. Stratification by age showed a positive relationship of ORF1p with age only in tumor samples. Quantitative immunofluorescence analysis colocalized LINE-1 ORF1p with pan-immune markers for T-cells, macrophages and neutrophils and revealed a strong positive correlation between ORF1p and CD68 in both LSQCC and LUAD, and to a lesser extent in non-tumor tissues. Together, these findings show that profiles of LINE-1 transcripts and ORF1p in NSCLC are strongly influenced by race and age, and possibly linked to intricate interactions regulated by macrophage infiltration into the tumor microenvironment.

Session Title: Epigenetics Poster Thursday Session

Board 1029T: Investigating the Role of Small Cajal Body-Associated RNAs (scaRNAs) in Congenital Heart Defects: A Study of SNORD94 Regulation of Alternative Splicing

Authors:

K. Jones[1], L. M. Holets-Bondar[2], N. Kibiryeva[1], M. Filla[1], J. O'Brien[2], D. Bittel[3]; [1]Kansas City Univ. Coll. of Med. and BioMed. Sci., Kansas City, MO, [2]Children's Mercy Hosp., Kansas City, MO, [3]Kansas City Univ., Kansas City, MO

Abstract:

Despite their prevalence, the underlying mechanisms that contribute to Congenital Heart Defects (CHDs) are poorly understood. Our lab has previously demonstrated that small cajal body-associated RNAs (scaRNAs), specifically SNORD94, play a crucial role in regulating alternative splicing during embryonic development. Abnormalities in spliceosomal regulation have been implicated in the pathogenesis of Tetralogy of Fallot, a common CHD. Here, we investigate the role of SNORD94 in cardiac developmental abnormalities using a novel approach involving CRISPR-Cas13, a RNA-targeting tool. We demonstrate that knockdown of SNORD94 in quail myoblast cells using CRISPR-Cas13 leads to aberrant alternative splicing and altered cardiac development. Our findings provide evidence for the first time that scaRNAs, particularly SNORD94, contribute to the development of CHDs, and highlight the potential of CRISPR-Cas13 as a tool for functional analysis in developmental studies. This research has significant implications for our understanding of the molecular mechanisms underlying CHDs and may ultimately lead to the development of novel therapeutic strategies for their treatment.

Session Title: Epigenetics Poster Thursday Session

Board 1030T: Machine learning reveals the diversity of human 3D chromatin contact patterns

Authors:

E. Gilbertson; UCSF, San Francisco, CA

Abstract:

Understanding variation in chromatin contact patterns across human populations is critical for interpreting non-coding variants and their effects on gene expression and phenotypic variation. However, experimental determination of chromatin contacts at a population-scale is prohibitively expensive. To overcome this challenge, we developed a machine learning method to quantify the diversity 3D chromatin contacts at 2 kilobase resolution from genome sequence alone. We validated that it accurately generalized on contact data from diverse individuals. We then applied this approach to thousands of diverse modern humans and the inferred human-archaic hominin ancestral genome. Patterns of 3D contact divergence genome-wide are qualitatively similar to patterns of sequence divergence, but we find that 3D divergence in local 1-megabase genomic windows does not follow sequence divergence. In particular, we identify 392 windows with significantly greater 3D divergence than expected from sequence. Moreover, 26% of genomic windows have rare 3D contact variation observed in a small number of individuals. Using *in silico* mutagenesis, we find that most sequence changes to do not result in changes to 3D chromatin contacts. However, in windows with substantial 3D divergence, just one variant can lead to divergent 3D chromatin contacts without high overall sequence divergence.In summary, inferring 3D chromatin contact maps across human populations reveals diverse contact patterns. We anticipate that these genetically diverse maps of 3D chromatin contact will provide a reference for future work on the function and evolution of 3D chromatin contact variation across human populations.

Session Title: Epigenetics Poster Thursday Session

Board 1031T: Metabolomic analysis indicates adenine, oleic acid, and glutamic acid are associated with flare remission, DNA methylation changes, and clinical subtypes in systemic lupus erythematosus

Authors:

M. Horton[1], D. Johnson[2], J. Nititham[1], K. Taylor[3], P. Katz[3], C. Ye[3], J. Yazdany[3], M. Dall'Era[3], L. Barcellos[4], J-L. Li[2], K. Overdahl[2], A. Jarmusch[2], L. Criswell[1], C. Lanata[1]; [1]NIH/NHGRI, Bethesda, MD, [2]NIH/NIEHS, Durham, NC, [3]Univ. of California, San Francisco, San Francisco, CA, [4]Univ. of California, Berkeley, Berkeley, CA

Abstract:

Background: Systemic lupus erythematosus (SLE) is a heterogeneous disease that follows a flare-remitting pattern. Current treatments neither adequately prevent flares nor disease-related organ damage. We previously identified changes in DNA methylation over time associated with flare remission (291 CpGs) and used methylation changes over time to define 3 patient clusters. The objectives of the current study are to integrate metabolomics into this analysis to identify whether changes in metabolites are associated with 1) flare remission and 2) changes in DNA methylation previously associated with SLE remission and subtypes.
Methods: We utilized 40 multi-ethnic SLE patients recruited during an active flare with whole blood DNA methylation profiles (Illumina EPIC array) and untargeted metabolomics data from plasma (UHPLC-HRMS/MS) at the baseline flare and a follow-up visit ~3 months later. We observed metabolites (with annotation) that changed between visits differentially by whether a patient remitted at the follow-up visit (disease activity score=0) or not using *limma*. Significant metabolite changes (FDR $q<0.05$) were tested for their association with methylation changes at 291 sites using correlation coefficients. We also tested whether changes in metabolites were associated with 3 methylation patient clusters using linear regression.
Results: Sixteen SLE patients remitted by follow-up. Remitters and non-remitters did not differ significantly by race, ethnicity, age, disease activity or symptoms at flare, or medications. We identified 3 metabolite changes associated with remission status: oleic acid (difference in change comparing remitters to non-remitters: 2.52, 95% CI: 1.39, 3.35), adenine (1.61, 95% CI: 0.88, 2.34), and glutamic acid (-1.63, 95% CI: -2.42, -0.83). Changes in these were associated with previous methylation-based patient clusters. We identified 7 methylation changes highly correlated with changes in metabolite levels (coefficient $>|0.7|$ and Bonferroni $p<0.05$). These included cg17847344 (within the body of *EBF1*, a key transcription factor of B-cell specification and commitment) and oleic acid (correlation= -0.79, $p=1.06\times10^{-7}$) as well as cg10685380 (within the body of *IL12B*, which encodes a cytokine that acts on T and natural killer cells) and adenine (0.70, $p=1.26\times10^{-6}$).
Conclusions: Results show that changes in metabolites were associated with remission status and were correlated with changes in methylation at SLE-relevant loci. The integration of DNA methylation and metabolomics might help us better understand underlying biological pathways relevant for SLE flare remission and result in targeted therapies.

Session Title: Epigenetics Poster Thursday Session

Board 1032T: MeVars (DNA Methylation Variants): Identification, tissue-specificity, reprogramming effects and disease susceptibility.

Authors:

N. Kommu[1], A. Anne[1], L. Kumar[2]; [1]BITS Pilani, Hyderabad Campus, Hyderabad, India, [2]Natl. Inst. of Technology, Kurukshetra, India

Abstract:

About one-third of the CpG sites in the Infinium Human 450K Methylation array were previously reported by us to show significant methylation differences among normal cerebellum samples. Here we analyzed genome-scale DNA methylation data of > 1000 samples from different tissues of ectodermal, endodermal and mesodermal origins, and identified many regions that displayed normal methylation variation with high effective values. These variants, referred to as Methylation Variants (MeVars) are enriched in genes associated with different neurological disorders and different forms of cancer. An extremely large proportion of MeVars showed tissue-specificity; systemic variants are less frequent. This somatic variation was significantly reduced in primed iPSCs derived from the corresponding tissues and in semen. A sizeable proportion of MeVars are shared across ethnic groups however, there appears to be some degree of ethnic specificity. Analysis of MeVars in the candidate genes *AURKC* and *NAPRT1* associated with male infertility and schizophrenia, respectively showed a negative correlation between the methylation and transcript levels. Using data on oral tissues, identified 62 MeVars that were correlated with gene dysregulation, further providing evidence of their functional effects. Significant differences in the occurrence of MeVars were observed between normal and affected (autism spectrum and schizophrenia) brain samples. About 25-30% of MeVars occurred in schizophrenia candidate genes of which a subset was associated with significant alterations in gene expression. Thus, MeVars are conserved across ethnic groups, largely contribute to tissue-specific interindividual variation, susceptible to both *in vitro* and *in vivo* reprogramming events and distinguish between genetically normal individuals. A subset of MeVars occurring in candidate genes of Mendelian and complex disorders confer disease-susceptibility by influencing gene expression and altering the functional gene dosage.

Session Title: Epigenetics Poster Thursday Session

Board 1033T: Molecular cross-talk in vascular calcification

Authors:

D. McDaniel[1], L. S. McDaniel[2], C. K. Moore[1], W. F. Campbell[1]; [1]Univ. of Mississippi Med. Ctr., Jackson, MS, [2]Univ Mississippi Med Ctr, Jackson, MS

Abstract:

Vascular calcification (VC) is a life-threatening condition associated with pathophysiological mechanism initiated by genetic variations, chronic inflammation, oxidative stress, demonstrated with osteogenic transdifferention (Ost-Diff) of vascular endothelium known as (EndoMT). Together with calcium and phosphate abnormality, the ECs participates in vascular & coronary artery calcification. We hypothesized that the VC calcification is assisted by three major entities; 1) genetic factors, 2) chronic inflammation, specifically due to the heterogeneity of macrophages, and 3) metabolic pathways such as glycolysis. Genome-wide association studies in populations have identified many genes associated with VC & CAC, but most recently, in addition to endothelin-1 (Endo-1), four novel genes have been identified as important factors for calcification. The genes are insulin like growth factor binding protein (IGFBP3), Fibroblast growth factor (FGF23), protein-coding gene (AR1D5B) & the 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 enzyme (PFKFB3 gene). All of these genes are associated in different ways with calcification & are potential clinical targets. Macrophages promote VC by releasing inflammatory factors and phagocytosis contributing to calcium deposition in the arteries. The specific mechanism by which macrophages effect the formation of VC have been elucidated by new ideas such as switching M1 to M2 for therapeutic strategies (Yalan L., et al. 2020). Glycolysis mediated by PFKFB3 promotes Ost-Diff of vascular cells, but PFKFB3 regulation is modulated by Metformin that controls glycolysis. Thus, chronic inflammation, metabolomes & genes lead to calcification.

Session Title: Epigenetics Poster Thursday Session

Board 1034T: Multi-Ancestry Epigenetic Investigation of Generalized Anxiety Disorders in the Million Veteran Program ⭐

Authors:

P. Gupta[1], D. Levey[2]; [1]Yale Sch. of Med., West Haven, CT, [2]Yale Univ./VA Connecticut Hlth.care System, West Haven, CT

Abstract:

Background: Anxiety is a common reaction to stress which is characterized by feelings of tension and worried thoughts which may be accompanied by physical changes such as increased heart rate. When anxiety becomes excessive or chronic it can interfere with daily life and lead to anxiety disorders. Generalized Anxiety Disorder (GAD) is a specific anxiety disorder characterized by chronic excessive worry, physical symptoms such as fatigue, sleep disturbances, and irritability. GAD is underrepresented in epigenetic studies. Methods: We accessed deidentified data from electronic health records available for participants in the Million Veteran Program to identify inpatient and outpatient International Classification of Disease (ICD) codes for GAD. We used principal component analysis to stratify by genetic ancestry into groups of European (EUR), African (AFR) and Hispanic (HIS) ancestry. Using a large sample to understand the role of epigenetic regulation in GAD, we performed EWAS with EPIC array data from approx. 45,000 individuals. We performed polygenic risk scores and epigenetic risk scores in independent cohorts to identify predictive power. Results: We identified 33 methylome-wide significant (P value < 5.77e-08) differentially methylated positions associated with GAD. Some of the strongest associations were found for CpG sites near the transcription start site (TSS) of *FURIN* gene (proteolytic enzyme) on chromosome 15, multiple CpG sites near TSS and in exon of *F2RL3* gene (thrombin receptor expressed on platelets) on chromosome 19, CpG sites in exons and 3' untranslated regions of *KAZN* gene (Kazrin, periplaktin interacting protein) on chromosome 1. Conclusions: We were able to identify a substantial number of associations between GAD and methylation change. We found 33 total differentially methylated sites, including one near the gene FURIN, which has been previously associated across multiple methodologies with various mental health traits, including anxiety disorders. This study furthers our understanding of the underlying biology of generalized anxiety disorders.

Session Title: Epigenetics Poster Thursday Session

Board 1035T: Multilocus Imprinting Disturbances (MLID) analysis in a cohort of Beckwith-Wiedemann and Silver-Russell Syndromes: Comparison between specific MS-MLPA and methylation microarray

Authors:

P. Lapunzina[1], A. Parra Parra[1], J. Tenorio[2], M. Cazalla[1], L. Miranda[1], N. Gallego[1], P. Arias[1], C. Silván[1], The Spanish Overgrowth Registry Initative (SOGRI), J. Nevado[1]; [1]INGEMM-HULP-CIBERER, Madrid, Spain, [2]INGEMM-HULP-CIBERER, ADNTRO Genetics and BITGENETIC, Madrid, Spain

Abstract:

Background/Objectives: Imprinting disorders comprise a group of diseases which molecular basis is due to defects in the pattern of methylation in specific regions, which either inactivate or activate specific genes. To date, several syndromes associated with defects in the methylation pattern have been identified, including Beckwith Wiedemann syndrome and Silver-Rusell syndrome. Multilocus imprinting disturbances (MLID) represented a group of individuals with more than one locus with altered methylation pattern, which can be associated with an increased risk of having a more heterogeneous severe phenotype and in which genetic counseling in extremely useful for mothers in order to prevent for new cases. **Methods:** A cohort of 54 patients with BWS (n=99) and SRS (n=13) who already have a methylation defect at the pathology-associated locus (11p15) were analyzed. Methylation analysis was performed by two approaches: 1- MS-MLPA using MRC-Holland kit ME034. This kit allows the analysis of 11 different imprinted regions. (2) 950k methylation microarray which can detect methylation defects at whole genome scale. Also, available DNA from mothers of patients with MLID have been also analyzed by WES to identify candidate variants associated with the increased risk of having children with imprinting disorders. **Results:** 20% (15 BWS + 1 SRS) of the patients analyzed had a MLID, inclungind a variety of different regions affected in the well known imprinted loci. Addittionally, methylation microarray confirmed an abnormal pattern of methylation in order regions not previously reported, suggesting a contribution of these region in the phenotype of the patients. **Conclusion:** A significant percentage of patients with BWS and SRS syndromes have methylation alterations in more than one locus. These additional alterations may be associated with a much more heterogeneous clinical features of the patients, confirming the multilocus syndrome. There are maternal genes associated with multilocus syndromes in their offspring. A consensus should be established for the clinical and molecular testing and reporting of patients with multilocus imprinting disturbances.

Session Title: Epigenetics Poster Thursday Session

Board 1036T: Nuclear regulatory disturbances precede and predict the development of Type-2 diabetes in Asian populations

Authors:

J. Chambers[1], P. Jain[1], H. Ng[1], D. Tay[1], T. Mina[1], D. Low[1], N. Sadhu[1], F. Tai[1], A. Gupta[1], P. Elliott[2], E. Riboli[2], E. Lee[1], J. Lee[3], J. Ngeow[4], R. van Dam[5], W. Koh[6], X. Sim[7], M. Loh[1]; [1]Nanyang Technological Univ., Singapore, Singapore, [2]Imperial Coll. London, London, United Kingdom, [3]Inst. of Mental Hlth., Singapore, Singapore, [4]Natl. Cancer Ctr. Singpore, Singapore, Singapore, [5]Natl. Univ. of Singapore, National University of Singapore, Singapore, [6]Duke-NUS Med. Sch., Singapore, Singapore, [7]Natl. Univ. of Singapore, Singapore, Singapore

Abstract:

Introduction
Type 2 diabetes (T2D) is a chronic disease typically characterised by the presence of adiposity, insulin resistance and impaired glucose metabolism, with the burden of diabetes especially high in Asia. This increased risk for T2D is not fully accounted for by traditional risk factors or by genetics. Recent studies have highlighted the presence of disturbances in genome regulatory mechanisms as early manifestations of T2D risk. In this study, we aimed to advance understanding of the molecular regulatory disturbances underlying development of diabetes amongst high-risk Asian populations.
Methods We first carried out epigenome-wide association for incident T2D using baseline samples collected at enrolment for 5,709 Asian individuals, from three prospective population-based cohorts. DNA methylation was quantified in genomic DNA from whole blood, using the Illumina EPIC or 450K array. Epigenome-wide significance was set at P<8.62E-8, as determined via permutation testing. To further explore potential genetic and environmental factors influencing methylation at our Sentinel CpG sites, we used data from an independent multi-ethnic Asian population cohort with deep clinical and molecular phenotyping (n=2,342). This was supplemented by functional genomic analyses and fine-mapping via targeted sequencing of selected top loci. We then tested the potential utility of DNA methylation for prediction of T2D via a Methylation Risk Score (MRS) approach.
Results We identified 420 CpG sites across 315 independent loci to be significantly associated with incident T2D (P: 8.51E-08 to 9.08E-46), with conditional analyses identifying 324 Sentinel CpGs. We noted significant enrichment for association of the Sentinel CpGs with multiple measures of adiposity and lipid metabolism and associated polygenic risk scores (PRS). Functional genomic analyses, along with methylation quantitative trait loci (mQTL) and network analysis provided multiple levels of evidence that Sentinel CpGs identify disturbances in nuclear regulatory pathways preceding the development of T2D and related metabolic phenotypes, linking immune activation to metabolic dysregulation. We further demonstrated that MRS was as predictive for future T2D as established PRS scores, with an independent and additive relationship. Fine-mapping identified multiple additional CpG sites that were more strongly predictive of T2D than evaluation based on low-coverage microarray alone.
ConclusionOur study discovered novel methylation markers for T2D, and also allowed us to identify nuclear regulatory disturbances that precede and predict the development of T2D in Asian populations.

Session Title: Epigenetics Poster Thursday Session

Board 1038T: PTSD and Epigenetic Aging: A Longitudinal Meta-analysis

Authors:

X. Zhao[1], S. Katrinli[2], B. M. McCormick[3,4], M. W. Miller[3,4], N. R. Nugent[5], A. H. Wani[6], A. S. Zannas[7], PGC-PTSD Epigenetics Workgroup, M. W. Logue[1,3,4], C. M. Nievergelt[8,9], A. K. Smith[2], M. Uddin[6], E. J. Wolf[3,4]; [1]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [2]Emory Univ., Atlanta, GA, [3]VA Boston Hlth.care System, Boston, MA, [4]Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [5]Alpert Brown Med. Sch., Providence, RI, [6]Univ. of South Florida Coll. of Publ. Hlth., Tampa, FL, [7]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [8]Univ. of California San Diego, La Jolla, CA, [9]VA San Diego Hlth.care System, San Diego, CA

Abstract:

Posttraumatic stress disorder (PTSD) has been associated with advanced epigenetic age in cross-sectional studies, but the longitudinal association between them remains unclear. Using data from 1,367 subjects across seven military and civilian cohorts participating in the Psychiatric Genomics Consortium PTSD Epigenetics Workgroup, we conducted meta-analyses to examine if new-onset PTSD diagnosis and changes in PTSD symptom severity over time were associated with changes in epigenetic age acceleration/deceleration (age-adjusted DNA methylation age residuals per the Horvath and GrimAge metrics) over two time points via linear regressions. Meta-analysis revealed that the interaction between Time 1 (T1) Horvath age residuals and new-onset PTSD diagnosis was significantly associated with Time 2 (T2) Horvath age residuals (meta $\beta = 0.16$, meta $p = 0.02$, p-adjusted = 0.03 via an FDR correction across two age algorithms), indicating that the association between Horvath age residuals at the two time points was greater among individuals with new-onset PTSD diagnosis. The interaction between T1 Horvath age residuals and changes in PTSD symptom severity over time was also significantly associated with T2 Horvath DNAm age residuals (meta $\beta = 0.24$, meta $p = 0.05$), suggesting that the relationship between Horvath age residuals at the two time points was stronger among those with increased PTSD symptom severity from T1 to T2. No significant associations were observed for GrimAge residuals. These results implied that individuals with new-onset PTSD and those experiencing a maximum possible symptom severity change (from having no symptoms at all to having all PTSD symptoms at the most severe level) would age faster by an extra 1.9 months and 2.9 months, respectively, over the time interval. This study provided meta-analytic evidence that new-onset or worsening of PTSD over time could result in greater epigenetic age acceleration at follow-up than would be expected based on baseline age acceleration, highlighting the need for intervention studies to determine if PTSD treatment has the potential to slow the pace of epigenetic aging.

Session Title: Epigenetics Poster Thursday Session

Board 1039T: The nuclear epigenome mediates the effect of mitochondrial DNA variation on cardiovascular disease risk

Authors:

P. Win[1,2], M. Lekhi[1], J. Nguyen[1], Y-H. Choi[3], C. Castellani[1,2,3,4]; [1]Dept. of Pathology and Lab. Med., Western Univ., London, ON, Canada, [2]Children's Hlth.Res. Inst., Lawson Res. Inst., London, ON, Canada, [3]Dept. of Epidemiology and Biostatistics, Western Univ., London, ON, Canada, [4]McKusick-Nathans Inst., Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Abstract:

Mitochondria (mt) drive cellular energy production and variation of mtDNA is associated with all-cause mortality, cardiometabolic traits, and atherosclerosis, a primary contributor to cardiovascular disease (CVD). mtDNA-associated CVD risk is independent of other known associated CVD risk factors and therefore may represent unaccounted-for CVD risk. mtDNA copy number (mtDNA-CN), heteroplasmy, haplotype, and haplogroup represent mtDNA variation measures. mtDNA-CN refers to the number of mt genomes per cell, heteroplasmy is defined as the ratio of mutated to wildtype mtDNA, haplotype describes polymorphic sets of nDNA that tend to be inherited together, and haplogroups classify similar haplotypes that share a common ancestor. Recent evidence has identified associations between mtDNA variation and remodeling of the nuclear DNA (nDNA) epigenome through global metabolite variability. We assessed the extent to which nDNA epigenetic modification mediates the impact of mtDNA variation on CVD risk.

We used the Canadian Longitudinal Study on Aging (CLSA), a middle-aged prospective longitudinal human cohort with matched genomic (Affymetrix Axiom array, N=26,622), epigenomic (Infinium EPIC Microarray, N=1,479), and metabolomic (Metabolon HD4, N=9,500) data, including two longitudinal timepoints for CVD and related outcomes. mtDNA-CN, haplogroup, and haplotype were determined using the Axiom array. Following quality control, linear mixed model regression and survival analyses, adjusted for relevant covariates, were used to identify associations. Causal mediation effects were identified using the R package *mediation*. The significance of any mediation effect was tested using a bootstrapping method with 1000 simulations.

Our analysis identified 738 and 88 methylation CpG sites to be significantly associated with haplogroup and haplotype, respectively (N=1,336, $P<1x10^{-7}$). Haplogroup L was found to be associated with the development of peripheral vascular disease (PVD) (N=26,622, $P=3x10^{-2}$) and a causal mediation effect was established for the nuclear methylation site cg22944890 in PVD.

Further assessment of mtDNA features and metabolomic profiles is ongoing. Gene ontology results revealed the molecular function 'guanyl-nucleotide exchange factor' ($P=3x10^{-4}$) to be enriched in haplogroup-associated CpGs. Elucidating the mechanisms driving mtDNA variation on disease risk in the context of nuclear DNA dynamics will aid in development of approaches towards better diagnosis, treatment, and prevention of CVD.

Session Title: Epigenetics Poster Thursday Session

Board 1040T: To stop or not to stop? An epigenomic landscape of transcription termination alteration networks implies their functional role in solid cancers

Authors:

K. Chwialkowska[1], M. Niemira[1], A. Bielska[1], A. Zeller[1], M. Kuzmicki[1], J. Szamatowicz[1], A. Michalska-Falkowska[1], J. Reszec[1], B. Kedra[1], P. Knapp[2], J. Niklinski[1], M. Moniuszko[1], A. Kretowski[1], M. Kwasniewski[1]; [1]Med. Univ. of Bialystok, Bialystok, Poland, [2]Univ. Clinical Hosp. in Bialystok, Bialystok, Poland

Abstract:

Background: Readthrough transcription results from termination failure and leads to downstream transcription (DoG), as well as read-in into next genes. These transcripts play a role in regulating chromatin states and modulating gene expression. Recently, transcription readthrough was shown to modulate normal cellular functions (Caldas et al., 2024). Moreover, it was presented that DoG RNAs play a role in tumour-promoting and tumour-suppressor pathways (Avilla-Lopez et al., 2024). Still, there is a knowledge gap regarding the specific functional role and the extent of epigenetic regulation of readthrough transcription during carcinogenesis. *BRD4*, a master regulator of transcription elongation, is a new cancer therapeutic target (Cameron et al., 2023). The study aimed to evaluate the relationship of readthrough transcriptome with DNA methylome (DNAm) profile and assess its significance in oncogenic pathways.
Methods: Stranded total RNA-Seq and genome-wide DNA methylation profiling (RRBS) were performed in our internal non-TCGA-based specimens of solid tumours and matched non-neoplastic tissues bio-banked as fresh samples under rigorous procedures. Both analyses were performed on the same specimens. Readthrough transcripts were detected using a sliding window algorithm. CpG methylation profiles were stratified into differential regions using circular regularization procedures. Enrichment analyses were calculated under hypergeometric distribution. All statistical tests were under $\alpha=0.05$ with False Discovery Rate (FDR) multiple

testing correction.
Results: In total, the project covers >2000 transcriptome and methylome profiles from >500 patients with ovarian, colon, and lung cancer, as well as glioma. Readthrough levels were related to the specific immune cells infiltration. Downregulation of DoG was associated with altered CpG methylation (P<1.0×10^-3). The up- and down-regulation of read-in genes was more severely affected by altered DNAm profiles (P<1.0×10^-50). For e.g., in colorectal cancer, upregulated read-in genes exhibiting DNA demethylation formed strong functional networks (P<1.0×10^-4) extending to the top-3: *PLK1*, *HJURP*, *CDK1*, as well as *APC*. There was a significant over-representation of processes related to mitotic division, differentiation, and immune responses.
Conclusion: Obtained data suggest that readthrough transcription aberration in solid cancers is associated with altered DNA methylation level in a non-stochastic manner and might play a biologically significant role during carcinogenesis, affecting differentiation, loss of cell identity, and driver/suppressor-related pathways.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3002T: A Diversity-First Approach to Understanding Primary Open-Angle Glaucoma

Authors:

L. Cruz[1], T. G. Kinzy[2], B. Gorman[3], Y. Bradford[4], C. Jiang[5], A. Waksmunski[1], R. Salowe[6], C. Halladay[7], C. Nealon[8], Million Veteran Program, L. Zangwill[9], R. Ayyagari[10], R. Weinreb[9], M. Ritchie[4], S. Verma[11], H. Choquet[12], J. O'Brien[13], M. Hauser[14], D. Crawford[1], S. Iyengar[1], N. Peachey[8], J. Cooke Bailey[15]; [1]Case Western Reserve Univ., Cleveland, OH, [2]East Carolina Univ., Greenville, NC, [3]Booz Allen Hamilton, Lexington, MA, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Kaiser Permanente Northern California, Div. of Res., Pleasanton, CA, [6]Univ. of Pennsylvania, Marlton, NJ, [7]Providence VAMC, Providence, RI, [8]VA Northeast Ohio, Cleveland, OH, [9]UC San Diego Hlth., San Diego, CA, [10]Univ California San Diego, San Diego, CA, [11]Univ. of Pennsylvania, Cherry Hill, NJ, [12]Kaiser Permanente - Div. of Res., Oakland, CA, [13]Univ. of Pennsylvania Perleman Sch. of Med., Philadelphia, PA, [14]Duke Univ Med Ctr, Durham, NC, [15]East Carolina Univ. Brody Sch. of Med., Greenville, NC

Abstract:

Introduction Genetic risk for the leading cause of irreversible blindness, primary open-angle glaucoma (POAG), is overwhelmingly studied in European (EUR) populations. POAG

disproportionately affects African (AFR) and Hispanic (HIS) populations, e.g. earlier onset and worse clinical outcomes. Vision-preserving treatment requires early detection in high-risk individuals, which necessitates a better understanding of POAG genetic architecture, critical to ensuring vision health equity.
Objectives The persistent paucity of POAG research in non-EUR is alarming; our study prioritizes underrepresented populations with highest POAG disease burden to examine POAG genetics in diverse groups. We accessed the Million Veteran Program (MVP), a large ancestrally diverse national cohort of United States Veterans to perform genome-wide association (GWA) and genetic admixture-informed (GAI) analyses. We meta-analyzed GWAS with other POAG datasets.
Methods We applied a validated EHR-based computable phenotype algorithm to MVP participants to define POAG cases and controls with TOPMed-imputed genetic data of 11933 AFR-ancestry (5099 cases, 6834 controls) and 4589 HIS-ancestry (885 cases, 3704 controls) individuals. We conducted ancestry-specific single-variant GWA studies (GWAS) of POAG adjusted for age, sex, and 10 principal components. We also performed admixture analysis via TRACTOR in MVP AFR. We performed ancestry-specific fixed-effects meta-analysis in AFR (8385 cases, 13160 controls) from four studies: MVP, Genetic Epidemiology Research in Adult Health and Aging [GERA], Duke African American POAG, and Primary Open-Angle African American Glaucoma Genetics [POAAGG]) and in HIS (1296 cases, 8482 controls) from MVP and GERA.
Results Three loci reached genome-wide significance (GWS; P<5.0x10-8) in the MVP AFR GWAS: CDKN2B-AS1, CHEK2, and IGFL2. CHEK2 and IGFL2 were also GWS in TRACTOR. Known POAG loci were GWS in AFR meta-analysis: ABCA1, CDKN2B-AS1, CHEK2, DGKG, XRCC6P1. In MVP HIS GWAS, no variants attained GWS; however, when meta-analyzed with GERA, TMCO1 and CDKN2B-AS1 attained GWS for the first time in HIS.
Conclusion We report the first POAG genetic discovery analyses in a biobank dataset of consistently-phenotyped AFR and HIS, and the largest AFR and HIS sample sizes in POAG genetic meta-analyses. We confirm known POAG loci. Though CDKN2B-AS1 and TMCO1 were previously identified in EUR, our phenotype definition did not require severe vision loss to detect GWS in HIS. Our study underscores the essential role of biobanks, consistent phenotyping, and diversity in enhancing our understanding of the genetic factors contributing to POAG.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3003T: A pangenome approach to study short read data for the population genetics of HLA ⭐

Authors:

O. Garcia[1], E. Ho[1], A. Guarracino[2], E. Garrison[2], O. Cornejo[1]; [1]Univ. of California Santa Cruz, Santa Cruz, CA, [2]Univ. of Tennessee Hlth.and Sci. Ctr., Memphis, TN

Abstract:

The human leukocyte antigen (HLA) or major histocompatibility complex (MHC) region on Chromosome 6 is the most crucial immune-related area of the human genome, mediating adaptive immune responses by presenting antigens to T lymphocytes. Given their essential role in presenting foreign peptides to the immune system, their study is important to understand how different people respond to infectious diseases, develop autoimmune diseases, and respond to organ transplantations. The MHC is the most diverse region of the genome with over 35,000 classical HLA alleles recognized. This genetic diversity, however, poses challenges to the characterization of contiguous haplotypic genetic diversity while using standard short read sequencing techniques that map to linear reference genomes. We anticipate that the HLA is a great candidate to show how the development of multi-genome reference panels can greatly contribute to characterizing the haplotypic diversity of complex regions. We introduce a graph genotyping method to explore HLA region structural variation using 4,407 global individuals with 30X genome sequences from the Human Genome Diversity Project, Simons Genome Diversity Project, and the 1000 Genomes Project. We extracted the reads from the BAM file separately, from the following regions: the chromosome 6, the alternate contigs, and unmapped regions. We modified the Human PanGenome Reference Graph to include long read sequences from The 1000 Genomes Project ONT Sequencing Consortium to improve the mapping quality for highly divergent sequences. After the genotyping, we examined the structural variation in the MHC region across different populations using various statistical packages. We further evaluated the approach by assessing, in admixed individuals, how admixture impacts the locus and whether there was an under or overrepresentation of specific ancestry in the MHC region when compared to global admixture levels. Lastly, we built phylogenetic trees to determine the relationships between populations in the HLA regions. Our developments for better mapping practices for the HLA region show a drastic improvement in mapping which will aid in discovering novel variation of the HLA region, which will improve the understanding of the population genetics of the region. Future applications of this will not only improve our knowledge of the region for studying immune related functions, but also aim to improve accuracy for HLA typing, improving transplant outcomes. Our approach represents a framework to improve the mapping of other complex regions such as the Killer Immunoreceptor Cell region when working with short reads.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3004T: Australian Indigenous Population Structure and Its Impact on Unique Genetic Variation

Authors:

A. Farlow[1,2], M. Silcocks[2], S. Leslie[2]; [1]Australian Natl. Univ., Canberra, Australia, [2]The Univ. of Melbourne, Parkville, Australia

Abstract:

The Indigenous peoples of Australia have a rich linguistic and cultural history. However, how this is reflected in genetic diversity remains largely unknown due to their limited engagement with genomic studies. Working with communities under a novel system of Indigenous governance, we analysed the genomes of 159 individuals from four remote Indigenous communities including, for the first time, people who speak a language other than from the most widespread language family in Australia.

Our analysis reveals strong population structure, indicating deep divergence times between communities. For instance, the populations of the Tiwi Islands and Elcho Island in Northern Australia have been genetically isolated for around 35 thousand years, despite only 500 km (300 miles) of current geographic separation.

Beginning around 50,000 years ago and ending during the last glacial maximum approximately 20,000 years ago, each Indigenous population experienced a dramatic increase in effective population size. This was followed by a rapid decrease and then stable but relatively small effective population sizes throughout the last 10,000 years.

This dynamic history, coupled with early divergence from Eurasian groups, has generated the highest proportion of novel genetic variation seen outside of Africa. Remarkably, one-quarter of these variants are absent from global reference panels, including gnomAD. This is crucial because rarity in gnomAD is a metric used to prioritise potentially pathogenic variants during clinical diagnostics.

Our findings underscore the necessity of including Indigenous Australian populations for a more complete accounting of human genetic variation, and that considering the complex ancestry of the Australian population is critical during both medical research and personalised medicine.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3005T: Characterizing evolutionary constraint of different time scale on genome function and diseases

Authors:

J. Wang[1,2,3], E. M. Koch[4,5], S. K. Reilly[6,7], S. Sunyaev[4,5], S. Gazal[1,2,3]; [1]USC, Los Angeles, CA, [2]Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [3]Ctr. for Genetic Epidemiology, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [4]Harvard Med. Sch., Boston, MA, [5]Dept. of BioMed. Informatics, Harvard Med. Sch., Boston, MA, [6]Yale Univ., New Haven, CT, [7]Dept. of Genetics, Yale Univ., New Haven, CT

Abstract:

Measuring evolutionary constraints across the genome allows us to quantify the functional importance of genomic regions by inferring if positions have changed more slowly than expected under neutral drift due to negative selection. Recent studies have estimated constraints in mammals and primates (through phylogenetic cross-species comparison), as well as in humans (through genetic diversity comparison across thousands of humans), allowing us to quantify the strength of negative selection at different time scales. Characterizing constrained regions of the human genome has the potential to identify which biological processes and phenotypes were impacted by different types of constraints, and help interpret non-coding genome function by identifying bases that are more likely to be functional and impact diseases. Here, we compared estimates of constraint in mammals, primates and humans from various sources. Despite low correlations between the scores, we observed that constrained positions identified by all metrics were enriched in relevant functional regions (e.g. exons, UTRs, promoters), informative for predicting pathogenic variants in ClinVar and Expression Modulating Variants in MPRA analyses, and enriched in disease heritability. Mammal-constrained variants were more informative for classifying ClinVar pathogenic variants (dominated by coding SNPs), primate-constrained variants were more enriched in disease heritability (dominated by non-coding SNPs), and human-constrained variants had consistently lower enrichments. These findings suggest that distinct biological processes have been under evolutionary constraint at different time scales.

To characterize those processes, we first estimated the fraction of constrained bases within scATAC-seq peaks from 222 cell types in fetal and adult tissues (Zhang et al. 2022 Cell). We observed that fetal tissue peaks tended to be more constrained overall, peaks from immune cell types more constrained in humans, and those from brain cell types more constrained in primates. Gene constraint estimates from their exons showed that genes more constrained in humans tend to have immune functions. We found no enrichment in specific pathways for genes differentially constrained between mammals and primates, suggesting that selective pressure related to primates operated primarily on the non-coding genome.

Altogether, our results highlight that genetic constraint has targeted different biological processes at different time scales. In addition, immune function enrichment within human constrained regions may indicate that human constraint also captures footprints of positive selection.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3006T: Confirmatory testing for Hemoglobin Disorders Included in the Philippine Newborn Screening Program

Authors:

M. Abadingo, **M. Alcausin**, J. Caneba, E. Maceda, C. Padilla; Univ. of the Philippines Manila, Manila, Philippines

Abstract:

Background. Hemoglobin disorders have been included in the Philippine expanded newborn screening (ENBS) panel since 2014. The inclusion of hemoglobin disorders in the panel is based on the extrapolation of Filipino data from the California Newborn Screening Program (CNSP). High-performance liquid chromatography (HPLC) is used for screening. Confirmatory tests for hemoglobin disorders have evolved since its inclusion. Capillary electrophoresis (CE) and molecular tests are utilized as confirmatory tests for hemoglobin H (Hb H) disease. At present, only molecular tests are utilized for confirmation of Hb H disease. CE was a confirmatory test for beta-thalassemia and other clinically significant hemoglobin variants. Eventually, molecular test was added as a confirmatory test for clinically significant hemoglobin disorders other than Hb H disease. The objectives of the study are 1.) to describe the transformation of the confirmatory testing for the hemoglobin disorders included in the ENBS, and 2.) to review the hemoglobin disorders confirmatory test results from 2014-2023. **Methodology and Results**. A review of algorithms for screening and confirmation of hemoglobin disorders was done. Data on confirmatory tests for hemoglobin disorders were collected from seven (7) newborn screening centers (NSCs) in the Philippines. Of 3,181 cases who screened positive for hemoglobin disorders, 2,929 (92.07%) were confirmed cases of Hb H disease. Of the 2,929, 2,911 (99.4%) have a deletional mutation. The second most common confirmed disorder is Hemoglobin E disease, with 145 cases (4.56%), followed by beta thalassemia major (1.98%). **Conclusion**. Confirmatory testing is needed to establish the diagnosis of hemoglobin disorders initially detected through newborn screening. Molecular diagnosis plays a crucial role in the diagnosis and management of hemoglobin disorders, specifically in the genetic counseling of cases.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3007T: Efficient adaptive target detection and characterization with minimal training by transfer learning

Authors:

S. Arnab, P. Sigler, A. Campelo dos Santos, M. DeGiorgio; Florida Atlantic Univ., Boca Raton, FL

Abstract:

Natural selection leaves detectable patterns of altered spatial diversity within genomes, and identifying affected genomic regions is crucial for understanding species and population evolution. Recent approaches to uncovering adaptive footprints leverage machine learning techniques applied to raw population genomic data. Convolutional neural networks (CNNs) are particularly effective for this task due to their ability to handle large input arrays and maintain correlations among elements. However, CNNs require large training datasets and are computationally intensive to train. Shallow CNNs, while simpler and less computationally demanding than deep CNNs, may fail to capture the complex patterns associated with the many factors that influence genomic variation. Deep CNNs, on the other hand, can capture these intricate patterns, but at the cost of requiring extensive training data and computational resources. To address these challenges, transfer learning can be employed, which conditions on a CNN pre-trained on a large, diverse dataset, that is then fine-tuned on a smaller, task-specific dataset. This approach reduces the need for extensive training data and improves training efficiency while maintaining high model performance. In a transfer learning architecture, the pre-trained deep CNN acts as a feature extractor, leveraging learned representations from a larger dataset to inform the task-specific model. In this study, we develop TrIdent, which uses transfer learning to enhance the detection and characterization of adaptive genomic regions from image representations of multilocus variation. We assess the performance of TrIdent under an array of genetic, demographic, and adaptive settings, as well as unphased data and other confounding factors. Our results demonstrate that TrIdent has significantly improved detection of adaptive regions compared to some recent approaches that operate on similar data. Additionally, we explore the interpretability of the TrIdent model through class activation maps, and we retool TrIdent to infer selection parameters underlying identified adaptive candidates. Using whole-genome haplotype data from Europeans and Africans and TrIdent trained to detect selective sweeps, we were able to recapitulate well-established sweep candidates and predict novel genes associated with cancer as highly supported sweeps. These results highlight the potential of TrIdent to contribute valuable insights into adaptive processes from genomic data.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3008T: Evaluating Functional and Evolutionary Genomic Predictors of Polygenic Score Portability Using Machine Learning and Explainable AI

Authors:

U. Hazra, J. Lachance; Georgia Inst. of Technology, Atlanta, GA

Abstract:

The portability of polygenic predictions across populations remains a significant challenge due to factors such as differences in allele frequency, linkage disequilibrium, and gene-by-environment interactions. Although these oft-mentioned explanations are supported by theory, they have not been examined at scale using empirical data. Here, we enhance the transferability of polygenic scores by leveraging machine learning to identify functional and evolutionary genomic features that contribute to whether trait associations are consistently replicating across populations. Utilizing large-scale biobank data (e.g., UK Biobank and BioBank Japan), we developed a machine learning model to evaluate how different features of SNPs impact the transferability of polygenic scores. For each SNP, genetic variance contributions and p-values from ancestry-specific GWAS were used to quantify how well associations replicate across populations, creating a range of class labels for classification tasks. These SNPs were then annotated using functional (e.g., GC content, local recombination rate, and DNAse I sensitivity) and evolutionary genomic (e.g., allele age, GERP, Fst, XP-EHH, and iHS) statistics to create a comprehensive feature matrix. We then trained a random forest model on these features and validated using accuracy metrics and hyperparameter tuning. While a good prediction model can help identify SNPs that will transfer across populations, it is also essential to know which features drive the transferability of genetic predictions. Thus, explainable AI techniques, specifically SHAP (SHapley Additive exPlanations) values, were employed to interpret the model's findings. This method allowed us to quantify the contribution of each functional and evolutionary feature to the model’s predictions. By interpreting SHAP values, we identified genomic features that significantly impact whether GWAS results replicate across populations. These features were then used to generate predicted replication scores, enabling researchers to pre-assess the transferability of a polygenic risk scores.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3009T: Evolutionary dynamics of intrinsically disordered regions of human transcription factors

Authors:

S. Song, J. M. Akey; Princeton Univ., Princeton, NJ

Abstract:

Transcription factors (TFs) regulate gene expression by binding to DNA and recruiting cofactors through their activation domains, which are often intrinsically disordered regions (IDRs). IDRs lack a fixed three-dimensional structure and instead rely on the biochemical properties of their sequence for molecular interactions, some of which have been shown to drive phase separation. TFs are especially enriched for IDRs, which may facilitate the formation of transcriptional condensates. We identify IDR regions in human TFs using MetaPredict and find that disorder content is correlated with protein-protein interaction network size, possibly due to the increased ability of TFs with longer IDRs to recruit more cofactors. Gene ontology enrichment analysis reveals that highly ordered TFs are enriched in immune response pathways, while highly disordered TFs are enriched in cell fate pathways. We explore the evolutionary forces acting on TF IDRs and the functional effects of genetic variants within these regions using the gnomAD (v4.1) and 1000 Genomes (Phase 3) datasets. We observe significantly higher nucleotide diversity in IDRs and find evidence that higher recombination rates may be one force contributing to this increased genetic diversity. We also observe that the ratio of non-synonymous to synonymous evolution (dN/dS) is elevated in IDR regions. Furthermore, we performed McDonald-Kreitman tests and found that positive selection dominates in IDR regions, while negative selection dominates in ordered regions. Furthermore, missense variants in TF IDRs are less deleterious than those in structured regions, as indicated by metrics such as CADD, REVEL, SIFT, ClinVar, and HGMD. In conclusion, a distinct evolutionary regime of positive selection and relaxed purifying selection operates in TF IDRs, which may play a crucial role in TF adaptability. Our study demonstrates that accounting for disorder and structure is critical when considering the evolution of TFs.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3010T: Founder variants in the Mexican population: a Systematic Review

Authors:

T. Metta-Harari[1], S. Harari-Arakindi[1], I. Espino-Gutiérrez[2], L. Taja-Chayeb[3], J. Garcia-Ortiz[4], T. Wegman-Ostrosky[3]; [1]Univ. Anahuac, Mexico City, Mexico, [2]Univ. Natl. Autónoma de México, Mexico City, Mexico, [3]Inst. Natl. de Cancerologia, Mexico City, Mexico, [4]CIBO IMSS, Guadalajara Jalisco, Mexico

Abstract:

Founder variants are defined as genetic alterations, at DNA or chromosomal level, that are observed with high frequency in genetically homogeneous populations, in which a common ancestor was a carrier for the variant. Founder variants significantly influence the prevalence of genetic disorders in specific populations. Due to the high costs of comprehensive genetic testing, testing for specific founder variants is a practical diagnostic tool for the Mexican population. This systematic review aims to summarize and describe the Mexican founder variants and their association with specific health conditions. A systematic literature search in LILACS, COCHRANE, Scopus and PubMed was conducted. Information about the genes, variants and haplotypes was extracted from the papers that meet the inclusion criteria. Based on the evidence provided, variants originating in the Mexican population were classified into two categories: strong or mild evidence of founder variant. Founder variants originating outside of Mexico, found among the Mexican population were classified into another category. Altogether, 41 studies were selected, describing 26 genes and 28 founder variants.The associated genes and founder variants described are the following: BRCA 1 c.548-?_4185+?del, G6PC3 c.210del, FANCG c.511-3_511-2delCA, PSEN1 c.1292C>A, KRIT 1 c.1363C>T, LDLR c.1078G>C and c.2271del, MC4R c.806T>A, FKRP c.1387A>G, NLRP7 c.2248C>G and c.2810+2T>G, GCDH c.389T>C, FKRP c.1387A>G, ATP6V0A4 c.1231G>T, VHL c.298-299insA, BLM c.2506_2507del, SRD5A2 c.635C>G, GAA c.1987del, TGM1 c.1054C>G, EVC c.1678G>T, PAX6 c.732-735delAACA, PABPN1 (GCG)11 or (GCG)9 expanded alleles, F9 c.881G>A, ATXN7 c.89AGC[233], PANK2 c.881A>T, CYP1B1 c.1159G>A, PAH c.60+5G>T, USH2A c.2299delG. Haplotype analysis revealed that although some variants are frequent in the Mexican population, their origin might be European and brought to Mexican territory around the Spanish conquest. However, it is still important to screen for these variants among the Mexican population, as a large portion of the population is Mexican-Mestizo.The knowledge of founder variants in a population allows for an understanding of genetic architecture, which has clinical, historical, and cultural implications. This is the first systematic review describing founder variants within the Mexican population.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3011T: Functional and evolutionary relationships of chitinase-like protein YKL-39 with chitotriosidase

Authors:

F. Oyama[1], T. Kanaizumi[2], K. Suzuki[1], K. Okawa[1], E. Tabata[1]; [1]Kogakuin Univ., Tokyo, Japan, [2]Kogakuin Univ., Hachioji, Japan

Abstract:

Chitinase-like proteins (CLPs) are structurally similar to chitinases but lack chitin-degrading activity. YKL-39 (also known as CHI3L2) is a type of CLP elevated in various inflammatory diseases, including osteoarthritis (OA), rheumatoid arthritis (RA), glioblastoma, multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), Alzheimer's disease (AD), and certain cancers. Consequently, YKL-39 has garnered attention as a potential biomarker for these diseases. Understanding the pathophysiological role of YKL-39 could provide significant insights into disease mechanisms and therapeutic targets. However, the precise role of YKL-39 remains unclear due to its lack of chitinase activity. YKL-39 is located on chromosome 1, near the genes for active chitinases, chitotriosidase (CHIT1), and acidic chitinase (CHIA), yet the evolutionary and functional relationships among these proteins are not fully understood. This study investigates these relationships by creating chimeric and mutant forms of YKL-39 and CHIT1. We engineered mutants, including one with a catalytic motif mutation (S138D, I140E) named MT-YKL-39, and additional mutants with CHIT1-type (MT-YKL-39-V137L) or CHIA-type (MT-YKL-39-V137F) substitutions. Contrary to expectations, further mutations in the catalytic motif did not activate the enzyme, suggesting that YKL-39's inactivation involves multiple factors beyond the catalytic motif. Further analysis of chimeric forms with CHIT1 indicated that amino acid substitutions in regions encoded by exons 4, 6, 7, and 8, along with the two amino acid mutations in the catalytic motif, play a crucial role in the inactivation of YKL-39. Therefore, the loss of chitinase activity in YKL-39 is attributed to a combination of these specific amino acid substitutions.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3013T: Functional impact of 2.7M structural variants across global populations

Authors:

X. Zhao[1,2], E. Pierce-Hoffman[2,1], R. Collins[3], M. Walker[2,1], A. Lee[1,2], A. Yenkin[4], N. Kurtas[1,2], L. Wang[4], D. Ben-Isvy[4], P. W. Darnowsky[2], K. Chao[2], S. Gabriel[2], E. Banks[2], A. O'Donnell-Luria[2], H.

Rehm[1,2], B. Neale[1,2], M. Daly[1,2], K. Samocha[1,2], K. Karczewski[1,2], H. Brand[1,2], M. Talkowski[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Dana-Farber Cancer Inst., Boston, MA, [4]Harvard Med. Sch., Boston, MA

Abstract:

Structural variants (SVs) involve rearrangements of >50 bp of the human genome and contribute to myriad human diseases and traits. Here, we present an atlas of ~2.7M SVs across 63,046 unrelated individuals in the Genome Aggregation Database (gnomAD) and 97,940 individuals in the All of Us Research Program (AoU). We applied GATK-SV on short-read whole genome sequencing (WGS) and discovered ~11K SVs per individual genome across diverse SV types: deletions, tandem duplications, insertions, inversions, multiallelic copy number variants, reciprocal translocations, and complex rearrangements. We benchmarked these data against 112 matched long-read WGS and >12K matched array samples, achieving >96% precision. On average, we observed SVs causing loss-of-function (LoF) alteration in ~188 genes, full genic copy gain (CG) in ~34 genes, and copy gain of exons in ~24 genes per genome.Expanding beyond coding-based sequences, we examined the potential functional regulatory impact of SVs by performing an unbiased noncoding combinatorial association selection (NCAS) model that evaluated unique combinations of SV types and noncoding annotations for purifying selective pressure. The SVs of each type were annotated with 68 noncoding functional categories, 16 noncoding genic elements defined in GENCODE, and their overlap with 8 conservation metrics, resulting in 198,720 unique categories. Using a 5% false positive rate threshold, we assessed purifying selective pressures on SVs within each category estimated by the adjusted proportion singleton (APS) of SVs after correcting for SV type, size, and genomic location using a logistic regression model. We then generated a null distribution of APS derived from 1,000 genome-wide permutations and established a Bonferroni correction using counts of effective tests calculated from eigen decomposition (n = 12,473) of the correlation between permutations. From these analyses, 616 categories have significantly higher than permuted APS (p value < 4.01e-6). Among these loci, we observed specific combinations of noncoding SVs and annotations experiencing reasonably strong intolerance to SVs, including deletions overlapping conserved DNase I hypersensitive sites and simultaneously altering TAD boundaries and lncRNAs (APS = 0.2, p value = 3.18e-6) that could play previously unappreciated roles in genome function. Analyses to further classify the 616 categories and identify common factors that drive global significance are actively ongoing.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3014T: Genetic ancestry reveals European population contribution in Brazilian communities of African descent in Central Brazil

Authors:

S. Oliveira[1], S. Paiva[2], A. Rivara[3], M. Nobrega[1], M. Klautau-Guimarães[1], S. Batista dos Santos[4], L. Madrigal[5]; [1]Univ.e de Brasilia, Brasília, Brazil, [2]Inst. Federal de Educação, Ciência e Tecnologia do Tocantins, Araguaina, Brazil, [3]Dept. of Chronic Disease Epidemiology, Sch. of Publ. Hlth., Yale Univ., Yale University, CT, [4]Laboratório de Genética Humana e Médica, Univ.e Federal do Pará, Belem, Brazil, [5]Univ South Florida, Tampa, FL

Abstract:

Brazil boasts a well-known multiethnic admixed population. Throughout history, Sub-Saharan Africans were brought to Brazil as slaves, contributing significantly to its ethnic diversity. Runaway or abandoned slaves established isolated communities known as quilombos, representing resilience and autonomy. Despite their prevalence in diverse regions, quilombos in Central Brazil have received relatively less attention due to geographic isolation, economic constraints, and communication barriers. Our study aims to shed light on the genetic ancestry of three quilombo communities in Central Brazil, offering insights into their unique histories. Conducted as a cross-sectional population-based investigation, our study focused on Cocalinho (N=54 individuals) and Pe do Morro (N=58 individuals) communities in Tocantins, and Kalunga (N=132 individuals) in Goias. Utilizing data from 61 biallelic INDEL genetic markers and employing STRUCTURE v2.3 software for analysis, alongside statistical analyses conducted via SAS v9.4, we aimed to estimate genetic differentiation (FST), genomic ancestry, and other pertinent population genetics parameters. While conventional racial classifications may imply varying African ancestry levels, our findings reveal nuanced patterns within these mixed populations. Notably, Cocalinho exhibited an average admixture estimate of 36.75% African, 20.86% Amerindian, and 42.38% European ancestry. In contrast, Pe do Morro showed 29.82% African, 18.49% Amerindian, and 51.69% European ancestry, while Kalunga displayed the highest African contribution at 63.17%, alongside 10.24% Amerindian and 26.59% European ancestry. Populations closer to urban centers displayed higher European ancestry, underscoring the impact of historical and geographical factors on genetic composition. These findings underscore the heterogeneous nature of Brazilian Afro-derived communities, each shaped by unique demographic processes and interactions with their environment and neighboring populations. By unraveling the genetic tapestry of Central Brazil's quilombos, our study contributes to a deeper understanding of Brazil's intricate social and historical landscape.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3015T: Genetic similarity groups detected by Identity-By-Descent clustering reveals disparities in disease prevalence in a Quebec population-based biobank ⭐

Authors:

J. Pelletier[1], V. Chapdelaine[2], D. Taliun[2], C. Bhérer[2]; [1]McGill Univ., Montréal, QC, Canada, [2]McGill Univ., Montreal, QC, Canada

Abstract:

Scalable Identity-by-Descent (IBD) clustering methods are emerging as robust alternatives to traditional Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) for identifying subgroups of individuals based on genetic similarity. While IBD captures recent genetic sharing, including rare variants, PCA predominantly reflects common variants with higher frequency differentiation between population subgroups. In this study, we use IBD clusters to characterize variation in disease prevalence and control for population stratification in single-variant association tests. We used the prospective CARTaGENE population-based cohort, comprising 29,330 participants aged 40 to 60 years recruited from urban centers in Quebec, Canada, with genomic and phenotypic data. Using genotyping array data (n=356,441), we identified IBD segments shared between the participants. We utilized iterations of the Louvain clustering algorithm in combination with Hudson's FST to delineate 21 genetic clusters, reflecting the diverse ancestries of Quebec populations, including 13 clusters more genetically similar to Europeans from the 1000 Genomes Project. These clusters varied in size, ranging from 371 to 7,474 participants, representing the French Canadian majority and well-known smaller diaspora populations in Quebec with significant 20th-century migration histories. IBD clustering revealed regional subgroups within the French-Canadian population, mirroring established fine-scale population structures related to historical migration patterns and founder effects. We estimated the prevalence of 65 well-studied diseases and traits. Significant differences in disease prevalence among clusters were identified for high blood cholesterol, gallbladder disorder and type 2 diabetes, adjusting for age, sex, and cluster size. We conducted GWAS within a European genetic ancestry subset (N = 24,991) using IBD clusters, age and sex as covariates. The analysis confirmed known associations, such as the well-documented association between *TCF7L2* and Type 2 Diabetes. In future analyses, we will investigate IBD segments and sharing within the regions amongst the top GWAS hit, looking for pathogenic variants in coding regions and their enrichment in IBD clusters. These findings underscore the utility of IBD clustering for elucidating genetic structure and its impact on disease prevalence, providing a complementary approach to traditional methods in genetic epidemiology.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3016T: GenLib.jl: Unveiling fine population structure through large-scale kinship coefficients' visualization

Authors:

G-P. Morin[1], C. Moreau[1], A. Barry[2], S. Girard[1]; [1]Université du Québec à Chicoutimi, Chicoutimi, QC, Canada, [2]Inst. Natl. de la recherche scientifique, Laval, QC, Canada

Abstract:

The population of Quebec, Canada, is known for its founder effect originating mostly from French immigration since the 17th century. This effect is particularly pronounced in the Saguenay—Lac-Saint-Jean (SLSJ) region of Quebec, known for its very rapid population growth from a limited number of founders and for the higher prevalence of certain hereditary diseases. To uncover hidden structures within such populations and reveal finer-scale structure, we have visualized kinship coefficients among individuals. These coefficients represent the likelihood that two randomly chosen alleles are identical by descent from a common ancestor. However, calculating kinship coefficients across large populations can be computationally expensive using traditional methods. To address this, we developed GenLib.jl, a Julia package featuring a new algorithm that efficiently computes exact kinship coefficients from large pedigrees in terms of both time and memory usage. Building upon previous approaches using parallel recursion or iterative memoization, our implementation adopts a hybrid strategy. This allows efficient computation of millions of pairwise coefficients per second using multithreading and moderate RAM. Using genealogical data for the province of Quebec and SLSJ from the BALSAC population register—that includes all individuals married in Quebec since the beginning of the European colonization until the 1960s—we obtained symmetrical kinship coefficient matrices. For the approximately 76,000 individuals who married in SLSJ between 1935 and 1960, the matrix contains almost 3,000,000,000 pairwise kinship coefficients that were computed within a dozen of minutes. To visualize this high-dimensional data, we employed PaCMAP, a novel dimensionality reduction technique. This visualization successfully identified clusters of individuals sharing a common demographic background, such as the same region or municipality of origin, or a similar genetic contribution from founding ancestors. Notably, this is the first time genealogical kinship coefficients have been calculated on such a large scale, and the first time we can observe the fine structure of the whole SLSJ population. The observed population stratification suggests that the studied populations are more heterogeneous than previously thought. This heterogeneity could have implications for the distribution of genetic diseases found in the region.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3017T: GenoBERT: A full-attention small language model for enhanced genotype imputation

Authors:

L. Huang[1,2], K-j. Su[1], C. Qiu[1], H. Shen[1], H. Hong[3], P. Gong[4], H-W. Deng[1], C. Zhang[2]; [1]Tulane Univ., New Orleans, LA, [2]Univ. of Southern Mississippi, Hattiesburg, MS, [3]Div. of Bioinformatics and Biostatistics, Natl. Ctr. for Toxicological Res., US Food and Drug Admin., Jefferson, AR, [4]Environmental Lab, U.S. Army Engineer Res. and Dev. Ctr., Vicksburg, MS

Abstract:

Genotype imputation, which involves filling in missing genetic data, facilitates a more comprehensive analysis across various genomic studies. Previous imputation methods traditionally relied on high-quality reference panels consisting of individuals from the same ancestral groups as the target samples for imputing common or less common variants. However, the high dimensionality of genotype datasets presents substantial challenges. Recent deep learning methods, especially Convolutional Neural Networks (CNNs) based models, although don't rely on reference panel data, often fall short in imputing datasets with a high proportion of non-homozygous reference alleles. To address this, we present GenoBERT (Genotype Bidirectional Encoder Representations from Transformers), a novel, full-attention-based small language model specifically designed for reference-free genomic data imputation. This model is built on the efficient weight-sharing strategy of the ALBERT (A Lite BERT) model, optimized for the diversity of genetic data. GenoBERT is structured with 12 encoder sublayers, incorporating a multi-head attention layer with pre-normalization and a fully connected neural network layer. Each sublayer contains 6 attention heads and utilizes rotary position encoding to integrate essential positional information in genetic sequences. By treating genotypes as sequences, we developed a novel segmentation approach to reduce dimensionality and augment dataset size, which facilitates the efficient training of our model on SNP segments. When tested on the GEUVADIS (Genetic European Variation in Disease) genotype dataset comprising 373 Caucasian samples, GenoBERT demonstrated superior performance comparing to a sparse convolutional denoised autoencoder model as the baseline, with an impressive 94.75% concordance rate and a 97.43% R^2 score, significantly exceed the baseline model's 88.89% concordance rate and 92.30% R^2 score, exclusively on missing SNPs. Beyond genotype imputation, GenoBERT shows promising potential in genotype representation modeling for downstream trans-omics and multi-omics profile prediction tasks. Plans are underway to extend GenoBERT to our in-house LOS dataset to further its utility in analyzing genetic influences on complex traits and diseases.

---PAGEBREAK---

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3018T: GrafAnc: Enhanced Genetic Ancestry Inference at Subcontinental Levels

Authors:

J. Jin, L-S. Wang, W-P. Lee; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Accurate inference of genetic ancestry from human genetic data is an essential aspect in many fields such as genetic association and human evolution studies. While principal component analysis (PCA) is currently the most widely used ancestry inference approach, its results may not be reproducible, reliable or generalizable when applied on different datasets. We developed GrafAnc, a novel approach that infers genetic ancestry at subcontinental levels quickly and robustly. GrafAnc is an advanced version of GRAF-pop. Compared with GRAF-pop, which can only cluster subjects at continental level, GrafAnc utilizes more statistical scores to infer subcontinental population structure. Using genomic data of some subcontinental populations from publicly accessible databases like the 1000 Genomes Project as references, GrafAnc calculates new statistical scores, such as EU1, EU2, EU3 (for Europeans), and EA1, EA2, EA3 (for East Asians), to determine more detailed population structure. For each continental population, several pairs of subcontinental populations were carefully selected from the references, and for each population pair, GrafAnc calculates the differences between genetic distances from test subjects to the two reference populations. These scores are then normalized so that they are generalizable and robust to missing genotype data. We evaluated the performance of GrafAnc using data from databases including All of Us, UK BioBank, Human Genome Diversity Project (HGDP), as well as those from ancient humans. GrafAnc successfully clusters subjects into subcontinental or country level groups. Unlike many ancestry inference tools, including GRAF-pop, which often cluster European subjects into a single group, GrafAnc separates Finnish and Southern Europeans from others and reveals subcontinental population structure as well. Moreover, GrafAnc can distinguish subjects of Middle East, North Africa, and East Africa from Europeans and other populations. For Asians and Pacific Islanders, GrafAnc can separate Japanese, Korean, Northern and Southern Chinese, Vietnamese, Filipinos, etc, from one another. Importantly, ancestry inference results generated by GrafAnc are reproducible, comparable, combinable, and plottable across different datasets and studies, ensuring

consistency and facilitating broader genetic research. GrafAnc is also very fast and easy to use. It can finish ancestry inference in less than one hour for a dataset with tens of thousands of subjects. Datasets in VCF or PLINK formats can be directly passed to GrafAnc with zero or very little preprocessing.

---PAGEBREAK---

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3019T: Gut microbiomes from diverse human populations induce unique transcriptional host response

Authors:

S. Arif[1], S. Nirmalan[2], A. Alazizi[2], H. Mair-Meijers[2], M. Poyet[3,4], M. Groussin[3,4], F. Luca[2], R. Blekhman[1]; [1]Univ. of Chicago, Chicago, IL, [2]Wayne State Univ., Detroit, MI, [3]Univ. of Kiel, Kiel, Germany, [4]Global Microbiome Conservancy, Kiel, Germany

Abstract:

Intestinal microbiomes of industrialized populations are compositionally different and less diverse than their non-industrialized counterparts, and are associated with immune dysregulation and gastrointestinal disease. However, as most current knowledge is based on correlative studies, it is not known if interpopulation variation in the microbiome impacts host physiology, and through what mechanisms. We used an in vitro system to quantify the impact of inter-population variation in the gut microbiome on human colonic epithelial cells. We treated host cells with live, stool-derived gut communities from non-industrialized and industrialized populations from Rwanda, Ghana, Nigeria, Malaysia and the United States and used RNA sequencing to quantify host gene expression response. We found that industrialized microbiomes induced larger and more significant transcriptional responses in host cells compared to non-industrialized microbiomes. Further, the abundances of specific bacteria drive specific host gene expression patterns. For instance, we found that taxa enriched in industrialized groups regulate the expression of host gene pathways involved in Wnt signaling and Notch signaling, while taxa enriched in non-industrialized groups induced host pathways related to cell structure. These findings suggest a mechanism by which interpopulation variation in the gut microbiome associated with different life-styles can affect host physiology, and open the door to further investigation of the impact of industrialization on host health mediated by the gut microbiome.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3021T: Haplotype-sharing among 860K individuals in Britain and Denmark reveals migration pattern across the North Sea ⭐

Authors:

X. Zhang[1], T. Fitzgerald[1], K. Banasik[2], P. Holm[3], S. Brunak[2], E. Birney[1]; [1]European Bioinformatics Inst., Cambridge, United Kingdom, [2]Univ. of Copenhagen, Copenhagen, Denmark, [3]Trinity Coll. Dublin, Dublin, Ireland

Abstract:

Background
The North Sea has been a strategically important area with an extensive history of people and goods flowing through and around. However, it remains unclear how historical migration events across the North Sea have affected the genetic structure of the modern human population in its coastal nations. Here we use the pairwise haplotype sharing from 858,635 British and Danish individuals, to assemble population genetic evidence and infer the migration history between British and Danish populations within the past 1,000 years.
Methods
Using nationwide healthcare registries in Denmark, we analyzed haplotype sharing data from 370,259 Danes. We also extracted close pedigree relationships from a national multi-generation family registry of all Danes (about 7 million) to validate the genetic relationships. To study gene flow between Britain and Denmark, we used haplotype sharing data between 370,259 Danes and 488,376 British individuals from the UK Biobank to detect regions with excess haplotype sharing across the two populations.
Results
Compared with previous studies that show the Danish population as highly genetically homogeneous, we identified fine-grained population structure within Denmark characterized by a Jutland-Zealand cline and a North Jutland-South Jutland cline. We also discovered a significant decline in the Jutland population post-Black Death compared to Zealand, which continued for at least five generations. Using 860K IBD sharing data between British and Danes, we identified two major migration links over time. Overall, Danes from South Jutland are most closely related to the British. The English from the eastern coast, spanning from Yorkshire and Lincolnshire to the South East, have the most IBD sharing with Danes. This coast-to-coast genetic link is most likely driven by trading and fisheries history. Until approximately 200 years ago, this link shifted towards city-to-city connections, with increasing IBD sharing between metropolitan areas across countries, such as Copenhagen and London, likely driven by the industrial revolution.

Conclusion

Our work provides a novel fine-scale view of the population genetic history of the modern Danish population. We discovered two significant migration links between Denmark and Britain associated with different economic activities. Additionally, we presented one of the largest cross-country biobank IBD sharing analyses on two closely related European populations, demonstrating the use of multiple nationwide biobanks as a powerful tool to study historical relationships within and across populations.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3022T: Hidden Population Structure in the All of Us Research Project ⭐

Authors:

R. Shemirani[1], C. Caggiano[2], S. Hanks[1], K. Hou[3], B. Pasaniuc[4], S. Asgari[1], E. Kenny[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Icahn Sch. of Med. at Mt Sinai, New York, NY, [3]Univ. of California, Los Angeles, Los Angeles, CA, [4]UCLA, Los Angeles, CA

Abstract:

The All of Us (AoU) Research Project is an essential resource in biomedical research, due to its comprehensive coverage of diversity in the US. Available information for participants indicates broad population categories. However, this lack of granularity may be inadequate to fully understand the impact of population structure in AoU. Founder effects are a prominent example of when undetected enrichment of cryptic relatedness within a sub-group of a larger cohort may confound analysis.

To systematically uncover evidence of founder effects in AoU, we conducted unsupervised label-free network analysis in participants with available genomic data (N=245,395) using segments of the genome shared among individuals identical-by-descent (IBD). We developed a scalable pipeline to identify IBD haplotypes from genotype data, filtered for high quality IBD segments, and assigned participants to clusters based on their IBD connections. Individuals who were 1st or 2nd degree relatives (N=15,376) were excluded from all downstream analysis.

We found that 19% of all pairs of individuals shared ≥1 IBD haplotype ≥6cM in AoU, indicating that each individual in the pair inherited this segment from a recent common ancestor. Analysis of the full network of 15 billion pairwise IBD haplotypes, revealed 480 clusters where IBD sharing within the cluster was significantly enriched relative to outside of the cluster (Wilcoxon test $p<1.81\times10^{-5}$). Of these, 146 clusters revealed evidence of founder effects, assessed by calculating the mean of the log of the sum of IBD sharing within each cluster; founder clusters

were defined as being 1 standard deviation away from the mean of all clusters. These founder clusters ranged in size 104-12,798 individuals, and in total, 35% of AoU participants were members of a founder cluster. All founder clusters were sub-groups of broader population groups provided by the AoU. We compared the distribution of 22 phenotypes, 6 socio-economic and demographic factors, and 7 polygenic risk scores (PRS) within members of each founder cluster to members of the broad group excluding the founder cluster. We demonstrated, for example, that one large founder cluster (N=12,798) within the AoU EUR cohort (N=123,281) had significantly higher prevalence of prostate cancer, breast cancer, and type 2 diabetes ($p < 0.001$) after adjusting for covariates, however also demonstrated decreased PRS predictive power for the same conditions ($p < 2.22 \times 10^{-82}$).
This work demonstrates the degree of hidden population structure within AoU that can be detected, even in the absence of population labels, and is associated with differences in disease prevalence, demographics, and PRS accuracy.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3023T: Increased representation of human genetic ancestries improves genetic intolerance metrics

Authors:

A. Han[1], C. Sands[1], D. Matelska[2], J. Butts[3], V. Ravanmehr[1], F. Hu[2], G. Esmeralda[1], Q. Wang[4], S. Petrovski[5], D. Vitsios[2], R. Dhindsa[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, [3]Rice Univ., Houston, TX, [4]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, [5]Dept. of Med., Austin Hlth., Univ. of Melbourne, Melbourne, Australia

Abstract:

Rapidly growing genomic databases have made it possible to identify regions of the human genome that are intolerant to genetic variation and thus likely relevant to human disease. However, a majority of the individuals represented in these large datasets are of European ancestry. It is well-established that this disproportionate representation affects genomic health equity. For example, we and others have shown that individuals of non-European ancestry are substantially less likely to receive genetic diagnoses due to uncertainty of standing variation in these populations. Here, we further examine how this unequal representation impacts the performance of several gene-level and sub-genic intolerance metrics. We computed varying genic and sub-genic intolerance scores (RVIS, Gene-level MTR, Sliding-window MTR, LOF O/E, LOF-FDR) using two large whole-exome sequencing datasets: the UK Biobank (n=460,551) and

gnomAD v2.1 (n=125,748). Our analysis demonstrated that metrics derived on data from ancestral groups with the highest genetic diversity, specifically African and Admixed American ancestral groups, are particularly effective at identifying haploinsufficient genes and disease risk genes. These differences were most apparent in RVIS and missense-specific intolerance scores (Gene-level MTR and Sliding-window MTR). Creating randomly sampled datasets with differing levels of ancestral representation, we found that both sample size and increased genetic diversity are crucial to the performances of these intolerance scores. In comparison to RVIS and MTR metrics, LOF intolerance metrics seemed more dependent on sample size. Overall, our work underscores that increasing ancestral diversity in genomic databases is not only a concern for health equity but also critical to improving our ability to identify genetic variants associated with human disease.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3024T: Initial analysis of 160,000 whole genomes from the Emirati Genome Program

Authors:

J. Quilez[1], D. Sanchez[1], C. Cremin[1], H. Wu[1], H. Sajad[1], **P. Jones**[1], J. Mafofo[1], G. Katagi[1], H. Al Mabrazi[1], V. Kusuma[1], S. Pejathaya[1], S. Elavalli[1], F. Aldhuhoori[1], I. Eltantawy[1,2], M. Ibrahim[1,2], A. Alsuwaidi[1], K. Thangarajan[1], A. M. Yousif[3], M. Olbrich[4], M. S. Al Ameri[3], I. Chishty[1], S. A. Al Marzooqi[3], W. M. Abdulrahman[3], A. I. Al Mannaei[3], M. Mousa[4], H. Al Safar[4], I. Iqbal[3], F. Al Marzooqi[1], E-K. Albarah[1], A. Al Awadhi[1], T. Magalhaes[1]; [1]M42, Abu Dhabi, United Arab Emirates, [2]IROS, Abu Dhabi, United Arab Emirates, [3]Dept. of Hlth.of Abu Dhabi, Abu Dhabi, United Arab Emirates, [4]Khalifa Univ., Abu Dhabi, United Arab Emirates

Abstract:

The Emirati Genome Program (EGP) will perform ~30X whole-genome sequencing (WGS) of all ~1 million Emiratis in the United Arab Emirates by 2025. Enrollment is voluntary and with no inclusion or exclusion criteria for age, gender or health status. DNA for WGS is derived from blood, mostly, and less-invasive buccal swabs for some participants (e.g. children). Unique to the EGP is the use of three high-throughput sequencing (HTS) technologies: Illumina and MGI (both short-read sequencing) and Oxford Nanopore Technologies (ONT) long-read sequencing - each EGP participant is sequenced in one of them. This multi-platform strategy was selected to maximize sequencing speed and to build the largest long-read WGS cohort worldwide.
So far, ~600,000 Emiratis have enrolled into the EGP and been sequenced. We see no gender-bias (females = 49.5%; males = 50.5%) and coverage of all age groups (0-75+), with enrichment

for the 18-40 years old range. All three HTS platforms are rather equally represented (Illumina = 107,920, MGI = 44,321 and ONT = 9,142). We implemented and optimized analysis pipelines for each of the three HTS platforms to absorb our sequencing. Not surprisingly, we found Illumina's DRAGEN as the most mature analysis solution, especially through the clustering of multiple DRAGEN servers. For MGI we used both cloud-based and on-premises analysis workflows using, respectively, Sentieon and MGI's ZBOLT. We adapted our ONT analysis pipeline to match the fast advancement in the analysis tools for this technology. For both MGI and ONT, we upgraded our pipelines to match, when possible, the callers included in DRAGEN or similar ones. We defined minimum criteria for 7 parameters to define high-quality individual WGS datasets. We selected these parameters based on project requirements (e.g. 30X mapping coverage) as well as best practices from other large-scale genome sequencing programs such as the UK Biobank and All Of Us. We applied such high-quality standards to an initial freeze of 150,000 EGP WGS datasets, resulting in ~144,000 (~95%) passing such filters. The three platforms performed similarly well (% of samples passing filters: Illumina = 93.3%, MGI = 97.3% and ONT = 95.2%). As expected, the pass rate for blood (98.2%) was better compared to buccal swab (90.0%). Indeed, most failed samples were buccal swabs with a mapping rate to the human genome reference sequence below target (70%), likely caused by a relatively high contamination from non-human species. In summary, our work shows (i) the progress of the EGP towards its ~1 million genomes target, (ii) its readiness in terms of sequencing and analysis infrastructure and (iii) the high quality of the EGP genomics datasets.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3025T: Integration of ribosome profiling experiments identifies a large pool of novel translated open reading frames neutrally evolving in the human genome

Authors:

J. Lee[1,2,3], A. Wacholder[1,2], M. Chikina[1], M. Zhang[4], A-R. Carvunis[1,2]; [1]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [2]Pittsburgh Ctr. for Evolutionary Biology and Med., Pittsburgh, PA, [3]Joint CMU-Pitt Ph.D. Program in Computational Biology, Pittsburgh, PA, [4]Carnegie Mellon Univ., Pittsburgh, PA

Abstract:

Human annotated proteins are translated from sequences spanned by open reading frames (ORFs) that are largely conserved. In contrast, recent advances in ribosome profiling have uncovered a vast number of unannotated human noncanonical ORFs (nORFs) that are

evolutionarily novel, existing mostly in the primate lineage. Because they are generally short (<100 amino acids) and taxonomically restricted, nORFs have been missed by current functional annotations. Thus, they are largely uncharacterized, and their prevalence in the human genome is an open question. In this work, we expand the known set of nORFs by integrating 1,067 publicly available human ribosome profiling experiments, and investigate their evolutionary roles and contribution to phenotypes. Our study utilizes SNP information from TOPMed whole genome sequencing (WGS) data (N=132,345), GWAS summary association statistics across 59 diseases and complex traits (avg. N=351,446), and fine-mapped SNPs from 94 UK Biobank diseases and complex traits (maximum N=361,194).We identify over 66,000 nORFs, including 51,461 not previously reported. We further study their functional impacts and reach three conclusions. First, the nORFs have a nonsynonymous-to-synonymous SNP ratio close to, but significantly less than the background value of 0.998 (adjusted pN/pS=0.960; bootstrapped p-value=0.00380), suggesting they are evolving close to neutrally. Second, by meta-analyzing LD score regression (LDSC) results across 59 nonredundant GWAS phenotypes with genetic correlation <0.1, we further find that nORFs are depleted with heritability, but provide independent information relative to the 96 functional annotations in the baseline-LD model (LDSC τ=-6.45±1.15×10-7, p=1.91×10−8). Third, compared to annotated coding sequences, nORFs have a substantially lower enrichment of fine-mapped SNPs (10.61±0.28 annotated coding sequences vs. 4.79±0.29 nORFs; p-value=0.0), but are more enriched for fine-mapped SNPs than negative control (1.57±0.19; p-value=0.0). In conclusion, though a growing number of case studies are reporting nORFs with biologically important function, our results suggest that the bulk of human nORFs are evolving neutrally as a whole with low contributions to complex traits and disease.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3026T: Local ancestry analysis identifies signatures of selection in the admixed Luo population from Kenya

Authors:

S. Jones[1], D. Harris[1], M. Hansen[1], A. Platt[1], S. Browning[2], S. Williams[3], C. Fokunang[4], A. Njamnshi[4,5], NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium, S. Tishkoff[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Washington, Seattle, WA, [3]Case Western Reserve Univ, Cleveland Heights, OH, [4]Univ. of Yaoundé I, Yaoundé, Cameroon, [5]Brain Res. Africa Initiative (BRAIN), Yaoundé, Cameroon

Abstract:

The Luo are a Nilo-Saharan-speaking population from Kenya with a high proportion of Bantu ancestry. They migrated from the southern region of modern Sudan into the western region of Kenya near Lake Victoria in multiple waves from at least 1,000 years ago, experiencing significant environmental shifts between these regions (e.g. low malaria to high malaria endemicity). This migration led to interactions and admixture with neighboring Bantu-speaking groups and shifts in their subsistence from traditional pastoralists to predominantly farmers. Our objective was to search for signatures of natural selection by identifying regions of excess differences in the proportion of Bantu and Nilotic ancestries in the Luo genomes. Using high-coverage whole genome sequencing data from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program we inferred global genomic ancestry using ADMIXTURE and local ancestry using FLARE in 91 Luo individuals using the Cameroonian Tikari Bantu speakers (n=63) and the Kenyan Pokot Nilotic speakers (n=66) as reference populations. Our analyses revealed that approximately 80% of the Luo's global ancestry was Bantu, despite speaking a Nilotic language. We identified multiple genomic regions that showed significant deviations from the genome-wide average contribution of Bantu or Nilotic ancestries. Notably, regions on chromosomes 6 and 7 exhibited a high excess of Nilotic ancestry, including genes such as *HLA-DQA1* and *HLA-DQB1*, which are involved in immune response. Conversely, regions on chromosome 2 showed a significant excess of Bantu ancestry, including genes such as *SLC24A4*, which plays a role in cellular ion transport and pigmentation. These findings suggest potential local adaptations, with further investigation required to understand their implications in the context of potential adaptation to novel diets and high malaria endemicity. This work was supported by grant numbers: ADA 1-19-VSN-02, NIH grants 1R35GM134957, R01AR076241, and 1X01HL139409-01.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3027T: Long-read *de novo* assembly and comparative analysis of six howler monkey genomes within genus *Alouatta*

Authors:

B. Chen[1], P. de Freitas[2], L. Portela[2], A. Goldberg[1]; [1]Duke Univ., Durham, NC, [2]Federal Univ. of São Carlos, São Carlos, Brazil

Abstract:

As our closest relatives, primate genomics and evolution inform human genetic interpretation. Though only one species in the genus *Homo* exists today, multiple monkeys have wide radiations, allowing us to compare closely related species within a genus. Despite a growing number of primate genomes available, most genera are represented by a single reference

genome. Here, we generate *de novo* reference assemblies using long-read sequencing for 6 Howler monkeys (genus *Alouatta*) to understand within-genus, between-species evolution. With between 9 to 14 species howler monkeys are some of the most widely distributed platyrrhines, extending from southern Mexico to northern Argentina, inhabiting a range of environments. Prior phylogenetic investigations into howler monkeys have exhibited discrepancies across studies or remained inconclusive, based a mix of on cytogenetic, morphological traits, and a restricted set of molecular markers. Attempts to address these questions have been limited because only a single poor-quality reference genome of mantled howler monkey (*Alouatta palliata*, Mexico) is currently available (assembly contig count: 1152695; N50: 51.304 Kbp). Our assemblies contain roughly ~2000 contigs, with N50 of ~15Mb and L50 of ~60. Based on phylogeny constructed with OrthoFinder, IQtree, and ASTRAL, we find the clade grouping *A. belzebul* and *A. belzebul /discolor* as sister taxa, with *A. belzebul/ululate*, *A. guariba*, *A. seniculus*, and *A. caraya* as respective sister lineages to this clade. We identify loci that show species-specific signatures of selection, as well as adaptive loci shared across the genus. In addition, we are testing species-specific and genus-specific structural variation that may be associated with howler-specific phenotypic traits.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3028T: Massive expansion of a Denisovan-introgressed tandem repeat in MUC19 in Americans ⭐

Authors:

E. Kaufman[1], E. Nizamis[1], D. Peede[2], F. A. Villanea[3], E. Huerta-Sanchez[2], P. Valdmanis[1]; [1]Univ. of Washington, Seattle, WA, [2]Brown Univ., Providence, RI, [3]Univ. of Colorado Boulder, Boulder, CO

Abstract:

The enrichment of genomic variants in a specific population has several potential explanations, including genetic drift, bottleneck events, and selective adaptation. Mucins are proteins characterized by glycosylated tandem repeats which form mucous to protect organs and to serve as a barrier to pathogens. Using a computational tool we developed, Isolate and neaten VNTRs (InVNTR), we identified a massive coding variable number tandem repeat (VNTR) expansion in Mucin 19 (MUC19) in 70 phased long-read samples from the Human Pangenome Reference Consortium and the Human Genome Structural Variation Consortium. This expansion occurred exclusively on the background of a haplotype introduced from Denisovans. The MUC19 locus harbors signals of adaptive introgression, and several prior reports provide evidence for the region around MUC19 as a signal of positive selection in individuals with

Indigenous North American ancestry, especially in Mexico. Remarkably, we found stable length of this 30bp repeat ranging from 8,178 - 10,008bp in non-American populations but several alleles from American samples had a 22,692bp repeat, the result of five extra copies of a 3,171bp in-frame copy number expansion event. This expansion appears to occur in a stepwise manner as several individuals of East Asian descent have three copies of the ~3kb expansion, while the full expansion has reached 0.15 minor allele frequency in Admixed American populations from the 1000 Genomes Project. Using the NIH All of Us Research Program's genomic and electronic health record (EHR) data, we ran a phenome wide association study (PheWAS) on SNPs in high linkage disequilibrium (LD) with the expansion, including rs74565412, which captures a signal of positive selection in North Americans. When correcting for sex, age, and the first 3 principal components, we found nominal associations in individuals with primarily Latin American genetic ancestry including disorders involving the immune mechanism (OR = .76, p = 0.002, n cases = 708, n controls = 27,027). These findings may explain the signature of previously reported positively selected SNPs in Central Mexico and suggest that VNTR expansion may explain signs of positive selection in the context of ancient genome introgression events.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3029T: Peruvian Population Genomics: Diversity, Admixture Dynamics, and the Impact of Indigenous American Ancestries on Complex Phenotypes ⭐

Authors:

V. Borda Pua[1], O. Caceres[2], C. Sanchez[2], C. Padilla[2], D. Veliz-Otani[3], M. Mendes de Aquino[4], C. Silva de Carvalho[5], I. Alvim[6], B. A. Maron[7], M. Gouveia[8], Peruvian Genome Project Consortium, E. Tarazona-Santos[9], H. Guio[10], T. O'Connor[1]; [1]Univ. of Maryland Sch. of Med., Baltimore, MD, [2]Inst. Natl. de Salud, Lima, Peru, [3]Univ. of Maryland Baltimore, Baltimore, MD, [4]SickKids Hosp., Toronto, ON, Canada, [5]UFMG/NCI, Rockville, MD, [6]UFMG, Belo Horizonte, Brazil, [7]The Univ. of Maryland-Inst. for Hlth.Computing, Rockville, MD, [8]NIH, Bethesda, MD, [9]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil, [10]INBIOMEDIC, Lima 14, Peru

Abstract:

Latin American populations have a recent history of extensive migration and admixture, shaping their genetic and phenotypic diversity. Specifically, admixed Peruvians are recognized as having predominantly Indigenous American ancestry and tend to be shorter. Additionally, like in many low and middle-income countries, the rate of obesity is increasing in Peru. As part of the

Peruvian Genome Project, we generated genome-wide data for 336 admixed individuals combined with 515 previously described samples representing 13 Peruvian regions to explore population structure, demographic history, and the relationship between Indigenous American ancestries and complex phenotypes (i.e., height, body mass index [BMI], and waist-to-hip ratio). We identified that Indigenous American ancestries in admixed Peruvians mirror the north-south interaction known from archeology and linguistics previously described for Peruvian Natives. We detected the most recent admixture events between European and Native Peruvians dating to 8-10 generations ago which occurred with populations already admixed between distinct original populations. We observed differences in continental ancestry proportions between autosomes and chromosome X, suggesting sex bias in admixture with a predominant contribution of Indigenous American females. Moreover, small stature correlated positively with a higher proportion of Indigenous American ancestry but no specific locus reached genome-wide significance. For BMI, an intronic variant in 4q25 corresponding to an uncharacterized non-coding RNA, LOC105377376, was significantly association with this trait. Furthermore, we identified a significant association between an INDEL in the locus 7p21.1 and the waist-to-hip ratio. Interestingly, both significant associations are common in Afroperuvians and African populations from the 1000 Genomes Project, but rare in other populations (MAF <5%). Our fine-scale population study provides subtle details of the genetic history and diversity, serving as one more step to understanding the broader human genetic variation landscape.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3030T: *PGG*.PGP: an academic web server for personal genome analysis

Authors:

B. Wang[1], X. Zhao[2], Y. Shi[1], X. Zhang[3], H. Chen[1], M. Wei[1], Y. Cui[2], S. Xu[4]; [1]Sch. of Life Sci. Fudan Univ., Shanghai, China, [2]Human Phoneme Inst. Fudan Univ., Shanghai, China, [3]Sch. of Life Sci. and Technology, ShanghaiTech Univ., Shanghai, China, [4]Fudan Univ., Shanghai, China

Abstract:

As personal genomic data becomes increasingly accessible to a broader population, the need for comprehensive genome interpretation intensifies. Addressing this need, we have developed PGG.PGP, a comprehensive and freely accessible web platform aimed at facilitating the detailed analysis and interpretation of individual genomic data, thereby contributing to the evolving field of personalized medicine. Users can upload their genotype data to access a variety of genetic analyses and interpretations. First, PGG.PGP provides genetic ancestry dissection of an

individual genome under various contexts using state-of-the-art methods. The platform also reports paternal and maternal lineages, archaic introgression detected from an individual genome, with further annotation of functionality, phenotype, or disease association. Secondly, PGG.PGP utilizes both single nucleotide variant (SNV) and structural variant (SV) data to offer comprehensive resources from popular databases like PharmGKB, ClinVar, SNPedia, and PGG.SNV and PGG.SV. Additionally, cutting-edge variant effect predictors such as EVE and M-CAP are employed to assess risk. Thirdly, to enable users to be aware of the potential risks or effects of the genetic variants identified in their genomes, PGG.PGP also estimates PGS for a wide range of diseases and phenotypes as cataloged in the most recent PGScatalog database. Fourthly, PGG.PGP offers HLA typing based on a well-established pipeline available in PGG.MHC, provides users valuable information about their HLA type critical for matching patients and donors in bone marrow or cord blood transplants. Furthermore, PGG.PGP enhances the accuracy of genotype imputation based on an in-house reference panel comprising 19,554 high-quality genomes with deep sequencing, thereby providing more comprehensive and detailed genetic analysis results. By offering these diverse functionalities, PGG.PGP assists individuals in exploring their unique genetic makeup and bridges the gap between complex genetic data and actionable health-related insights, paving the way for personalized health strategies. This feature is particularly valuable for personalized medicine. To our best knowledge, PGG.PGP stands as the sole web server freely accessible, providing genetic interpretation to both the general public and the scientific community.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3032T: Robust detection of ancient trans-species balancing selection of varied footprint sizes

Authors:

A. Campelo dos Santos, M. DeGiorgio; Florida Atlantic Univ., Boca Raton, FL

Abstract:

Balancing selection is an evolutionary process that maintains genomic diversity by increasing the density of polymorphisms of moderate frequency neighboring the locus under selection. A notable case of a genomic region evolving under this process is the major histocompatibility complex (MHC), known to harbor signals of long-term balancing selection due to trans-species polymorphisms among great apes. However, such trans-species polymorphisms can result from technical issues, and only a few approaches for detecting long-term trans-species balancing selection avoid using these sites. Yet these methods suffer as they must employ a user-defined

genomic window size and only achieve optimal performance when this size matches the narrow footprint expected for signals of long-term balancing selection. As a result, they are sensitive to varying selection footprint sizes, losing power when analysis windows are too large and increasing false discovery when windows are too small. Conversely, recent likelihood test statistics, termed *B* statistics, were constructed and demonstrated to robustly detect long-term balancing selection without using genomic windows. However, these statistics only consider polymorphism from a single species, making their power suboptimal if polymorphism information from other species is available. Here, we introduce *FLAT-ABS* (FLexible Assessment of Trans-species Ancient Balancing Selection), an application that extends the *B* statistic likelihood framework to operate on polymorphism and substitution information present across an arbitrary number of species without using sites that are polymorphic in more than one species. We thoroughly tested the performance of these new statistics in detecting trans-species balancing selection of different modes, ages, strengths, and equilibrium frequencies, as well as against other settings such as sample sizes and variable recombination and mutation landscapes. Overall, we find that *FLAT-ABS* has comparable or higher power than other trans-species balancing selection statistics computed at their optimal window sizes, with significantly better performance when such methods employ suboptimal window sizes, which are unknown in advance. Our application of *FLAT-ABS* to human and chimpanzee whole-genome data detected candidate footprints of long-term balancing selection affecting great apes, such as the previously established MHC region and *ESR1* gene. Finally, *FLAT-ABS* identified a slew of novel candidate genes, some of which are involved in contemporary deleterious phenotypes, providing new pathways to a deeper understanding of the long-term evolutionary history of great apes.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3033T: Structural variation hotspots in the human genome and their evolutionary implications

Authors:

Z. Yang[1], S. Li[1], X. Tan[2], Chinese Pangenome Consortium (CPC), S. Xu[1,3]; [1]State Key Lab. of Genetic Engineering, Human Phenome Inst., Zhangjiang Fudan Intl. Innovation Ctr., Ctr. for Evolutionary Biology, Sch. of Life Sci., Fudan Univ., Shanghai, China, [2]Key Lab. of Computational Biology, Shanghai Inst. of Nutrition and Hlth., Univ. of Chinese Academy of Sci., Chinese Academy of Sci., Shanghai, China, [3]Sch. of Life Sci. and Technology, ShanghaiTech Univ., Shanghai, China

Abstract:

Structural variations (SVs) play an important role in human evolution and disease etiology, but little is known about their distribution and evolutionary mechanism in the genome. Here, we analyzed high-quality third-generation sequencing data of 126 samples representing 5 continental populations as provided by the Human Pangenome Reference Consortium (HPRC, n=47), the Chinese Pangenome Consortium (CPC, n=58), and the Human Genome Structural Variation Consortium (HGSVC, n=21). We detected SVs from haplotype-resolved assemblies and uncovered 495,349 SVs, of which 2,099 SV hotspots encompassing ~208 Mbp on T2T-CHM13. With the centromere and telomere being excluded, the SV hotspots account for only 7.61% of the whole genome but are enriched with 80.8% of all SVs detected. Correspondingly, the SV density within SV hotspots is 51-fold higher than the non-hotspots region. Interestingly, about 9% of the SV hotspots are T2T-CHM13-specific compared with the GRCh38. We further distinguished recurrent SVs from ancestral SVs by differentiating nucleotide diversity levels in the flanking region of SVs, and estimated the recurrent SV fragment carried by each person accounts for about 0.3% of the entire genome. We suggest that both population demographical history and natural selection played important roles in shaping the distinct patterns of the inversions, outstanding examples include a recurrent one flanking FLNA-EMD genes and an ancestral one inner SUGCT. Notably, the variable number of tandem repeats (VNTR) might be a critical driving force of SV hotspots, an outstanding example is PTPRN2 which ranks the top region enriched with SVs. Moreover, mobile element insertions (MEI) across the genome are highly polymorphic and substantially contribute to the structural architecture of the human genome. We built the first comprehensive map of SVs and SV hotspots based on telomere-to-telomere (T2T) human genomes and more than 200 high-quality haplotype-resolved TGS assemblies, and our results indicated the complex mechanisms behind the heterogeneous distribution of SVs in the human genome, which further imply disease susceptibility differences among populations and individuals.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3034T: Surprising source of confounding impacts the sign of GWAS effect estimates on immune function

Authors:

S. Smith, S. Rybacki, A. Harpak; Univ. of Texas at Austin, Austin, TX

Abstract:

An often neglected facet of genetic architecture is the direction of allelic effects. A systematic tendency towards trait-increasing or decreasing alleles (i.e. a “sign bias”) in genome-wide

association studies (GWAS) may reflect 1) a mutational bias, where de novo mutations tend to impact the trait in one direction; 2) natural selection, which might systematically favor effects in one direction; 3) ascertainment bias and statistical artifacts; or some combination of these three phenomena. Here, we develop an inference method to investigate sign bias on effect estimates across the allele frequency spectrum and apply it to hundreds of continuous traits and diseases in the UK Biobank and Biobank Japan. Across traits, we observe no sign bias with respect to derived alleles. However, the effects of minor alleles on immune-related traits tend to show a positive sign bias. For example, lymphocyte count across the folded allele frequency spectrum. Given the absence of sign bias in derived alleles, we propose that an underappreciated mode of population structure drives the sign bias in immune-related traits, rather than a mutational bias or selection. Our results suggest population structure can substantially confound inference of genetic architecture in a previously unappreciated manner.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3035T: The Genetic History of the Qiang People

Authors:

W. Zeng, S. Xu; Fudan Univ., Shanghai, China

Abstract:

The Qiang people, residing in the transitional high mountain region from the Sichuan Plain to the Tibetan Plateau in southwest China, are one of the most ancient populations. However, their genetic origins and the mechanism shaping their genomic diversity remain largely unknown, especially due to the lack of high-quality genome data. Here, we made the first effort in whole-genome sequencing of 20 Qiang genomes to high coverage (~32×), and analyzed them along with Tibetan highlanders, lowlanders Han, as well as other available genomes. Our results indicated that genetic admixture with Han and Tibetan has largely shaped the genetic makeup of the Qiang people and further driven the formation of sub-populations. Combining with ~450 ancient Asian genomes, we traced back the most recent common ancestor of the Qiang people to be ancient Yellow River farmers ~5300 years ago, supporting the common origin of the Han Chinese and Qiang. Notably, both genetic admixture and positive selection contributed to the high differentiation of genetic variants associated with ethanol metabolism and pigmentation between the Qiang sub-populations. Moreover, in contrast to highlander Tibetan and lowlander Han Chinese, the high prevalence of missense variants in a blood pressure regulation gene may contribute to the special mountain environment adaptation for the Qiang population. Overall, our results advanced a deeper understanding of the genetic history of the Qiang population,

and highlight the admixture and natural selection in shaping the genetic diversity of populations who reside in high mountain regions.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3036T: The Impact of Associative Overdominance on Genetic Diversity

Authors:

S. Sun, R. Hernandez; Univ. of California San Francisco, San Francisco, CA

Abstract:

The impact of purifying selection on neutral diversity at linked sites has been extensively explored within the framework of codominance. When codominant deleterious variants are subject to weak to moderate selection, there's a reduction in variation at linked neutral sites, known as background selection (BGS). However, if deleterious variants act recessively, an unexpected increase in variation can occur, akin to classic overdominance effects. This phenomenon, termed associative overdominance (AOD), is less thoroughly understood compared to BGS. Using forward simulations incorporating a gamma distribution of fitness effects, we demonstrate how AOD influences genetic diversity via the pairwise nucleotide diversity (π) and the site frequency spectrum (SFS) statistics. Our analysis reveals that the SFS resulting from classic overdominance models differs from those influenced by AOD. Additionally, we illustrate how bottleneck events and the presence of concurrent codominant alleles further alters the effects of AOD. Notably, bottleneck events or the inclusion of codominant alleles alongside fully recessive ones transforms the SFS shaped by AOD to resemble those produced under a milder selection regime.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3037T: Understanding founder variants in Kuwait to expedite diagnosis and reduce burden of genetic diseases.

Authors:

D. Marafi[1], H. Alajmi[1], M. Alajmi[1], N. Alabduljalil[1], R. Alsafi[2], R. Elshafie[3], S. Alkanderi[3], A. Alahmad[3], H. Alrohaif[3], A. Elshafey[3], H. Alsharhan[1], M. Alali[3], B. Albash[3], M. Marafie[3], L.

Bastaki[3]; [1]Kuwait Univ., Jabriya, Kuwait, [2]Ministry of Hlth., Kuwait, Kuwait, [3]Kuwait Med. Genetic Ctr., Kuwait, Kuwait

Abstract:

Introduction Founder variants are disease-causing variants occurring in isolated groups due to a common ancestor carrying pathogenic variant. Thus, a higher carrier rate for the founder variant is observed within the population isolate and accounts for a higher prevalence of rare disease compared to the rest of the population. Knowledge of the existing founder variants within a group or a geographic region can facilitate diagnosis of undiagnosed cases, guide management and offers an opportunity for prevention through screen of at-risk couples. Around 40% of the cases undergoing exome analysis for genetic disorders in Saudi Arabia are found to have a founder variant accounting for the disease. Such information is lacking in Kuwait. We aim to identify the common founder variants in Kuwait and understand their distribution in order to facilitate diagnostic effectiveness and estimate their burden to plan preventive measures on a national scale. **Methods:** We reviewed the Kuwait Medical Genetic Center database and surveyed clinicians and laboratory technicians for variants observed recurrently within families or tribes in Kuwait. We also reviewed the literature extensively for previously reported founder variants in Kuwait. **Results:** We arrived at 62 founder variants in Kuwait. Most of the variants underlie autosomal recessive disease traits (60, 97%) while only two variants (3%) are associated with autosomal dominant disease traits. Of the founder variants identified, only 8 variants (13%) were novel and specific to Kuwait, while 46 variants (72%) have been previously reported in other countries from the Arabian Peninsula, and 8 variants (13%) were previously reported across the Middle East North Africa region. Approximately, 68% of variants were observed in isolated tribes in Kuwait, while the rest were observed in multiple tribes with shared co-ancestry or permissible inter-triable marriage. Commonly observed variants in multiple tribes were in the following genes: *TBCE* accounting for Sanjad-Sakati syndrome, *ADAT3* related intellectual disability, and *SLC19A3* associated with biotin-thiamine-responsive basal ganglia disease. While these disorders are rare worldwide, they prevalence in the Kuwaiti population exceeds that of rare disease (0.8-9 cases per 100,000 Kuwaitis). **Conclusion:** Our data confirm that founder variants are common in Kuwait and associated with a significant health-care burden. Majority of the founder variants underlie autosomal recessive disease traits. Our findings enable prevention through national genomic-based premarital screening and counseling of carrier couples to reduce genetic disease burden within the country.

Session Title: Evolutionary and Population Genetics Poster Thursday Session

Board 3038T: What can admixture tell us about the genetics of health disparities?

Authors:

A. Zaidi[1], K. Bird[2]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Univ. of California Davis, Davis, CA

Abstract:

A major goal in human genetics is to understand whether health disparities are driven by differences in average genetic risk. To this end, researchers often test for correlations between health-related phenotypes and individual ancestry in admixed populations. Such correlations might be suggestive of the presence of highly differentiated or 'population-specific' alleles that could then be mapped by admixture mapping. However, much of the early motivation of this approach was based on the assumption that there might be one or a few causal loci driving the difference in genetic risk. How likely are we to find 'population-specific' alleles for polygenic traits that covary with ancestry? To address this, we use analytical theory and simulations to describe the quantitative behavior of ancestry-trait correlations as a function of admixture history and the genetic architecture of the trait.

We show that the interpretation and 'mappability' of ancestry-trait correlations is more complex, even in the absence of confounding environmental effects. Ancestry-trait correlations are expected to be modest in most admixed populations because such correlations are less an indicator of the presence of differentiated alleles and more a measure of the linkage disequilibrium (LD) between causal loci and ancestry informative markers. This LD decays rapidly after admixture due to recombination and independent assortment. Thus, only in the extreme case, e.g. when the genetic risk is driven by a few highly differentiated large-effect alleles, do we expect substantial ancestry-trait correlations. We argue that such cases are rare (e.g. DARC for neutrophil count and APOL1 for end-stage renal disease) and would have already been discovered given the current size of admixed cohorts.

For polygenic traits, ancestry-trait correlations are even weaker because the heritability is distributed across a large number of loci and it is unlikely that they will all be highly differentiated between populations in the same direction, even under moderate selection. The only known exception is skin pigmentation - an oligogenic trait under very strong divergent selection. We believe that most traits are unlikely to have this type of architecture and if they did, we would have mapped much of the heritability in existing admixed cohorts.

We argue that the presence of widespread correlations between ancestry and health-related traits among admixed individuals (e.g. African Americans and Latinos) is an indicator of environmental confounders (e.g. structural racism and access to healthcare) rather than evidence of 'population-specific' alleles and should be interpreted with caution.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7052T: A Comprehensive System for Monitoring Pediatric Genetic Testing Outcomes Using EHR Data: Insights from the eMERGE IV Study

Authors:

P. Maripuri, D. Karavite, J. Thayer, S. Terek, J. Purcell, J. Green, A. Khan, J. Connolly, M. Harr, M. Hess, B. Luo, R. Grundmeier, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Background: Monitoring outcomes from genetic testing is uniquely challenging as it involves predictive results for future risks, and impacts on family members, often compounded by external lab results integrated into electronic health records (EHR). The workflow from a genetic testing result can span months, capturing actions and interactions with patients/caregivers, genetics and primary care providers, and other specialists. Data representing these actions can occur in various areas within the I, including the patient portal.
Methods: We developed a return of results (ROR) system for a Genome-Informed Risk Assessment (GIRA) as part of the eMERGE IV study. GIRA combines polygenic risk scores with patient and family histories to assess risks for asthma, obesity, type I and type II diabetes in a pediatric setting based on defined study thresholds. Critical points in the ROR workflow were identified where EHR data could inform diverse clinical and process outcomes. Data and domain experts collaborated iteratively to create and test queries for these outcome measures.
Results: Our ROR outcome monitoring system utilizes a range of EHR data sources, including data access logs, to capture caregiver and clinician actions. The system tracks outcomes such as notifications and responses by caregivers and clinicians, access to GIRA results, genetic counselor/primary care encounters, caregiver-primary care interactions, and primary care responses including interactions with clinical decision support, lab and medication orders, consults and referrals, visit documentation, and delivery of patient education resources.
Conclusion: Our study demonstrates that while Epic can be a valuable resource for capturing and monitoring genetic testing outcomes, addressing several of its complexities is essential to maximizing its utility. To ensure comprehensive data retrieval, supplementation with top-down approaches is required due to the diverse and interconnected nature of the outcomes. Implementing a multi-site monitoring system is particularly challenging, making it crucial to consider these complexities early in the planning stages to optimize workflow and data collection. Although automation is desirable, it presents its own set of challenges. Nonetheless,

the system we developed for monitoring the return of GIRA results is scalable and adaptable, provided there is input from all major stakeholders, including data experts, domain experts, clinicians, and caregivers. This collaborative approach ensures the system can effectively capture and utilize a broad spectrum of EHR data to inform clinical and process outcomes.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7053T: A Framework for Integrating Polygenic Risk Scores in the Electronic Health Record: The eMERGE Phase IV Study

Authors:

R. Freimuth, H. Bangash, A. Miller, J. Petrzelka, O. Ogunsan, M. Hamed, L. Bade, J. Sutton, D. Davidson, D. J. Huschka, Y. Ali Nur, P. Caraballo, I. Kullo; Mayo Clinic, Rochester, MN

Abstract:

Purpose: Polygenic risk scores (PRS) can improve risk stratification for common diseases and inform interventions for primary prevention. In the eMERGE-IV study, PRS results and additional relevant genomic data for 10 common diseases are being returned to participants as part of a genome-informed risk assessment (GIRA). We enumerate the lessons learned during this process at our eMERGE site.
Methods: As part of a genomics-enabled learning health system (gLHS), we developed a framework to integrate GIRA which includes PRS, monogenic risk and family history, into the EHR using the Quality Enhancement Research Initiative (QUERI) Implementation Roadmap. In the pre-implementation phase, the study team identified institutional champions and obtained technical/financial resources. As part of the implementation phase, research team members underwent training for various study processes and technical solutions for PRS integration were developed and prototyped to align with clinical workflows. The sustainment phase will begin after all results are returned and metrics of clinical decision support (CDS) utilization will be assessed.
Results: As of June 4th, 2024, 2198 of 2250 total participants were enrolled in the eMERGE-IV study at this site; of these, 631 participants were returned a GIRA using the framework described; of the GIRAs returned, 303 were categorized as 'high-risk' with 11 (1.7%) positive for a pathogenic/likely pathogenic monogenic variant, 124 (19.7%) with high PRS, and 168 (26.6%) with a positive family history. Implementing a framework for PRS integration led to important lessons learned: i) timeline and priorities of a federally funded genomics research study differs from real-world clinical and institutional capacities and timelines; ii) engagement and

collaboration with clinicians and research and IT leadership are imperative to harmonize federal processes with institutional ones and establish realistic timelines for implementation; iii) avoiding silos and facilitating communication across various teams involved in the study is necessary to prevent downstream delays in study steps; iv) to ensure adoption of PRS in practice, there is a need to provide clinicians with CDS and education resources; v) implementation science models can account for multi-level contextual factors in the real-world and guide genomic medicine implementation in practice.
Conclusion: We describe use of an implementation framework to integrate PRS results in the EHR as part of the eMERGE IV study. The lessons learned from this process can guide deployment of PRS for common diseases in the clinical setting.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7054T: A scoping review investigating the impact of genetic counselor involvement in genetic and genomic test order review

Authors:

C. Cook[1], C. Pistawka[1], A. M. Elliott[1,2,3], G. Study[1]; [1]Univ. of British Columbia, Vancouver, BC, Canada, [2]BC Children's Hosp. Res. Inst., Vancouver, BC, Canada, [3]Women's Hlth.Res. Inst., Vancouver, BC, Canada

Abstract:

Background: The rise in the availability and complexity of genetic technologies, paired with more genetic tests being ordered by non-genetic healthcare providers (HCPs) has resulted in an increase in inappropriately ordered tests. Utilization management (UM) is the evaluation of the medical necessity, appropriateness, and the efficiency of the use of health care services. Genetic counselors (GCs) are ideally suited to assess the appropriateness of a genetic test considering the patient's medical and family history. **Purpose:** To provide an overview of the literature on GC involvement in UM initiatives to determine the impact of having GCs involved in these processes. **Methods:** Five databases (MEDLINE, EMBASE, CINHAL, EBM reviews and Web of Science Core Collection), grey literature and professional organization guidelines were searched (English literature, published since 2010). Title/abstract screening and full-text review was performed by two independent reviewers with disagreements resolved through discussion with a third reviewer. A data extraction form was used to record key variables including; country of interest, outcomes, clinical indication for testing, degree and timing of GC involvement, type of institution and other (e.g., costing). We completed a descriptive analysis of the variables of

interest and a narrative account of the extracted data based on prevalent themes identified during primary data extraction. **Results:** A total of 51 studies were included and used for data extraction and analysis. The United States (n=47, 92%) was the country of interest for the majority of studies. For studies that described cases of test misorders or a UM program (n=44), 28 (64%) included multiple clinical indications, 11 (25%) focused on hereditary cancer testing and 5 (11%) focused on other clinical indications (e.g. cardiovascular, nephrology, hemophilia B, hemochromatosis, testing in NICU). GCs were involved either as the sole reviewer (n=18, 41%), with multiple GCs (n=8, 18%), or as part of a multidisciplinary team (n=16, 36%). Most commonly evaluated outcomes included diagnostic rate (n=13), economic efficiencies (n=27), cancellation rate (n=21) and impact on patient management (n=31). A number of studies (n=9) described that GC involvement resulted in improved ordering practices of non-genetic HCPs. **Conclusion:** Employment of GCs in the laboratory has been implemented widely as a solution to test misorders. These studies describe a variety of ways in which GCs can be integrated into testing workflows to not only reduce the number of inappropriate tests but to also have wider impacts on nongenetic HCPs ordering practices and the patient experience.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7055T: A stakeholder informed ethical framework for epigenetic editing ⭐

Authors:

R. Major, A. Davis, J. Cadigan; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

The field of epigenetic editing has expanded drastically over the past several years, presenting an exciting opportunity to manipulate gene expression without altering the underlying genetic code. Unlike classic genome editors, which create permanent changes to the genome, epigenetic editors consist of a nuclease-dead Cas9 (dCas9) and a guide RNA (gRNA) fused to an epigenetic effector domain. Although classic genome editors have long been debated in the literature, epigenetic editing has not warranted the same level of attention. To understand how stakeholders perceive emerging issues related to genome editing use and governance, we conducted 92 semi-structured interviews with gene editing scientists and policy professionals from 23 different countries. One emergent issue that we did not anticipate was the variation in how these interviewees discussed the ethical considerations of epigenetic editing compared to classic genome editing. Interviewees identified unique benefits, such as the “naturalness” of

epigenetic editing compared to classic genome editing, and unique risks, such as the potential for more far-reaching off-target effects than classic genome editing. Interestingly, interviewees also presented unique scenarios for epigenetic editing usage that extend beyond typical therapeutic genome editing scenarios. These results implicate a gap in the larger discussion around the ethics of genetic therapies, which we have used as a starting point for an ethical framework that could be applicable when developing epigenetic editing therapeutics. By incorporating the considerations of stakeholders in tandem with technical and ethical considerations, we can create a roadmap grounded in data that can be addressed before epigenetic editing proceeds to the clinic.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7056T: Assessing access to clinical genetic services among young children with birth defects in Texas based on parental recall

Authors:

D. Betancourt[1], C. Shumate[1], C. Yantz[1], H. Gandhi[1], M. Drummond-Borg[1], C. Singletary[2], D. Riconda[3], A. Agopain[4]; [1]Texas Dept. of State Hlth.Services, Austin, TX, [2]The Univ. of Texas MD Anderson Cancer Ctr. UTHlth.Houston Graduate Sch. of BioMed. Sci., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Baylor Coll. of Med., Dept. of Molecular and Human Genetics, Houston, TX, [4]Dept. of Epidemiology, Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract:

Background: Birth defects are known to be caused by genetic factors. Genetic testing and counseling can help families of children with birth defects understand their child's condition and guide their medical treatment. The Birth Defects Epidemiology and Surveillance Branch (BDESB) routinely connects young children from the Texas Birth Defects Registry (TBDR) born with select birth defects to agency social workers for assistance accessing health and service programs. **Methods:** Beginning in November 2022, social workers asked parents they contacted from the TBDR: "Has the parent or child met with a geneticist or genetic counselor (or visited a genetics clinic)?" Social workers documented whether the parents responded "yes" or "no" on a data collection form that was returned to BDESB. Responses were tabulated, overall and by the child's phenotype (spina bifida or encephalocele, cleft lip or cleft palate, Down syndrome, critical congenital heart defects), time of birth defect diagnosis (during pregnancy, at birth/delivery, after delivery), insurance status at time of contact, and select maternal characteristics. Additionally, a list of genetics clinics (prenatal, pediatric, cancer, specialty) in

Texas was compiled and used to assess if responses varied between parents residing in state regions with low (0-5 clinics), medium (6-10 clinics), or high (11+ clinics) number of genetic clinics. A chi-square test was used to assess if these characteristics were significantly associated ($p<.05$) with parent's reported access of genetic services.
Results: Among the 400 parents reached by social workers, 261 (65%) reported that they or their child had not seen a geneticist, genetic counselor, or visited a genetics clinic. Significantly higher percentages of parents reporting that they or their child had not seen a geneticist, genetic counselor, or visited a genetic clinic were observed among those whose child had spina bifida or encephalocele (78%), those with no insurance (100%), and among mothers residing along the Texas-Mexico border (76.5%). Parents who resided in regions with a higher number of genetic clinics (11+) were significantly more likely to report they had met with a geneticist, genetics counselor, or visited a genetics clinic compared to parents residing in regions with fewer clinics (50% vs. 37%).
Conclusion: In Texas, more than half (65%) of parents of children with select birth defects reported that neither they nor their child has met with a geneticist, genetic counselor, or visited a genetics clinic. Our findings highlight population subgroups with particularly low access. Further research on barriers to accessing genetic services is warranted.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7057T: Barriers to genetic research enrollment in community settings: Lessons from SeqFirst recruitment

Authors:

O. Sommers[1], K. MacDuffie[2], T. Theoryn[3], P. Murali[3], K. Anderson[1], A. Scott[4], M. Sikes[4], C. Davis[5], K. Buckingham[1], K. McWalter[6], P. Kruszka[6], J. X. Chong[7,5], J-H. Yu[8,1,3], T. Wenger[4,1], K. M. Dipple[4,1], D. Doherty[1,4], M. Bamshad[1,4,5]; [1]Dept. of Pediatrics, Univ. of Washington, Seattle, WA, [2]Seattle Children's Res. Inst., Seattle, WA, [3]Inst. for Publ. Hlth.Genetics, Univ. of Washington, Seattle, WA, [4]Seattle Children's Hosp., Seattle, WA, [5]Brotman Baty Inst. for Precision Med., Seattle, WA, [6]GeneDx, Gaithersburg, MD, [7]Univ. of Washington, Dept. of Pediatrics, Seattle, WA, [8]Univ. of Washington Sch. of Med., Seattle, WA

Abstract:

SeqFirst is a research study with a goal of increasing early and equitable access to a precise genetic diagnosis for critically ill neonates and children with intellectual or developmental disabilities. SeqFirst-neo enrolls newborns from a neonatal intensive care unit (NICU) at Seattle

Children's Hospital (SCH), while SeqFirst-DDi enrolls children with developmental delay from an SCH neurodevelopmental clinic (NDV) and two local Early Intervention Centers (EICs). In 14 months of recruitment through the NICU (Jan 2021-Feb 2022), 126 participants were enrolled (average 9.0/month, range 1-13). After 31 months of NDV recruitment (Oct 2021-May 2024), 117 participants were enrolled (average 3.7/month range 0-9). Notably, after 38 months of EI recruitment (Mar 2021-May 2024), 11 participants were enrolled from one EIC, and 1 from the second EIC, for a combined average rate of 0.3/month (range 0-2). Additionally, a new SeqFirst effort to expand enrollment from a community-based pediatric clinic has struggled to gain traction, despite enthusiastic organizational support. Once referred to SeqFirst-DDi, participants from different settings enroll at similar rates (EI: 60%, NDV: 47%), suggesting that the EIC enrollment barrier stems from provider referral. To explore provider differences, we performed qualitative interviews with neonatologists (n=8), NDV providers (n=11), and EI providers (n=1; enrollment challenges at EICs were also experienced when recruiting providers for interviews) involved in SeqFirst recruitment. The semi-structured interview guide asked about standard pathways for obtaining genetic services and provider experience with and attitudes towards SeqFirst. Comparing recruitment processes and analyzing recruitment facilitators reported in NEO and NDV interviews may reveal possible recruitment barriers at EICs. In the NICU and NDV, SeqFirst staff review the electronic medical record (EMR), identify possibly eligible patients, and alert providers to which patients should be approached for enrollment. In addition, whole genome sequencing (WGS) results are recorded in the EMR, allowing NEO and NDV providers to see and benefit from clinically relevant genetic results. Furthermore, SeqFirst investigators are members of SCH medical staff who have personal relationships with providers and serve as SeqFirst "champions" to promote study awareness. The facilitating role of champions was referenced in both NEO (n=6) and NDV (n=4) interviews, while there was no champion embedded in EICs. Further research on barriers in community settings is required to improve the implementation of genetic testing in the future.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7058T: Bridging the gap: Patient and healthcare professional perspectives on accessing and using cancer genetics services by racialized and ethnic communities

Authors:

S. Grewal[1,2], J. Sam[2], E. Reble[2], V. Aguda[2], A. Nsenga[2], S. Shah[2], M. Clausen[2], M. Aronson[3], K. Bell[4], K. Calpu[5], M. DeCoteau[6], A. Eisen[7], H. Etchegary[8], H. Feilotter[9], C. Hessels[10], F. Holness[5], C.

Law[5], A. Lofters[11], H. Nabi[12], S. Panchal[13], N. Persaud[14], T. Reece[5], A. Sayani[15], T. Tiano[16], A. Tone[17], Y. Bombard[1,2], Equity in Cancer Genetics Study Team; [1]Inst. of Hlth.Policy, Management & Evaluation, Univ. of Toronto, Toronto, ON, Canada, [2]Genomics Hlth.Services & Policy Res. Program, Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [3]Zane Cohen Ctr. for Digestive Diseases, Mount Sinai Hosp., Toronto, ON, Canada, [4]Provincial Genetics Program, Ontario Hlth., Toronto, ON, Canada, [5]Equity in Cancer Genetics Patient Partner, St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [6]Rethink Breast Cancer, Toronto, ON, Canada, [7]Odette Cancer Ctr., Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [8]Clinical Epidemiology Unit, Faculty of Med., Mem. Univ., St. John's, NL, Canada, [9]Pathology and Molecular Med., Queen’s Univ., Kingston, ON, Canada, [10]Li-Fraumeni Syndrome Association Canada, Nanaimo, BC, Canada, [11]Peter Gilgan Ctr. for Women’s Cancers, Women’s Coll. Hosp., Toronto, ON, Canada, [12]Oncology Div., CHU de Québec-Université Laval Res. Ctr., Québec City, QC, Canada, [13]Marvelle Koffler Breast Ctr., Mount Sinai Hosp., Toronto, ON, Canada, [14]MAP Ctr. for Urban Hlth.Solutions, Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [15]Equity-Mobilizing Partnerships in Community, Women’s Coll. Hosp., Toronto, ON, Canada, [16]My Gut Feeling – Stomach Cancer Fndn. of Canada, Toronto, ON, Canada, [17]Ovarian Cancer Canada, Toronto, ON, Canada

Abstract:

Introduction: Genetic screening of high-risk individuals is critical for early cancer detection and risk reduction, as well as cascade testing for at-risk relatives. Yet, racial and ethnic disparities exist at every stage of the genetics service pathway. Racialized individuals are less likely to be referred for genetic testing and are more likely to receive inconclusive results due to underrepresentation in reference genomic databases. This limits their access to high-risk cancer screening and prevention, which may ultimately affect health outcomes. Limited efforts in Canada to identify racial disparities impedes the development and implementation of interventions to reduce these inequities.
Aim: We aim to explore the experiences, needs, barriers and enablers to accessing and using cancer genetics services among racialized patients and healthcare professionals (HCPs).
Methods: Semi-structured interviews are being conducted with racialized patients and HCPs involved in cancer genetics care across Ontario. Analysis follows an interpretive description approach.
Results: To date, we have interviewed 9 racialized patients and 7 HCPs (target: n=30 patients and 30 HCPs, to be completed by October 2024). Patients and HCPs noted that inequities are based on more than racial identity; socioeconomic status, geographic location and education level were mentioned as other intersecting factors impacting awareness of and access to genetics care. Both groups also noted the shortage and lack of diversity in the genetics workforce as factors exacerbating inequities. Patients felt the need to be their own advocates

when obtaining a referral, while referring HCPs noted challenges due to unclear eligibility criteria. Both discussed difficulties during family history collection due to missing information or secrecy around cancer in families. HCPs discussed the importance of cultural awareness regarding family structure and dynamics. They also felt that more could be done at the institutional level, and that Diversity, Equity and Inclusion training offered by institutions were a form of performative diversity. As a way to improve access to genetics services, HCPs suggested mainstreaming, which is genetic testing initiated and ordered by a non-genetics specialist. **Conclusion:** Multi-level, intersectional barriers exist to accessing cancer genetics services for racialized patients such as challenges obtaining a referral and limited family history knowledge. Suggested strategies to increase access include mainstreaming and improved cultural awareness among HCPs. This work will be used to co-develop interventions to optimize access and reduce disparities.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7059T: Cancer Epigenetics through a Sex and Gendered Lens: Reflections on a Comprehensive Scoping Review

Authors:

N. Palmour[1], K. Huerne[1], S. S. Jackson[2], R. Lall[1], A. Berner[3], C. Dupras[4], Y. Joly[1]; [1]McGill Univ., Montreal, QC, Canada, [2]NIH, Rockville, MD, [3]Barts Cancer Inst., Queen Mary Univ. of London, London, United Kingdom, [4]Univ. of Montreal, Montreal, QC, Canada

Abstract:

In the study of epigenetic mechanisms for various types of cancer, sex and gender are important considerations, impacting cancer susceptibility, prognosis, and therapeutic responses to cancer interventions. Epigenetics-based sex differences in cancer are particularly significant as they influence gene regulation and expression, ultimately affecting cancer outcomes. These sex differences of epigenetics may arise from a combination of inherited, behavioral, and environmental factors, including gender-based differences. Thus, both sex and gender are key considerations in epigenetics-based cancer research. Despite its importance, there has been minimal assessment of the state of sex and gender-based analyses (SGBA) in this field. For its unique ability to integrate environmental effects as biological states, epigenetics has the potential to help explain sex/gender-based differences in cancer expression. Scientific findings shape and are shaped by existent discourse on sex and gender, as such it is important to be attentive to the interpretation and communication by researchers of sex/gender findings in a

responsible, equitable, and representative manner.
This presentation reports the results of a PRISMA-ScR scoping review conducted with 111 peer-reviewed studies encompassing colorectal, gastric, head and neck, hepatocellular carcinoma, and lung cancers. The research critically evaluated the utilization of sex and gender data in the field of cancer epigenetics, in terms of its representation, discussion and data analysis since 2010. We document that in studies involving sex/gender as a biological or social variable, researchers often neglect to provide a working definition of sex/gender in their study or omit discussions on how sex/gender differences may impact clinical outcomes altogether. For historically underrepresented or misrepresented populations (e.g., women, intersex, transgender persons) in health research ensuring both a cognizant representation and analysis of sex/gender differences is critical from an ethical perspective to shaping future discourse on the matter. The need for clear guidelines on incorporating sex and gender as variables in epigenetics research is emphasized.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7060T: Community Engagement in Practice: A Case Study with the Pukapuka Community in Aotearoa, New Zealand

Authors:

O. Gray[1], J. Moors[1], T. Merriman[2,3], N. Rapana[4], R. Rapana[4], C. Beumelberg[4,5], M. Hendershott[1], K. Wasik[1]; [1]Variant Bio, Inc., Seattle, WA, [2]Dept. of Microbiol. and Immunology, Univ. of Otago, Dunedin, New Zealand, [3]Div. of Clinical Immunology and Rheumatology, Univ. of Alabama at Birmingham, Birmingham, AL, [4]Pukapuka Community Ctr., Auckland, New Zealand, [5]Etu Pasifika Auckland Ltd, Auckland, New Zealand

Abstract:

It is a well-documented failing of genomic research, that even as genomic studies and population biobanks increase in size and scope, they remain predominantly of European ancestry resulting in an incomplete understanding of the impact of human genetic variation on health and disease. Indeed, fewer than 1% of all GWAS include participants from Oceania, a proportion which decreases still further in functional genomics studies. This underrepresentation cannot be addressed merely through increased recruitment efforts. There is often understandable skepticism amongst Indigenous and minority populations about participating in genomics research, which, all too often, has perpetuated rather than resolved group harms and inequities. Long-term, community engagement throughout the lifespan of a

study - design, implementation, and results return - is of pivotal significance in cultivating a locally meaningful partnership, thus fostering enhanced equity and justice for individuals involved in genomic studies. Here, we present a case study of one such collaborative approach to genomics research undertaken with the Pukapuka community in Aotearoa New Zealand. Researchers from academia and industry, in the US and Aotearoa NZ engaged in a collaborative partnership with the Pukapuka Community to develop a functional genomics study which reflected community health priorities. The partnership emphasized building trust and empowering communities through in-depth community engagement, respecting cultural protocols, equitable benefits sharing, and the return of aggregate genomic and health findings to the community. Additionally, the study committed to building local capacity by engaging and training a Pukapuka community-based implementation team who successfully recruited over 300 people into a project which included a comprehensive biomedical health assessment, whole genome sequencing, as well as paired transcriptomics and metabolomics. A functional genomic study of this scale has never been carried out within the Pukapuka population and will greatly enhance our understanding of the functional impact of previously uncharacterized genetic variation. Importantly, following completion of sampling, the process of community engagement remains ongoing to ensure that community stakeholders guide the process for the return of results in a way that aligns with locally identified health concerns.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7061T: Comparative impact of rapid whole genome sequencing in pediatric and intensive care units in Quebec pediatric hospitals: the PRAGMatIQ study ⭐

Authors:

C. Varin-Tremblay[1], S. Levesque[2], J-f. Soucy[1], L. Witkowski[3], M-A. Dugas[4], C. Maftei[1], N. Anastasio[2], F. F. Hamdan[1], J. Gauthier[1], A-M. Laberge[1], J. L. Michaud[1]; [1]CHU Sainte-Justine, Montreal, QC, Canada, [2]CHU de Sherbrooke, Sherbrooke, QC, Canada, [3]McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [4]CHU de Québec-Université Laval, Québec, QC, Canada

Abstract:

Background: Rapid whole genome sequencing (rWGS) has been recognized for its potential to increase the diagnostic yield of the investigation of critically-ill children with a suspected rare genetic disorder. Most studies to date have focused on the application of rWGS in intensive care units (ICUs), demonstrating its ability to provide timely and actionable genetic diagnoses that

can significantly alter patient management and outcomes. The impact of rWGS in other pediatric care settings remains less explored.
Objectives: This study aims to fill the gap by comparing the diagnostic yield, time to diagnosis, and clinical management outcomes of rWGS in pediatric units and ICUs in pediatric university hospitals across Quebec.
Methods: Children admitted to inpatient units of one of Quebec's four pediatric university hospitals with an unexplained clinical presentation and suspected monogenic condition are eligible to participate in the study. At term, 750 trios will be recruited by medical geneticists over a three-year period and will receive a rWGS. A review of medical records and questionnaires for parents and clinicians will be used to assess the medical, psychosocial and economic impacts of rWGS. Recruitment and data collection is ongoing.
Results: As of June 2024, rWGS was completed in 337 patients and their parents with an overall diagnostic yield of 34%. 222 of those patients (66%) were recruited in the ICUs and 115 (34%) were recruited in pediatric units. Preliminary findings indicate a molecular diagnostic yield of 31% vs 42%, respectively. For the 74 patients with a positive molecular diagnosis for whom data collection is completed (45 ICU patients and 29 pediatric unit patients) clinical management was modified in 27 ICU patients (60%) and in 20 pediatric unit patients (69%). A familial impact was observed in 21 (28%) of those families. Conclusion: The implementation of rWGS in pediatric care units has demonstrated substantial benefits for hospitalized patients. These findings underscore the critical role of rWGS in improving clinical outcomes for hospitalized pediatric patients beyond ICUs and highlight the necessity of broader adoption and implementation of rWGS into pediatric acute care settings. Expanding the use of rWGS in other pediatric settings will ensure that more children with genetic conditions receive the prompt and accurate diagnoses they need.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7062T: Delayed diagnosis in neurogenetics disorders in Peru

Authors:

M. Illanes-Manrique[1,2], **M. Silva-Bullon**[1], B. Toledo-Pacheco[1], R. Yrene-Cubas[1], M. Galecio-Castillo[1], E. Sarapura-Castro[2,1], A. Rivera-Valdivia[2,1], M. Cornejo-Olivas[3,2]; [1]Neurogenetics Working Group,Univ. Cientifica del Sur, Lima, Peru, [2]Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [3]Neurogenetics Working Group, Univ. Cientifica del Sur, Lima, Peru

Abstract:

Background: Genetic diseases often face delayed diagnosis due to clinical, sociodemographic, and economic factors. Few studies analyze delayed diagnosis in rare diseases around the world. Low-middle-income countries like Peru have significant difficulties accessing health services, which exacerbates delays in diagnosing rare disorders, including neurogenetic diseases.Objective: To determine the factors associated with delayed diagnosis of neurogenetic diseases in patients followed at a specialized tertiary neurogenetic center in Peru.Methods: Cross-sectional study carried out at the Neurogenetics division of a national reference center for neurological disorders in Peru. A standardized questionnaire was applied to individuals older than 18 years of both genders affected by neurogenetic diseases confirmed by appropriate genetic testing. This study was approved by the local ethics committee at INCN.Results: A total of 86 participants (54.7% male, 39.5% married) completed the questionnaire. The mean age at assessment was 44.2 ± 12.4 years. The average years of education were 12.16 ± 3.1. About 55.8% of them live in Lima, the capital city. Approximately 66% of them are formal employees, and almost all of them (83 out of 86) have some type of public health insurance. The genetically confirmed neurogenetic disorders included: 45 participants (52.3%) with Huntington's disease, Myotonic Dystrophy type 1 (24.4%), dominant ataxias (8.1%), recessive ataxias (4.6%), Neurofibromatosis type 1 (3.5%), and 6 participants (6.9%) with other even rarer disorders. The mean age at onset was 33.6 ± 14.4 years. The mean age at first clinical assessment at any healthcare center was 36.337 ± 14.027 years, with most of them (88.4%) seeking consultation at public hospitals. The average diagnostic delay from the first symptom to genetic testing was 7.1 ± 6.9 years. A younger age at the first symptom was the only significant factor associated with delayed genetic diagnosis, adjusted for gender, number of other family members, and type of healthcare services.Conclusions: We determined a mean delay of 7.1 years for genetic diagnosis within a cohort of patients followed at a tertiary neurogenetic center in Peru. A younger age at the first symptom is associated with a more prolonged diagnostic delay. Further studies with larger sample sizes at the national level are required to confirm this observation.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7063T: Democratization and cloud-based genomic data sharing: views of platform developers and early adopters

Authors:

S. Fullerton[1], J. M. Dahlquist[1], S. C. Nelson[2]; [1]Univ. of Washington Sch. of Med., Seattle, WA, [2]Univ of Washington, Seattle, WA

Abstract:

Background: Several new cloud-based computing platforms have been established by the National Institutes of Health to facilitate collaboration and maximize the scientific utility of large-scale genomic and linked biomedical data. Compared with pre-existing data sharing mechanisms, these platforms offer new and potentially more efficient alternatives for sharing and analyzing data, promising to widen access to genomic information and promote data science equity. **Methods**: To better understand platform policies and procedures, we performed semi-structured interviews with 21 developers associated with five cloud platforms: the All of Us Research Hub, the AnVIL, BioData Catalyst, Genomic Data Commons, and the Kids First Data Resource, as well as the existing exemplar dbGaP. To learn more about the experiences of platform users, we interviewed 22 early adopters of the same five platforms. Interviews with developers focused on the design and promise of these new data sharing mechanisms, while interviews with early adopters focused largely on perceived benefits and challenges of cloud-based genomic data sharing and analysis. **Results**: The developers we interviewed routinely invoked the "democratizing" potential of cloud platforms, while also acknowledging current barriers to achieving that promise. Democratization was understood as a platform's ability to broaden or expand data access and sharing, including with non-traditional users and/or institutions. Early adopters were far more equivocal on the democratizing potential of such platforms, with interviewees noting both the steep learning curve and compute costs that users must negotiate, challenges which they felt could interfere with broader adoption. **Conclusions**: These observed differences of perspective among platform developers and early adopters raise important questions about the potential of cloud platforms to broaden both data access and use, and hence to meet claims underlying their substantial public investment. More research is needed to better understand whether and how cloud-based data sharing can widen access to genomic and linked biomedical information, thereby promoting data science equity.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7064T: Disparities in VUS among ethnoracial groups in the National Brain Gene Registry

Authors:

A. Sveden[1], J. Cole[2], S. Mahida[1], V. Lanzotti[3], M. Clarke[4], A. Bican[5], I. Oh[3], A. Gupta[3], B. Minor[3], H. Hazlett[6], D. Baldridge[3], M. Sahin[1], P. Payne[3], Brain Gene Registry Consortium, C. Gurnett[3], M. Chopra[1]; [1]Boston Children's Hosp., Boston, MA, [2]Children's Hosp. Colorado, Aurora,

CO, [3]Washington Univ. Sch. of Med., St. Louis, MO, [4]Albert Einstein Coll. of Med., New York, NY, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Univ. of North Carolina, Chapel Hill, NC

Abstract:

Background
Barriers to genomic testing for historically marginalized ethnoracial populations are well documented. Such groups also face higher variants of uncertain significance (VUS) burden compared to non-Hispanic/Latine White (White) patients, likely due to lack of diversity in population databases, suboptimal test selection by providers, and limited access to parental testing. Here, we describe revisions to ethnoracial descriptors for an intellectual and developmental disability (IDD) research registry, the Brain Gene Registry (BGR), and assess disparities relating to VUS among ethnoracial groups.

Methods
The BGR pairs genotypic and phenotypic data from individuals with clinical variants (pathogenic, likely pathogenic [PLP], or VUS) in IDD genes across 12 IDD Research Centers to advance IDD research. Following a literature review of recent guidelines and research group discussion, BGR ethnoracial descriptors were revised to be more inclusive, and self-reported identities were remapped to these descriptors for the first 585 BGR participants. Alongside ethnoracial descriptors, we analyzed the proportion of variants classified as VUS (VUS/VUS + PLP) and, for VUS, test type and inheritance.

Results
BGR ethnoracial descriptors originally followed minimum NIH requirements. These were revised to use inclusive language, add Middle Eastern/North African, and add subcategories for “Asian.” Revisions also allowed participants to select multiple ethnoracial descriptors.
With these revised descriptors, the first 585 BGR participants were described as 6% Asian, 5% Black/African American/African, 10% Hispanic/Latine, 1% Middle Eastern/North African, 8% Multiracial/multiethnic, 70% White, and 5% Unknown. Overall, VUS accounted for 225/585 (38%) of variants. There was a higher proportion of VUS among non-White groups (45%, 74/164) than White counterparts (32%, 125/389, $p = 0.004$). The Black/African American/African group had the highest proportion of VUS, with VUS accounting for 16/30 (53%) of variants. Within the VUS cohort, 20% (15/74) of VUS in non-White groups were identified through trio testing versus 37% (46/125) in the White group ($p = 0.011$). Only one non-White individual harbored a de novo VUS, compared to 17% (21/125) of White individuals ($p < 0.001$).

Discussion
Ethnoracial descriptors in the BGR were revised to align with recent recommendations and guidelines. The higher proportion of VUS in non-White groups is likely multifactorial, with

barriers to optimal test selection and parental testing as significant contributors. Moving forward, additional efforts for inclusivity and resolving VUS for non-White groups are critical.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7065T: Evaluating patient-centric care in cancer genetics counselling for HBOC syndrome using a measure of informed choice

Authors:

Z. Chua[1], N. Binte Ishak[1], Z. Zhang[1], J. Chiang[1], J. Ngeow[2,3]; [1]Natl. Cancer Ctr. Singapore, Singapore, Singapore, [2]Natl. Cancer Ctr. Singpore, Singapore, Singapore, [3]Nanyang Technological Univ., Lee Kong Chian Sch. of Med., Singapore, Singapore

Abstract:

Background The decision to undergo cancer genetic testing is complex, as the results have cascading implications to both the patient and their family. While cascade testing for hereditary breast and ovarian cancer (HBOC) syndrome is a tier 1 genomic application, there is a need to understand if patients are making an informed decision to undergo testing. Hence, we aimed to assess a measure of informed choice for genetic testing.
Methods We designed a survey to assess patient's knowledge and attitudes towards HBOC testing and captured their uptake of testing. An informed decision was defined as one made with adequate knowledge, and when attitudes are congruent to actual uptake of genetic testing. We also assess the presence of decisional conflict using the decision conflict scale for genetic testing.
Results We recruited 217 patients of which 189 (87.1%) accepted testing. Testers (74.2%) were more make an informed decision for testing than non-testers (8.0%), $X^2(1, N=194) =43.10$, $p<.01$. While difference is knowledge was not significant, attitudes scores towards testing was more positive for testers than non-testers, with non-testers having more ambivalent attitudes towards testing. Associations between financial assistance and testing uptake was found $X^2(1, N=201) =12.60$, $p<.01$. No difference in decisional conflict was found.
Conclusions Proportion of participants having adequate knowledge on HBOC testing was high (73.3%) post counselling. Informed decisions were predominantly influenced by attitudes, and the presence of financial assistance moderated the relationship. Greater financial assistance and addressing attitudes towards testing can help patients made better informed decision for HBOC testing.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7066T: Family Information Disclosure and Testing Practices within the Context of a Population Biobank: an Analysis of Survey Results from the Colorado Center for Personalized Medicine Biobank (CCPM Biobank)

Authors:

E. Todd, L. Axell, L. Ku, The Colorado Center for Personalized Medicine, K. Crooks; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Introduction The CCPM biobank provides research participants clinical genetic test results from its CAP accredited laboratory. When these results are shared with biobank participants, discussion includes implications for family members and recommendations to share this information. A follow-up survey is offered to participants to measure outcomes including disclosure to family members and cascade testing. Reported rates of information dissemination and cascade testing within families vary and it is important to understand if these rates in the setting of biobank-based population screening differ from those in traditional clinical settings. This abstract describes survey results from biobank participants focusing on responses related to family information disclosure and cascade testing practices. **Methods** CCPM biobank participants consent to receive clinical genetic test results, which are returned via a phone call with a genetic counselor. Discussion includes implications for the individual and their family members. Six or more months after receiving results, biobank participants are surveyed via the REDCap electronic database to collect information on outcomes and experience. **Results** As of April 2024, 377 biobank participants who have received results were sent surveys with 77 responses (21%). When asked if results were shared with family members, 5 (6%) answered ‘no’, 2 (3%) answered 'not yet but planning on in future’, 1 (1%) answered 'prefer not to answer’, 1 (1%) was ‘unsure’ and 68 (88%) answered ‘yes’. The average number of family members informed of the result was 2.5. When asked if family members pursued genetic testing, 38 (49%) answered ‘no’, 13 (17%) answered ‘unsure’, 17 (22%) answered ‘yes’, and 9 (12%) did not answer. When asked if family members saw their providers after being informed of biobank participant’s result, 35 (45%) answered ‘no’, 17 (22%) answered ‘unsure’, 15 (19%) answered ‘yes’, and 10 (13%) did not answer. **Conclusions** Of the biobank participants who received clinical genetic results, the majority, 88%, shared these results with relatives. The rate of cascade genetic testing noted in families of our biobank participants was 22%. Our study indicates that providing this information in a biobank setting leads to high levels of information

dissemination within families. The rate of cascade genetic testing in families is expected to increase over time as more family members access genetic testing.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7067T: Forced resilience: Indigenous people's perspectives on humanizing diagnostic genomic testing in British Columbia, Canada

Authors:

M. Ehman[1], S. Pollard[1,2], L. Montour[3,4], D. Weymann[1,2], D. Kirk[5,4], K. Brown[4], S. Epp[4], D. Wadsworth[4], N. A. Garrison[6], A-M. Laberge[7], J. Reading[2], N. R. Caron[8], L. Arbour[8,9], D. A. Regier[1,8]; [1]BC Cancer Res. Inst., Vancouver, BC, Canada, [2]Simon Fraser Univ., Vancouver, BC, Canada, [3]Mohawk Nation of Kahnawake, Kahnawake, QC, Canada, [4]Indigenous Advisory Council, Vancouver, BC, Canada, [5]Vancouver Community Coll., Vancouver, BC, Canada, [6]Univ. of California, Los Angeles, Los Angeles, CA, [7]CHU Sainte-Justine, Service de génétique médicale, Montreal, QC, Canada, [8]Univ. of British Columbia, Vancouver, BC, Canada, [9]Univ. of Victoria, Victoria, BC, Canada

Abstract:

Purpose: Healthcare services and diagnostic benefits of genomic testing (GT) are inequitably available to Indigenous children with suspected rare diseases. As a result, Indigenous children may experience an extended diagnostic odyssey. We sought to qualitatively understand factors contributing to this delayed diagnosis and identify opportunities to create valued and inclusive GT services within British Columbia (BC), Canada, from the perspective of Indigenous families.
Methods: We conducted virtual interviews and focus groups using a talking circle approach with Indigenous adults residing in urban settings across BC. We developed the study protocol through partnership with the BC First Nations Health Authority. Our Indigenous Advisory Council (IAC) further guided protocol and instrument development, analysis, and interpretation. Participants provided feedback on preliminary analyses via optional member checking surveys. We co-analyzed transcripts alongside the IAC using thematic analysis.
Results: Between February and March 2022, twenty-four First Nations and/or Métis individuals participated across five focus groups and one interview. Nine participants provided member checking feedback. Thematic analysis yielded two major themes describing participants' perceptions toward GT as a clinical service embedded within BC healthcare systems. Participants highlighted resilience as an imposed and normalized aspect of healthcare, identifying a lack of accessible and safe care, demands of self-advocacy, and experiences of

racism impeding willingness to undergo GT. Participants recommended humanizing practices grounded in respectful relations, advocacy support, cultural safety, and Indigenous perspectives on health and wellness.

Conclusions: Colonialism, systemic discrimination, and racism force Indigenous people's resilience in healthcare settings. In addition to physical access barriers, disrespectful interactions and distrust delay healthcare access, exacerbate data- or diagnosis-related harms, and create an unmet need for patient supports. These factors impede GT accessibility. Redressing this paradigm requires thoughtful and reconciliatory action across both physical and relational barriers. Humanizing healthcare practices offer one approach, defined by collective intent toward creating welcoming and safe spaces for Indigenous families. Building relations, increasing representation of Indigenous healthcare professionals, and responsibly integrating cultural services and understanding into service delivery may improve testing uptake and satisfaction among Indigenous families.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7068T: Improving Access to High-quality Healthcare for Incarcerated Patients: The Role of the Genetic Counselor ⭐

Authors:

A. Abram[1], R. Aatre[1], S. Roberts[1], B. Yashar[1], N. Waligorski[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Seattle Children's Hosp., Seattle, WA

Abstract:

Arrests in the United States are concentrated in low-income, predominantly non-white communities that are more likely to be medically underserved. With the expanding utility of genetics-based healthcare, it is salient to explore how a patient's incarceration status impacts the provision of this specialized care. To address this question, we evaluated genetic counselors' (GC) perceived self-efficacy to provide high-quality comprehensive healthcare to incarcerated patients. A four-part online survey, (1. demographics, 2. experiences with incarcerated patients, 3. & 4. perceived self-efficacy (SE) to provide care to an incarcerated patient and a non-incarcerated patient) based on the ACGC practice-based competencies and the GC self-efficacy scale (GCSES), was created and shared with the GC community. Results from 122 completed GC responses were evaluated using descriptive statistics and linear regression analysis. 61% of these respondents reported having no experience providing care to incarcerated patients and 65% of these GCs anticipated differences in their ability to counsel an incarcerated patient

compared to a non-incarcerated patient, with the greatest anticipated impacts involving resource identification, referrals and psychosocial care. In regression analysis, predictors of increased SE included experiences observing or providing care to incarcerated patients and receiving graduate or postgraduate training on providing care to incarcerated patients and accessing relevant resources. A p-value threshold ($p<0.05$) was applied for significance while a separate p-value threshold ($p<0.01$-0.05) was applied for suggestive relationships. Those who received incarcerated patient care training during graduate school showed a higher self-efficacy in GC skills involving resource identification, referrals and psychosocial care than those who did not (p-value 0.054). Among GC respondents who had provided care to at least one incarcerated patient, primary indications included cancer, prenatal, and cardiology and in a one-tailed t-test, self-efficacy was positively impacted for every additional incarcerated patient seen (p-value: 0.056, F-statistic 3.91). Our findings suggest that a disparity exists in GCs' ability to provide care for incarcerated patients and highlights potential educational deficits that may be addressed to help GCs navigate suggested barriers and provide more equitable care to our population of incarcerated patients.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7069T: Outcomes of clinical genetic testing for patients with kidney cancer referred to genetic counseling

Authors:

J. Wolfe[1], L. Byrne[1], M. Hussein[2], M. Abdel-Rahman[2,1]; [1]Div. of Human Genetics, Dept. of Internal Med., and Comprehensive Cancer Ctr., Columbus, OH, [2]Dept. of Ophthalmology and Visual Sci., The Ohio State Univ., Columbus, OH

Abstract:

Objective: 5-10% of kidney cancers are hereditary. In 2020, the National Comprehensive Cancer Network (NCCN) established guidelines specific to hereditary kidney cancer. At The Ohio State University Wexner Medical Center a genetics genitourinary clinic was initiated in 2019 to specifically take care of kidney cancer patients. This study was done to evaluate the outcomes of referrals to genetic counseling. We also evaluated germline mutation rates in an unselected cohort of kidney cancer patients from our institution. Methods: A retrospective chart review of kidney cancer patients who were seen at The Ohio State University Medical Center and were referred by their providers to Clinical Cancer Genetics based on their personal or family history of cancers. Genetic testing for these patients was part of their clinical care. We also tested an

unselected cohort of 154 kidney cancer patients seen at similar intervals and enrolled through the Total Cancer Care (TCC) protocol. Genetic testing for the second cohort was carried out in our research laboratory using a next generation sequencing panel of 100 cancer genes and results were confirmed by direct (Sanger) sequencing. Results: Out of the 388 patients referred to genetics, 276 (71%) attended their appointments. Out of those 237 (85.8%) were tested and the remaining declined. Additionally, majority of the genetics referrals were after the establishment of a GU clinic. Most of the patients referred to clinical genetics met the current NCCN criteria. Of those who were tested 40 (16.9%) had pathogenic/likely pathogenic (P/LP) variant in a known cancer gene including *ATM, AXIN2, BAP1, BRCA2, CHEK2, EGFR, FH, FLCN, HOXB13, LZTR1, MLH1, MSH2, MSH3, MSH6, MUTYH, NF1, NTHL1, PALB2, POT1, PTEN, SDHB, TP53, TSC1*, and *VHL*. For the unselected patient cohort 12/154 (7.8%) had P/LP variants in one of the known cancer genes including: *ATM, BLM, BRCA1, FANCC, FLCN, MUTYH, NBN, RAD50, RECQL4,* and *TP53*. Conclusion: The current NCCN criteria are sensitive for the detection of patients at high-risk for genetic predisposition to kidney cancer. The association of many of the cancer genes with kidney cancer needs to be further assessed. Access to a specific GU genetic counselor increased referrals for those with kidney cancers at The Ohio State Wexner Medical Center. We need to evaluate genetic appointment declines for themes that our GU patients are encountering to further genetic testing accessibility.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7070T: Pathways to precision cancer care: integrating genetic services into a nationwide cancer program at scale ⭐

Authors:

K. Tangney, D. Brockman, L. Servais, **C. Neben**, A. Zhou, S. Topper; Color Hlth., Burlingame, CA

Abstract:

Genetic testing and counseling for cancer predisposition have become more common in clinical settings, though access remains limited to patients who seek it and have resources to complete it. Here, we describe a novel end-to-end comprehensive, employer-sponsored cancer program - the Color Virtual Cancer Clinic - in which clinical genetic testing is coupled with assistance completing risk-adjusted cancer screening allowing for early detection, and present the outcomes of the first 1,900 participants. As part of this program, some employers offered genetic testing based on National Comprehensive Cancer Network (NCCN) guidelines, while others offered testing to their entire population. Within employer groups that offered genetic

testing to all, 55% met NCCN-based criteria for genetic testing, most commonly due to a family history of cancer. A total of 54% (n=1,018) participants completed genetic testing. Of those, 2.75% (n=28) were found to have a pathogenic or likely pathogenic variant in one of 29 actionable cancer genes, with ATM, BRCA2, and CHEK2 mutations being the most common. About one third (32%; n=9) of mutations were in participants who wouldn't have met NCCN testing criteria, highlighting the impact of accessible genetic testing on risk stratification and precision cancer care. All participants were offered post-test phone genetic counseling, virtual physician consults, and cancer care navigation. 167 participants with genetic test results completed physician consults (7% [n=11] positive and 93% [n=156] negative results) and 52 participants completed genetic counseling consults (54% [n=28] positive and 46% [n=24] negative results). Of the 28 participants with positive genetic results, 15 (53%) had gaps in their genetics-informed cancer screening. The Color Care team guided 7 participants with positive genetic results to appropriate management, facilitating 4 breast MRIs, 2 mammograms, 2 at-home PSA tests, 1 upper endoscopy, and 3 referrals to local high-risk breast and gastrointestinal specialists for additional management. The other 8 participants received genetic-informed cancer management counseling through the Color Virtual Cancer Clinic but chose to pursue cancer screening independently. Integrating cancer genetics in this way enables a novel access point to comprehensive cancer risk assessment and screening, reaching individuals through a program that is available remotely and at no-cost to participants. Color's Virtual Cancer Clinic allows broader access to genetic services and providers, allowing for immediate risk-adjusted screening and prevention for high-risk individuals through telemedicine and care navigation.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7071T: Patient perspectives in precision medicine: Identifying knowledge gaps to inform cancer care

Authors:

J. Malinowski[1], W. Peters[2], R. Bailey[1], J. Vanderpoel[3], L. Shea[1]; [1]Johnson & Johnson Innovative Med., Titusville, NJ, [2]Evidera, Bethesda, MD, [3]Johnson & Johnson Innovative Med., Horsham, PA

Abstract:

Background: Genomic literacy is a priority as targeted therapies for cancers and other conditions are developed. Whether patients with these conditions understand genomic concepts sufficiently to actively engage in shared decision-making for testing and treatment choice remains unclear. **Methods**: The Janssen Patient Engagement Research Council (PERC)

represents a diverse group of disease-aware participants living with chronic health conditions who provide their insights and feedback through a structured series of activities. PERCs are organized by condition; the Precision Medicine (PM) PERC was comprised of members with self-reported bladder cancer (BC), lung cancer (LC), and prostate cancer (PC). Four PM PERC focus group sessions were held in August 2023 using a semi-structured discussion guide. The Integrated Behavioral Model was applied to understand participants' beliefs, understanding, and rationale regarding genomic testing. All transcripts were coded in MaxQDA to identify themes and assist with synthesis of insights. Demographics and results are reported with descriptive statistics. **Results**: Twenty participants completed a PM PERC (BC, n=5; LC, n=7; PC, n=8); 12/20 (60%) reported having undergone testing. Participants were majority male (65%), White (55%), ≥65 years old (65%), and 65% had a bachelor's degree or higher. Topic areas explored were individual experience with testing, considerations for testing, precision medicine, and social-emotional support for testing. The potential to inform treatment and find a targeted therapy were considered the most important perceived benefits to testing. However, participants were confused by and unfamiliar with terminology and methods related to testing (e.g., germline, somatic, biomarker testing, precision medicine, fluorescence in-situ hybridization, gene-specific tests). In multiple sessions, patients identified structural barriers (e.g. insurance coverage for testing, integrated genetic counseling, access to testing). Several participants recounted significant family histories of cancer but had not considered or known of hereditary risk even after their diagnosis. Participants who were more recently diagnosed were more likely to report testing discussions with their healthcare provider. **Conclusion**: In this series of focus groups, we observed that terminology related to precision medicine and genomic testing was confusing for persons with cancer. Education, hereditary cancer risk assessment, and integrated genetic counseling may improve testing rates. Patients should be supported to overcome structural barriers to testing, including cost and insurance coverage.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7072T: Protocol for a type 3 hybrid implementation cluster randomized clinical trial to evaluate the effect of patient and clinician nudges to advance the use of genomic medicine across a diverse health system ⭐

Authors:

A. Raper, B. Weathers, T. Drivas, C. A. Ellis, C. Morse Kripke, R. Oyer, A. Owens, A. Verma, P. Wileyto, C. Wollack, W. Zhou, M. Ritchie, R. Schnoll, K. Nathanson; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: Germline genetic testing is recommended for a growing number of conditions, the results of which impact medical management. However, genetic testing is underutilized due to system, patient, and clinician level barriers. Behavioral economics provides a framework to create implementation strategies, such as nudges, to address multi-level barriers. Here we present a protocol for a randomized study of nudge effect on uptake of genetic testing, and the results of rapid cycle approaches for nudge development. **Methods**: Patients meeting genetic testing guidelines will be identified by electronic phenotyping algorithms for 10 conditions. Algorithms have been developed by iterative chart review until achieving ≥85% positive predictive value that a patient has one of the target diagnoses. A pragmatic, type 3 hybrid cluster randomization study will test nudges to patients and/or clinicians, or neither. Patient nudges are sent via text message, and clinician nudges are sent via electronic health record (EHR) as alerts. Clinician nudges prompt to either refer to genetics or order genetic testing. The primary outcome is uptake of germline genetic testing, and patient data from the EHR will be examined for moderators of nudge effect. Rapid cycle approaches informed by clinicians, patients, health equity, and behavioral economics helped optimize nudges. A clinician discrete choice experiment (DCE) tested four messages addressing behavioral health biases (status quo, focusing effect, impact, or omission), that may influence willingness to refer to genetics or order genetic testing. Of 79 clinicians approached, 43 (54%) completed the DCE, with significantly greater preference for a message addressing status quo bias (44%, $p<0.001$). Patient nudges were developed using feedback from two focus groups of five patients each. Four messages targeting biases were presented, and the preferred message addresses focusing effect bias. Participants suggested creating an educational video about genetic testing, which was created and linked in patient nudges. **Discussion**: This study is one of the first randomized trials to examine effects of patient and clinician nudges on uptake of genetic testing. The pragmatic design allows for a large, diverse patient sample, assessment of genetic testing uptake, and comparison of nudge combination effects. The trial involves optimization of patient identification, test ordering, and result reporting in an EHR-based infrastructure, and findings may help determine the impact of low-cost, sustainable implementation strategies that can be integrated into health care systems to improve the use of genomic medicine.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7074T: Psychological effects of a predictive neurodegenerative disease diagnosis

Authors:

T. Vander Groef, D. Burks, N. Minix, J. Paulsen; Univ. of Wisconsin-Madison, Madison, WI

Abstract:

The primary concern among physicians when providing an early predictive diagnosis for a neurodegenerative disorder is the potential for an increase in psychological distress, manifesting through depression, anxiety, or possibly suicide. This study found that predictive genetic testing does not lead to long-term distress, making it safe to disclose an early diagnosis with the proper disclosure techniques. A search was conducted in PubMed and EBSCO using the terms genetic testing, diagnosis disclosure, neurodegenerative disease, Alzheimer's disease, Huntington's disease, psychological effects, depression, and anxiety. 12 sources met the inclusion criteria as they were in English, published from 2014-2024, and all discussed the effects and process of disclosure for predictive testing, primarily for Huntington's and Alzheimer's disease. The results of this study found that when one receives a negative result, indicating they are not likely to develop a neurodegenerative disorder, they typically experience psychological benefits by providing emotional relief and settling curiosity. When the test result is positive, this study concluded that disclosing this information is not harmful to the patient long-term. Many patients report that their depression levels remained unchanged during the 3-month period following their diagnosis. Depression and anxiety symptoms increased for only a minority of patients, with 8% of caregivers stating that an early diagnosis had an overall negative effect on patients compared to 71% that said a diagnosis was neutral. Any increase in psychological distress that does occur following a positive test is typically transient and relatively mild, with few people reporting feelings of regret or catastrophic events, such as suicide. Learning a positive result has clinical benefits, including the ability to plan medical care, inform family members of their potential risk, and help with end-of-life planning, but these benefits may be limited based on the clinical validity of the testing results. This makes it crucial that any results that are disclosed are explained thoroughly using open communication and holistic explanations. This way, patients will feel supported and have a solid understanding of the implications and probability of their results. The consistency among the data indicates that predictive testing for neurodegenerative disorders and disclosing a possible future diagnosis in advance is safe and does not have a significant impact on long-term psychological well-being. The overall goal of this research is to create a universal, tailorable, disclosure procedure that can be applied to each individual condition.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7076T: Receiving Negative Results from DNA-based Population Screening for Adult-onset Diseases: Participants' Experience ⭐

Authors:

F. Russo[1], D. Chatterjee[2], N. DeMaria[2], M. E. Florido[2], M. Marasa[2], M. Sabatello[2], J. Wynn[3], **H. Rasouly**[2]; [1]McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [2]Columbia Univ., New York City, NY, [3]New York Presbyterian Columbia, New York City, NY

Abstract:

DNA-based population screening for adult-onset diseases holds promise for advancing personalized medicine and improving public health. Yet, as most individuals pursuing such screening receive negative results, it is critical that the return of results process assures that negative results and implications are clearly understood. We explored the experiences of adults who received negative results from DNA-based population screening for adult-onset diseases as part of the Electronic Medical Records (EMR) and Genomics consortium Phase 3 project (eMERGE-3) at Columbia University. The eMERGE-3 panel included the 56 genes from the first ACMG actionable finding list and additional genes deemed actionable by the eMERGE-3 consortium. All results were placed in the patients' EMR. Participants with negative results received a laboratory report and a standard counseling letter explaining the meaning and the limitations of their results. In addition, participants were randomized to receive (or not) a vignette depicting a conversation between individuals receiving negative results and their providers. The vignette reading level based on the Flesch-Kincaid test was sixth grade. Participants were asked to complete a baseline survey at time of enrollment and a second survey after learning about the results. A racially, ethnically, and socio-economically diverse cohort of 437 adult participants completed both surveys. Seventy-seven percent were motivated to participate in the study because they hoped to understand the cause of their current clinical condition, even as this is not the goal of population screening. In addition, most participants did not have a condition covered by the eMERGE-3 panel. A quarter of participants reported not feeling confident explaining their results to others (n=105, 24%), and those who did not receive the vignette were less confident than those who did (29% versus 19% respectively; p-value = 0.02). Open text responses about personal and family members' reactions to the results suggested that some participants perceived an exaggerated benefit from their result and did not understand the difference between screening for adult diseases, carrier screening and clinical genetic testing. Our findings highlight the complexity of returning negative results in the frame of screening programs for adult-onset diseases and raise concerns that

participants might forgo more appropriate clinical genetic testing after learning about their negative results. Future research is needed to compare the efficacy, acceptability, and appropriateness of different forms of ancillary materials on individuals' comprehension of negative results.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7077T: Research Participants and Stakeholders views on feedback of Genetics Research Findings: A qualitative study of the H3Africa Kidney Disease Research Network, Ghana.

Authors:

G. Larbie Gafa[1], V. Boima[2,1], I. Rockson[1], S. T. Chatio[3], A. Ghansah[4], B. L. Salako[5], A. Ojo[6], P. Tindana[7], D. Adu[1,2]; [1]H3-Africa Kidney Disease Res. Network, UGMS, Accra, Ghana, [2]Dept. of Med. and Therapeutics, Univ. of Ghana Med. Sch., Coll. of Hlth.Sci., Univ. of Ghana, Accra, Ghana, [3]Navrongo Hlth.Res. Ctr., Ghana Hlth.Service, Navrongo, Accra, Ghana, [4]Noguchi Mem. Inst. for Med. Res., Univ. of Ghana, Legon, Ghana, Accra, Ghana, [5]Coll. of Med., Univ. of Ibadan, Ibadan, Nigeria, [6]Kansas Univ. Med. Ctr., Kansas, Accra, Ghana, [7]Dept. of Hlth.Policy, Planning and Management, Sch. of Publ. Hlth., Coll. of Hlth.Sci., Univ. of Ghana, Accra, Ghana

Abstract:

Genetics and genomics research raise important ethical issues, particularly those related to obtaining valid consent, engaging with relevant communities in the implementation of research and feedback of research findings. There is evidence that variants of the Apolipoprotein L1 gene in Africans and people of African descent increase the risk of developing chronic kidney disease (CKD). A study carried out by the H3Africa Kidney Disease Research Network showed that 28.2% of the study participants carried 2 *APOL1* high-risk variants. It is also known that relatives of subjects with CKD may be at increased risk of developing CKD. Should participants and their relatives be informed about their risk of developing CKD? What findings should be given and who should be involved in the feedback process? The aim of this study was to seek the views of research participants, families and stakeholders on feedback of genetic research findings. This research employed an exploratory qualitative study approach for data collection. The study explored the views of genetic/genomic researchers, participants, families and members of research ethics committees on what count as good ethical practice in deciding what, who and how to return genetic research findings of the kidney disease research in Ghana through in-depth interviews, focus group discussions and deliberative workshops. Data was coded and

structured with NVivo software and thematically analysed. There was a consensus that relevant individual and aggregate genetics results should be fed back to participants and communities. Most participants preferred to receive their personal results from a doctor or a research scientist. There is an ethical imperative to return validated clinically relevant individual genetic research results to the kidney disease research participants and families and aggregate results to communities. We recommend the H3Africa Kidney Disease Research Network to educate participants and families on the concept of risk variants, train Genetic Counsellors and explore innovative strategies to support the feedback process.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7078T: Return processes and one-on-one disclosure uptake of a genome informed risk assessment which includes polygenic risk scores in eMERGE IV

Authors:

J. Odgis[1], K. W. Davis[2], B. M. Boyd[3], A. Allworth[4], M. Head[5], J. Cortopassi[6], J. J. Denton[7], L. Scarimbolo[1], S. Terek[8], H. Bangash[9], K. E. Bonini[1], B. Etheridge[6], M. Hamed[9], V. Hernandez[10], C. Hoell[11], P. N. Marathe[1], E. Perez[12], J. A. Petrzelka[9], N. S. Abul-Husn[1], M. H. Harr[8], G. Jarvik[13], E. W. Karlson[14], E. E. Kenny[1], I. J. Kullo[9], J. E. Linder[15], M. F. Murray[1], C. A. Prows[16], G. L. Wiesner[2], M. Sabatello[17], S. A. Suckiel[18]; [1]Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, [4]Genome Sci., Univ. of Washington, Seattle, WA, [5]Univ. of Washington, Seattle, WA, [6]Translational Pharmacogenomics Program, Dept. of Neurology, Univ. of Alabama at Birmingham, Birmingham, AL, [7]Dept. of Clinical and Diagnostic Sci., Univ. of Alabama at Birmingham, Birmingham, AL, [8]Ctr. for Applied Genomics, Div. of Human Genetics, The Children's Hosp. of Philadelphia, Philadelphia, PA, [9]Dept. of Cardiovascular Med., Mayo Clinic, Rochester, MN, [10]Integrated Collaborative Res., Mountain Park Hlth.Ctr., Phoenix, AZ, [11]Ctr. for Genetic Med., Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, [12]Brigham and Women's Hosp., Brigham and Women's Hosp., Boston, MA, [13]Div. of Med. Genetics, Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [14]Div. of Rheumatology, Inflammation, and Immunity, Dept. of Med., Brigham and Women’s Hosp., Boston, MA, [15]Inst. for Clinical and Translational Res., Vanderbilt Univ. Med. Ctr., Nashville, TN, [16]Human Genetics and Patient Services, Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [17]Ctr. for Precision Med. and Genomics, Dept. of Med.; Div. of Ethics, Dept. of Med. Humanities and Ethics;

Columbia Univ., New York, NY, [18]Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York City, NY

Abstract:

eMERGE IV is a multi-site genomics research program studying the utility of incorporating polygenic risk scores (PRS) for 10 conditions in clinical care. Participants receive a Genome Informed Risk Assessment (GIRA), which collates PRS, monogenic risk, family history, clinical risk factors, and integrated risk scores (IRS), along with care recommendations developed by clinical experts.

GIRAs are classified as “High Risk” if PRS, monogenic risk, IRS, or family history meet condition-specific criteria, or “Not High Risk” for coronary heart disease, atrial fibrillation, hypercholesterolemia, obesity, type 2 diabetes, chronic kidney disease, breast cancer, and prostate cancer. For pediatric GIRAs, high PRS triggers High Risk GIRAs (asthma, obesity, type 1 and type 2 diabetes). High Risk GIRAs for adults are triggered by PRS, monogenic risk, family history, IRS, or a combination. Disclosure of High Risk PRS results represents a unique area of investigation in eMERGE.

Methods to return High Risk GIRAs (RoR) vary across sites, using different modalities, personnel types disclosing results, and outreach strategies (scheduled appointments, unscheduled phone calls), allowing exploration of feasibility and effectiveness. Participants are offered one-on-one RoR documented in the electronic health record (EHR) for High Risk PRS and monogenic GIRAs. Not High Risk or High Risk due to family history GIRAs are returned via EHR, email, or mail.

As of May 2024, 25,241 participants have enrolled, with 2,017 participants (8%) receiving a High Risk GIRA. Average age of participants receiving a High Risk GIRA was 45 (range: 3-75); 59% reported female sex; and 19% self-reported as Hispanic/Latino(a), 23% Black, and 42% White. Of those requiring RoR, 1,714 (85%) completed the visit, while 303 (15%) were unreachable, declined, or deceased/incapacitated. RoRs included 1,358 High Risk for at least one PRS only, 130 positive monogenic findings only, and 38 with both High Risk PRS and positive monogenic findings. For RoR methods, 1,110 (65%) were disclosed by phone, 561 (33%) by videoconference, and 16 (1%) in person. Total RoRs by personnel included genetic counselors (N = 1,269, 63%), pharmacists (N = 394, 20%), study coordinators (N = 211, 11%), and study physicians (N = 47, 2%). Sites estimated that RoR sessions took an average of 28 minutes (range: 17-45).

The return processes for High Risk GIRAs provide insights into engagement and clinical utility of incorporating PRS in patient care. eMERGE IV will analyze healthcare processes and participant outcomes based on RoR method, personnel, and other variables, informing future implementation of genome-informed risk in clinical care.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7079T: SPOT: Sanger post-orthogonal tracker

Authors:

S. Chavan[1], S. Krishnamoorthy[1], M. Fisher[2], J. Shortt[1], N. Pozdeyev[2], A. Stewart[1], K. Arbogast[1], T. Frye[1], E. Todd[1], K. Crooks[1,3], N. Rafaels[1], The Colorado Center for Personalized Medicine; [1]Colorado Ctr. for Personalized Med., Univ. of Colorado Anschutz Med. campus, Aurora, CO, [2]Dept. of BioMed. Informatics, Univ. of Colorado Anschutz Med. campus, Aurora, CO, [3]Dept. of Pathology, Univ. of Colorado Anschutz Med. campus, Aurora, CO

Abstract:

The biobank at the Colorado Center for Personalized Medicine (CCPM biobank) currently returns High Impact pathogenic variants (HIPV) to the CCPM Biobank participants identified through 94,826 individuals genotyped from 72,428 samples from whole exome sequencing (WES) and 33,864 samples genotyped on the Multi-Ethnic Genotyping Array (MEGA). Filtering of the WES and MEGA data to the American College of Medical Genetics (ACMG) recommendations for reporting of incidental findings revealed 2416 and 738 patients, respectively, identified for confirmatory genotyping. It is time consuming and cumbersome to coordinate tracking these HIPV between the genetic counselors (GC) and the Biobank team and is imperative to use an informatics tracking solution to track these HIPV identified for Sanger confirmatory testing.

Considering HIPAA guidelines for secure tracking of these HIPV, we developed the *Sanger Post Orthogonal Tracker (SPOT),* an academic software solution, at the CCPM biobank. *SPOT* is a meticulously generated Microsoft Excel application hosted on OneDrive, leveraging the power of pivot tables to elegantly track and manage HIPV entries. This sophisticated tool not only streamlines workflow efficiency but also fosters enhanced cross-team collaboration, empowering users to effortlessly navigate and manipulate data for optimal decision-making. *SPOT* comprises of main table with HIPV and associated annotations with ability to dynamically change and track the status of each patient/variant entry. The novelty of SPOT comes from the use of Pivot tables which allow for dynamic analysis of data. Users can easily rearrange and filter HIPV variants to track progress and collaborate changes from multiple users across the GC and the Biobank laboratory teams, facilitating real-time modifications and updates. *SPOT* enables users to search by patient identifiers which renders HIPV associated with that patient and also allows search by variant of interest which renders one or many patients that have the HIPV of interest.

SPOT extends functionality from variant identification to interfacing with patients and facilitates

as a useful resource for biobank laboratory researchers and GC and thereby aiding the patients and contributing to CCPM's personalized medicine initiatives. CCPM has already provided high impact pathogenic variants to over 500 biobank participants. These clinical genetic test results have been used to educate and inform individuals of their hereditary disease risk information and has given hundreds of people in our community the opportunity to pursue health care tailored to their personal and familial genetic risk.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7080T: State of Affairs of Genomic Education in US Allopathic Medical Schools

Authors:

M. Drew[1], C. Kartchner[1], H. Ayoubieh[2]; [1]TTUHSC El Paso, El Paso, TX, [2]TTUHSC, Florida, FL

Abstract:

While genomic medicine is revolutionizing medical care, many physicians feel unprepared to tackle this facet of medicine in their medical practice because genetics is typically taught as a basic science in the pre-clerkship phase of medical school. The undergraduate medical curriculum is dense and there are limited opportunities to include genomic medicine due to time constraints and potentially, lack of expertise. Schools that offer genetic tracts or distinction programs recognize the importance of education as the keystone in a physician's training and will be equipping future physicians with the skills and knowledge necessary to deliver genomic medicine in their future practice. Every Allopathic medical school in the United States will be examined as it is not currently known how and when US allopathic medical schools include genomic medicine in their curricula and if they offer a concentration or distinction pathways for their students. The goal of this project is to perform a descriptive analysis of the curricula of all American Allopathic schools that are accessible through the web to identify where and how genomics is taught.

Two reviewers searched the curricula of 153 US Allopathic medical schools. Only curricula that are posted on the web were included. The curricula were searched for the following: 1- how genomics is taught; 2- if a distinction/ concentration or scholarly pathway is offered, program design, enrollment, and published outcomes; and 3- if further training in genetics is available at the same academic medical institution in the form of residency programs. Preliminary analysis revealed that there is a limited number of medical schools that offer scholarly pathways in genomics, they varied in their design and implementation and few reported on their outcomes.

This analysis will describe the state of undergraduate medical education in genomics and look at the alternate pathways that offer a deeper dive into genomic medicine for medical students.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7081T: The Genome Health Analyst: a new role in genomic medicine

Authors:

A. Cantor[1], C. Davis[2], K. Ruggles[1], A. Chakravarti[1]; [1]NYU Grossman Sch. of Med., New York, NY, [2]Sarah Lawrence Coll., Bronxville, NY

Abstract:

With the rise in DNA sequencing in medicine, the need for an integrated workforce is acute and incompletely served by individuals with either insufficient (e.g., M.S. graduates of genetic counseling programs) or too advanced (e.g., PhD computer scientists) training in the requisite areas. The current division of labor, where data analysis is performed by bioinformaticians and computational genomicists and clinical application is covered by medical geneticists, genetic counselors and other clinicians, is inadequate. To meet this need, we designed the Genome Health Analysis (GHA) training program to train a new class of health professional who will generate medically actionable information from personal genomes to improve an individual's future health outcome. We created the GHA as a novel, cutting-edge 2-year M.S. degree program to meet the growing demands of those working in genomic medicine. This program follows a competency-based education model, and includes didactic coursework, case-base skills lab and an immersive practicum, that trains students in human genetics and genomics, computer science and bioinformatics, and healthcare systems. Students will learn to perform unbiased genome-wide variant extraction and annotation, and assess clinical actionability in order to facilitate genomically-informed healthcare decisions. Graduates of the program will develop versatile expertise and experience in interdisciplinary environments that prepare them to work in many healthcare contexts, such as hospitals and clinics, genetic testing companies, biotech and pharmaceutical companies, organizations, and government agencies.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7082T: Understanding How Perception of Uncertainty Affects Decision Making for or Against Familial Carrier Testing in Families Affected by Disorders of Purine and Pyrimidine Metabolism

Authors:

S. Haines, K. Chaillet, O. Shchelochkov; NIH/NHGRI, Bethesda, MD

Abstract:

Background: Genetic counseling for rare diseases can be challenging, given the limited literature available and the variable presentation of many rare diseases, including disorders of purine and pyrimidine metabolism (DPPM). Current recommendations from the American College of Medical Genetics (ACMG) do not support targeted carrier testing of unaffected minors for recessive conditions. In rare diseases, with limited understanding of natural history and few treatment options, families frequently experience uncertainty and many pursue carrier testing for unaffected individuals even when it is not supported by the current ACMG recommendations. Uptake of carrier testing in DPPM families is unknown and factors influencing this decision are poorly understood. **Methods**: Sample selection will be unaffected family members enrolled in a DPPM natural history protocol (NCT06092346), including parents and siblings. Interested individuals will sign consents for recordings of the semi-structured interviews. Transcription software may be utilized, and we will analyze the transcripts to determine the main themes. A codebook will be created for qualitative data analysis. **Preliminary Results:** There are currently eight individuals with DPPM enrolled in the protocol, for a total of 7 families affected by Lesch-Nyhan disease (n=4), adenylosuccinate lyase deficiency (n=3, 2 are siblings), and phosphoribosylpyrophosphate synthetase deficiency (n=1). Eighty-seven percent of the families self-identified as white and non-Hispanic. The age range of the affected individuals is 2 years to 17 years of age. There are 6 unaffected siblings in the seven families. Three siblings are over 13 years of age. One unaffected sibling is a twin to the affected sibling. Thus far, of the 6 unaffected siblings, 3 have completed carrier testing before age 18. Two of the three requested and completed bioethics review Board approval before carrier testing. **Conclusions:** Preliminary results suggest that some DPPM families may cope with uncertainty through information-seeking by pursuing genetic testing for unaffected family members. Additional collected data will provide the opportunity to assess the role of carrier testing in families of rare diseases and reveal its effect on interactions with healthcare and other social systems, intrafamilial communication regarding the diagnosis, and coping strategies adopted by family members and caregivers. Through in-depth interviews, we will gain a better understanding of the decision-making processes and

downstream effects of carrier testing in families affected by DPPM, and rare diseases more broadly.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7083T: Utilizing *ABCC11* genotyping in high school genetics education: A hands-on approach

Authors:

T. Tokutomi[1], R. Takai[2], A. Yoshida[3], D. Paudel[4], S. Giri[4], T. Kitagawa[4], T. Arakawa[2], Y. Kuramitsu[4], T. Ohta[4]; [1]Dept. of Pediatrics, Kawasaki Med. Sch., Kurashiki, Japan, [2]Div. of Biochemistry Dept. of Oral Biology, Hlth.Sci. Univ. of Hokkaido, Tobetsu, Japan, [3]Dept. of Clinical Genetics, Sch. of Med., Iwate Med. Univ., Morioka, Japan, [4]Advanced Res. Promotion Ctr., Hlth.Sci. Univ. of Hokkaido, Tobetsu, Japan

Abstract:

Introduction Molecular genetics technology is advancing rapidly, due in part to university-level engagement. Hands-on exposure in high school through experimentation is beneficial but costly. The ABCC11 gene, which determines earwax type via Mendelian inheritance, offers an accessible model for teaching genetics. A 2-day course guided high school students through cloning their earwax genes and identifying DNA variations, enhancing their understanding of Mendelian traits via hands-on and computer-based learning. **Materials and Methods** Participants were selected based on strong performance in biology, and were given a preparatory session. Parental consent was obtained and the university ethics committee approved this study. The course included lectures and a conversation with a guest researcher. Approximately 25 students conducted the experiments simultaneously, working in groups as needed. Students extracted genomic DNA from oral epithelial cells, conducted PCR, and created pedigree charts using provided software. A satisfaction questionnaire was completed by 126 students who participated between August 2019 and January 2020. **Results** Isolation of DNA for PCR required approximately 3 h. The students amplified DNA, obtaining substantial amounts of PCR products for gel electrophoresis. The PCR products were digested with DdeI to allow visual genotyping via gel electrophoresis. Students discussed their results, deepening their understanding of molecular techniques and Mendelian genetics. In one session with 47 students, 38 had genotype A/A, four had G/A, and one had G/G, while four failed to detect their genotype. The G allele was dominant, and the students analyzed the data using Hardy-Weinberg equilibrium. Students also created pedigree charts with family phenotype information using the

f-tree application. **Discussion** Among the participants, 90 were very satisfied, 12 were satisfied, 16 were average, four were dissatisfied, and four did not respond. Over 80% of participants were very satisfied or satisfied. This course significantly enhanced genetic literacy and practical understanding, demonstrating its real-world relevance. Despite limitations including a small sample size and a short duration, the course effectively enhanced students' understanding of genetics.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7084T: Utilizing web analytics to assess patient engagement and prioritize educational targets for population-scale biobanks

Authors:

E. Kudron[1], E. Shalowitz[1], J. L. Martin[2], The Colorado Center for Personalized Medicine, D. P. Kao[1]; [1]Univ. of Colorado Sch. of Med., Aurora, CO, [2]Univ. of Colorado Skaggs Sch. of Pharmacy & Pharmaceutical Sci., Aurora, CO

Abstract:

Introduction: Some population-scale research biobanks preemptively return clinical genomic results to participants through the electronic health record (EHR) months after enrollment without the involvement of the patients' healthcare providers. This method of returning genomic results poses unique challenges to sustained participant engagement and education. Web content can help mitigate these issues by making educational resources widely accessible. Herein, we describe the frequency of participant visits to a large research biobank's webpages. Methods: Our biobank has developed online content to educate participants, which includes 10 pages about pharmacogenomic (PGx) results and 4 about secondary findings. Using website analytics, we assessed which webpages were most frequently viewed by participants between 6/1/23 and 5/31/24. Participants were inferred using tagged referrals from welcome emails, newsletters, and the online patient portal connected to the EHR, where results are returned and visible to participants. All other webpage visits were excluded from analysis as we could not accurately infer if these visitors were participants. Results: Of the 26,178 unique webpage visits, 5,511 (21%) were identified as participants, with 4,379 (80%) being new visitors and 796 (14%) returning users. The most frequently visited pages were: PGx results overview (n=1151, 21%); PGx gene pages (n=1126, 20%); secondary findings summary (n=418, 8%); return of all results (n=361; 7%); frequently asked questions (FAQs) (n=319, 6%); how the biobank works (n=116, 2%); and PGx-specific drug pages (n=30, 1%). Compared to returning users, new visitors were

significantly more likely to visit PGx results information, including gene pages (n=940 new vs. 139 returning; p=0.01) and results overview (n=971 new vs. 126 returning; p=0.00006). Conversely, returning users were significantly more likely than new users to visit pages about secondary findings (n=165 new vs. 165 returning; p<0.00001), return of results overall (n=249 new vs. 70 returning, p=0.0008), and FAQs (n=201 new vs. 101 returning; p<0.00001). Conclusion: We found that new biobank webpage users were more likely to view education about PGx results while returning users were more likely to seek information about secondary findings. When initiating the development of participant education, biobank programs should consider starting with the creation of robust PGx content to maximize new user engagement. By catering to participants' interests and needs, we can foster a more informed population, ready to leverage genomic information for better health outcomes.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Thursday Session

Board 7085T: What should you do when molecular lab results do not acknowledge likely allele status or current literature regarding resulting disease?

Authors:

Y. Furuta, R. J. Tinker, R. M. Dahlhauser, J. A. Phillips, III; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Introduction: EMILIN1, an elastin microfibril interfacer, is linked to connective tissue integrity. While the *EMILIN1*gene is listed in OMIM as causing an autosomal dominant (AD) disorder (OMIM # 620080), there are no current OMIM entries of clinically heterogeneous disorders with an autosomal recessive (AR) mode of inheritance variants. In 2022, biallelic loss-of-function variants in the *EMILIN1* gene were shown to cause an AR disorder characterized by cutis laxa, arterial tortuosity, aneurysm formation, and bone fragility in 5 individuals (PMID: 36351433). **Materials:** A 9-month-old male was referred with short stature, tortuosity in multiple arteries, pulmonary stenosis, and multiple fractures. **Molecular results:** DNA sequencing panels for osteogenesis imperfecta and bone fragility as well as connective tissue disorders revealed a heterozygous *COL9A2* variant of uncertain significance (VUS). Trio exome sequencing (ES) revealed a c.275_277delinsTT, p.(C92Ffs*22) maternal and a c.1759 G>T, p.(G587*) paternal *EMILIN1* variants of unknown significance (VUSs). However, the lab report did not acknowledge that these biallelic variants that encoded premature terminations relatively early in *EMILIN1* were likely pathogenic, and stated that there were no known

diseases, either AD or AR, that were associated with *EMILIN1* variants. The lab classified his biallelic *EMILIN1* terminator variants as VUSs and did not postulate a possible relationship with an AR disorder. Also, the close overlap of his pleiotropic features with those of multiple individuals who had the recently identified biallelic loss-of-function variants in the *EMILIN1* was not acknowledged. Thus, neither concern over the possible pathogenicity of his variants, nor the possibility that his very unusual phenotypic features meshed well with those seen in others with biallelic loss-of-function *EMILIN1* variants were reported. **Conclusions:** The very unusual pleiotropic features of this case are almost certainly caused by the detected biallelic terminator *EMILIN1* variants. However, his trio ES report did not: 1) acknowledge the likely pathogenicity of these variants, 2) acknowledge the current OMIM # 620080 of a known AD disorder, or 3) report current literature (PMID: 36351433) regarding his likely resulting disease. To be truthful and prevent harm from these lab oversites, we informed his parents that he very likely has this known condition, and gave them details about the condition, despite the s benign lab interpretation. This scenario illustrates the potential need for clinicians and genetic counselors to carefully review, and if clinically indicated, correct misinformed lab reports.

Session Title: Genetic Therapies Poster Thursday Session

Board 8081T: A drug repurposing screen identifies N-acetyl-L-leucine as a candidate therapeutic molecule for SYNGAP1-related developmental disorder

Authors:

K. Beebe, E. Coelho, C. Massey, H. Evans, C. Chow; Univ. of Utah, Salt Lake City, UT

Abstract:

Autosomal dominant mutations in SYNGAP1 cause a rare developmental disorder with a spectrum of symptoms across patients including intellectual disability, epilepsy, and hypotonia. There are currently no curative treatments for SYNGAP1 disorder. We recently completed a drug repurposing screen in a *Drosophila* model of SYNGAP1. Using an eye-based phenotype, we screened 1520 molecules from the Prestwick Chemical Library and identified a number of promising hits, including molecules in steroid and anti-inflammatory classes. Select hits were advanced for validation and N-acetyl-L-leucine (NAL) had the strongest validation data. N-acetyl-L-leucine is a modified form of the essential amino acid leucine that is safe and has been approved since 1957 in France under the name Tanganil for use as an antivertigo therapeutic. Further testing indicates that other forms of leucine or the enantiomer of NAL are unable to rescue the model, suggesting a very specific role for NAL. We will present transcriptomic data in

treated and untreated SYNGAP1 model *Drosophila* that will shed light on the mechanism of action of this rescue. We will also present tissue-specific analyses that will reveal in which tissues NAL might be most effective, with special focus on neurons and glia. This work suggests that NAL is a potential new therapeutic molecule for SYNGAP1 and reveals a previously unknown role for leucine in SYNGAP1 dysfunction.

Session Title: Genetic Therapies Poster Thursday Session

Board 8082T: A knockdown-replacement system to target pathogenic DEAF1 mutations

Authors:

O. Tewogbola (Falayi)[1], P. Jensik[2]; [1]Southern Illinois Univ., Carbondale, Carbondale, IL, [2]Southern Illinois Univ., Carbondale, IL

Abstract:

DEAF1-associated neurodevelopmental disorder (DAND) is characterized by a spectrum of neurological disorders, and caused by heterozygous or homozygous mutations in the deformed epidermal autoregulatory factor 1 (*DEAF1*) gene. Genetic-based therapies involving the reduction of a mutant transcript, or introducing a functional copy of the gene have been explored in single-gene disorders like Rett and Angelman syndrome. This study aims to develop an in vitro methodology to knockdown the majority of mutant *DEAF1* transcripts as well as replace DEAF1 with a shRNA-resistant *DEAF1* transgene (a knockdown-replacement system). Several DEAF1 targeting shRNAs were designed and shRNA-mediated percent knockdown of DEAF1 expression was determined using western blot and RT-qPCR. The effectiveness of selected shRNAs in reducing protein expression of 2 different patient-identified DEAF1 pathogenic variants, p.Q264P and p.R226W, was determined using western blot analysis. An shRNA-resistant DEAF1 transgene was developed by altering the *DEAF1* native shRNA target coding sequence with silent mutations (that preserve the amino acid sequence). shRNA-mediated knockdown of the shRNA-resistant *DEAF1* transgene was assessed by western blot. An AAV construct was generated that expresses both the DEAF1 targeting shRNA and shRNA-resistant *DEAF1* transgene. Stable cells over-expressing p.Q264P were transduced with AAV DEAF1 knockdown-replacement expressing particles to assess the rescue of *BRSK2* (a DEAF1-regulated gene).
A DEAF1-targeting shRNA (sh619) was selected based on knockdown percentage and sequence location relative to identified pathogenic DEAF1 mutations. Expression of sh619 in HEK293t cells stably expressing 2 different DEAF1 variants, p.Q264P or p.R226W, resulted in over 90%

reduction in mutant protein levels. As expected, the shRNA-resistant DEAF1 transgene showed no differences in cellular localization and transcriptional activity compared to WT DEAF1, and more importantly sh619 was unable to reduce shRNA-resistant DEAF1 protein levels. The mRNA and protein levels of BRSK2 (DEAF1-regulated gene) are significantly reduced in cells overexpressing p.Q264P. An 8-fold ($p<0.01$) protein and a 3.5-fold mRNA ($p<0.05$) increase in BRSK2 levels, respectively, was observed in p.Q264P-expressing cells after AAV DEAF1 knockdown-replacement transduction.
Our reports indicate that a knockdown-replacement combination construct can potentially rescue the expression of a DEAF1-regulated gene *BRSK2* in a DAND cell model. These results indicate a potential approach to restore the expression of dysregulated genes caused by pathogenic DEAF1 mutations.

Session Title: Genetic Therapies Poster Thursday Session

Board 8083T: An antisense oligonucleotide (ASO) strategy for restoring IP_3R1 expression in spinocerebellar ataxia type 15/16

Authors:

B. Masser, J. P. Stein, H. S. McLoughlin; Univ. of Michigan, Ann Arbor, MI

Abstract:

Currently, no disease-modifying therapeutics are available for any of the ~50 identified subtypes of spinocerebellar ataxia (SCA), a heterogeneous group of dominantly inherited neurodegenerative ataxias. SCA15/16 is caused by haploinsufficiency of the inositol 1,4,5-trisphosphate receptor, type 1 (*ITPR1*/IP_3R1), and is associated with slowly progressive cerebellar atrophy, particularly of Purkinje neurons, leading to late-onset gait disturbances, eye movement abnormalities, and difficulties with speaking and/or swallowing. Although haploinsufficiency is a readily treatable disease mechanism through adeno-associated viral (AAV) gene replacement strategies, IP_3R1 supplementation is challenged by the large calcium (Ca^{2+}) channel's size, with an 8.2 kb coding sequence, nearly double the packaging capacity of vectors. IP_3R1, like many ion channels, is also a "Goldilocks" protein, sensitive to both down- and upregulation, which is causative of SCA29 and is observed in mouse models of SCAs caused by toxic repeat expansions of a disease gene's polyglutamine tract. To overcome these obstacles, we propose the use of antisense oligonucleotides (ASOs), which allow for restoration rather than overexpression of the endogenous locus. We hypothesize that ASO binding to microRNA (miRNA) recognition elements (MREs) within the *ITPR1* 3' untranslated region (UTR) will inhibit endogenous miRNA-mediated translational silencing from targeting the remaining

allele, restoring IP_3R1 expression and function. Using an *ITPR1* 3' UTR dual luciferase reporter, we confirmed 10 putative *ITPR1*-miRNA interactions. A miR-masking ASO (MMO) was designed against the most active MRE and validated using our reporter system, demonstrating a restoration of reporter expression to control levels in MMO-treated cells. Current studies focus on assessing MMO protection of endogenous *ITPR1* transcripts. While the generation of *ITPR1*$^{+/-}$ and isogenic control human induced pluripotent stem cell (hiPSC)-derived cerebellar Purkinje neurons for patient-relevant insights is underway, we are utilizing the same hiPSC lines to generate *ITPR1*-expressing forebrain neurons, gaining a preliminary understanding of MMO performance. Ongoing histological and immunoblotting analyses are comparing IP_3R1 expression in MMO-treated *ITPR1*$^{+/-}$ neurons, relative to untreated cells. Ca^{2+} imaging experiments of MMO-treated neurons will further assess IP_3R1 functional restoration. This research not only provides preclinical support for the MMO strategy as an SCA15/16 therapeutic but is also broadly applicable to diseases caused by haploinsufficiency of a large protein.

Session Title: Genetic Therapies Poster Thursday Session

Board 8084T: Antisense oligonucleotides targeting linked-SNPs provide allele-specific knockdown to a dominant-negative *SPTAN1* pathogenic variant in a complex genetic region

Authors:

C. Wang[1], S. Li[1], S. Zhao[1], J. Sinson[1], D. Lanza[1], W. Wang[1], G. Zapata[1], K. Macakova[1], N. Ho[1], B. Vuocolo[1], G. Clark[1], S. Nagamani[1], L. Burrage[1], J. Heaney[1], H. Zoghbi[2], P. Liu[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. of Med./HHMI/NRI, Houston, TX

Abstract:

Developmental and epileptic encephalopathy 5 is a rare neurodevelopmental disorder associated with monoallelic variants in *SPTAN1*, from which the non-erythrocytic αII-spectrin subunit is translated. Individuals with the c.6908_6916dup (p.D2303_L2305dup) variant experience infantile epilepsy with refractory seizures, profound developmental delay, hypotonia, and microcephaly. Multiple patients harboring the duplication variant succumb to the disorder before 6 years of age. This pathogenic variant is located within the heterodimerization domain of αII-spectrin. Previous studies demonstrated a dominant negative consequence of the variant, causing aggregation of αII with β spectrins in patient-derived fibroblasts. We sought to explore the potential of antisense oligonucleotide (ASO)-mediated RNaseH-dependent allele-specific degradation as a therapeutic option. However, targeting the

variant is made challenging by the presence of in-tandem repeats within the duplicated region. Here, we report an alternative strategy to effectively knockdown the variant allele in complex genetic regions.
We enrolled 5 affected children with the duplication variant from 4 different families. Patient-derived neural stem cells (NSCs) and neurons (Ns) were generated from peripheral blood mononuclear cells or fibroblasts. We designed ASOs directly targeting the variant but failed to effectively knockdown the variant expression and discriminate between the normal and duplication alleles. This is likely due to the extended length of the duplicated sequence preventing the ASOs from being centered on the variant-allele-specific nucleotides and causing self-dimerization.
As an alternative approach, we performed genome sequencing and identified candidate heterozygous exonic and intronic linked single nucleotide polymorphisms (SNP) in each individual and designed targetable gapmer ASOs. We performed qPCRs to phase SNP candidates to the duplication allele. Priorities were given to SNPs that exhibited promising sequencing features supporting ASO potential or higher population frequencies, allowing a broader application for future translation. Our screen using ddPCR and RNA-seq generated several ASOs targeting the linked-SNPs with promising levels of allele-specific mRNA knockdown. Immunofluorescent imaging in NSCs and Ns treated with linked-SNP ASOs further demonstrated changes in αII protein levels.
Ongoing work includes functional electrophysiology studies in Ns, assessing transcriptional consequences of the variant, evaluating potential off-target effects of the ASO, and developing a mouse model for *in vivo* testing.

Session Title: Genetic Therapies Poster Thursday Session

Board 8085T: Bone-targeted anti-TGFβ antibody approach to improve efficacy and safety for the treatment of osteogenesis imperfecta ⭐

Authors:

I-W. Song[1], D. Nguyen[1], M. Adeyeye[1], Y. Chen[1], A. Tran[1], S. Musaad[2], Y. Mirabile[2], B. Greene[3], J. Cao[3], X. Ying[3], A. Park[3], E. Masterjohn[3], D. Honey[3], S. Liu[3], S. Nagamani[1], B. Lee[1]; [1]Baylor Coll. of Med., Houston, TX, [2]US Dept. of Agriculture/Agricultural Res. Service (USDA/ARS) Children's Nutrition Res. Ctr., Houston, TX, [3]Sanofi, Waltham, MA

Abstract:

Osteogenesis imperfecta (OI) is a genetic skeletal disorder for which, currently, there are no FDA-approved therapies. We, and others, have previously shown that excessive TGFβ signaling is

a key pathogenic driver in OI and that inhibition of TGFβ could be a potential mechanism-specific approach for treatment. In a phase 1 study with fresolimumab, an anti-TGFβ antibody, we reported single-dose was well-tolerated and found an interesting differential response between moderate OI type IV and severe OI type III, which suggested a dose-specific effect based on disease severity. Here, we report on the tolerability and pharmacodynamics of multi-dose fresolimumab cohort in the phase 1 study and the preliminary data of a bone-targeted anti-TGFβ antibody to improve efficacy and safety in OI mice. There are two dosing groups in the multi-dose fresolimumab cohort: 4mg/kg Q3M (Group 1) and 4mg/kg Q6M (Group 2). We enrolled total 3 individuals with moderate-to-severe collagen I-related OI. One participant completed 3 doses in group 1 and 2 participants completed 2 doses in group 2. We report here the data through 180-day trial period. Overall, the multi-dose was well-tolerated. Total 41 adverse events (AEs) reported with 11 graded study medication related. The most common study medication related AEs were mild gum bleeding (3) and mild epistaxis (2). The hemoglobin, white blood cell and platelet count were comparable to baseline at day 180. While the serum CTX level fluctuated throughout the reporting period, the P1N1 showed a 30-130% increase across 3 participants at day 180. Although this suggested a potential anabolic effect after multi-dose exposure, a mixed response was observed in lumbar spine aBMD and trabecular measurements. This alluded to the possible need targeting anti-TGFβ to bone to improve safety and efficacy. Therefore, we generated a bone-targeted anti-TGFβ antibody, 1D11-D10. After a single dose administration in WT mice, 1D11-D10 showed rapid distribution to bone and has a longer half-life of 76 days in bone compared to 16.3 days for untargeted 1D11. Importantly, 1D11-D10 showed higher increase in bone mass compared to untargeted 1D11 at 5mg/kg 3x/week in *Col1a2*$^{G610C/+}$ OI mice. These data supported the effective bone targeting and a potential increased efficacy and safety profile for 1D11-D10. The bone targeted anti-TGFβ therapy may offer wider therapeutic index for clinical translation in OI and other skeletal diseases with TGFβ dysregulation.

Session Title: Genetic Therapies Poster Thursday Session

Board 8086T: Development and application of antisense oligonucleotide (ASO) assessment guidelines for loss-of-function variants causing autosomal recessive and X-linked monogenic disorders

Authors:

L. Newton[1,2], D. Cheerie[1,2], M. C. Lauffer[3], D. Beijer[4,5], M. Meserve[6], M. Synofzik[4,5], A. Aartsma-Rus[3], M. Szego[2,7], K. Amburgey[2,8], B. Kalish[1], J. J. Dowling[2,1], G. Costain[2,1,8]; [1]Genetics and Genome Biology, SickKids Res. Inst., Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of

Toronto, Toronto, ON, Canada, [3]Dutch Ctr. for RNA Therapeutics, Leiden Univ. Med. Ctr., Leiden, Netherlands, [4]Div. of Translational Genomics of Neurodegenerative Diseases, Hertie-Inst. for Clinical Brain Res., Univ. of Tübingen, Tübingen, Germany, [5]German Ctr. of Neurodegenerative Diseases (DZNE), Tübingen, Germany, [6]Boston Children's Hosp., Boston, MA, [7]Joint Ctr. for Bioethics, Univ. of Toronto, Toronto, ON, Canada, [8]Div. of Clinical & Metabolic Genetics, Hosp. for Sick Children, Toronto, ON, Canada

Abstract:

Background: Of the more than 7000 rare diseases—conditions affecting fewer than 200,000 individuals—80% are genetic in etiology and 70% pediatric in onset. Despite our understanding of the underlying genetic causes of rare diseases, few therapeutic options exist, with less than 5% having an FDA-approved treatment. Their RNA altering mechanisms and proven effectiveness underscore the potential of antisense oligonucleotides (ASOs) as sources of treatment for rare genetic diseases. ASOs are short, synthetic, single-stranded oligonucleotides that can bind RNA via Watson-Crick base pairing. ASOs can modify protein expression through canonical exon-skipping, splice rescue, RNA knockdown, or the upregulation of wildtype gene product, thereby offering the opportunity to customize treatments for specific DNA variants. To accelerate ASO development, efficiency, and applicability, as well as to advance precision medicine approaches and the treatment of rare genetic diseases, a standardized approach for evaluating the eligibility of genetic variants for ASO treatment is necessary.
Methods: The N=1 Collaborative (n1collaborative.org/) has developed a consensus guideline for the assessment of diagnostic variants in monogenic disorders for eligibility towards ASO development. Guideline development was a multi-site effort, developed through iterative rounds of writing, testing, and revisions, incorporating the expert opinions of clinicians, geneticists, and researchers. These guidelines provide readers the ability to analyze and classify variants with regard to their amenability to ASO therapies using the best available knowledge and practice.
Results: Our guidelines provide step-by-step instructions on the assessment of variants, along with recommended bioinformatic tools, publicly available databases, resources, and example variants. Using our guidelines, variants are classified in a hierarchal manner, enabling us to identify the most eligible variants and optimize patient selection and resource allocation. The guidelines are now being used to evaluate all diagnostic results from genome-wide sequencing in a cohort of children with rare genetic conditions at The Hospital for Sick Children over a 5-year period.
Conclusion: Our initial findings underscore the importance of considering molecular, clinical, and bioethical factors to better reflect the number of genetic diagnoses amenable to genetic therapy. Results promise to deliver a precise count of patients whose variants are credibly amenable to genetic therapies.

Session Title: Genetic Therapies Poster Thursday Session

Board 8087T: Exon Specific U1 snRNAs as a therapeutic approach to rescue optic neuropathy in Familial Dysautonomia ⭐

Authors:

K. Kondabolu[1], M. Stern[1], E. Kirchner[2], J. Bolduc[1], F. Pagani[3], L. Vandenberghe[4], S. Slaugenhaupt[2], E. Morini[2], A. Chekuri[2]; [1]Massachusetts Gen. Hosp. Res. Inst., Boston, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]ICGEB, Trieste, Italy, [4]Harvard Med Sch, Boston, MA

Abstract:

Familial Dysautonomia (FD) is an autosomal recessive disorder caused by a T-to-C base transition in 5' splice site of intron 20 of Elongator acetyltransferase complex subunit 1 gene (ELP1). This mutation causes a tissue-specific skipping of ELP1 exon 20, leading to a premature termination codon and reduction of ELP1 levels predominantly in the central and peripheral nervous system. FD patients experience complex neurological symptoms along with progressive optic neuropathy due to selective loss of retinal ganglion cells (RGCs). Therapeutic strategies for FD include small molecule splicing modulators, antisense oligonucleotides, and exon-specific U1 small nuclear RNAs (ExspeU1). ExspeU1s were designed to interact with the 5' splice site (5'ss) of ELP1 exon 20 through specific base-pairing and restore correct exon inclusion by promoting recruitment of spliceosome machinery at mutated 5'ss. We investigated the therapeutic potential of ExspeU1 in rescuing visual function in FD. To initially evaluate the efficacy of ExspeU1 *in vitro*, HEK cells were transiently transfected with a minigene carrying human ELP1 genomic sequence spanning from exon 19-21 containing FD mutation and treated with increasing concentrations of plasmids construct expressing ExspeU1. We observed a dose-dependent improvement of Exon 20 inclusion with ExspeU1 treatment. Similarly, transduction of FD patient-derived fibroblasts with self- complementary adeno-associated virus 2 scAAV2-ExspeU1 also demonstrated improved exon 20 inclusion. To demonstrate the *in vivo* efficacy of ExspeU1, intravitreal administration of scAAV2-ExspeU1 was performed in humanized mouse model of FD. One eye of the FD mice was injected with 1µl of scAAV2-ExspeU1 at a titer of 1e11 vector genomes (VG) and the contralateral eye was injected with sterile phosphate-buffered saline (PBS) at age postnatal day 30. Our results indicate significant rescue of retinal nerve fiber layer (RNFL) thickness and retinal ganglion cell (RGC) count in the scAAV2-ExspeU1injected eyes compared to PBS injected eyes. Finally, we evaluated the rescue of visual function using pattern electroretinography (pERG) and optomotor response assay (OMR) in FD mice injected with the

scAAV2-ExspeU1. Our results indicate significant improvement in pERG amplitude and OMR responses in FD mice in scAAV2-ExspeU1eyes compared to PBS injected eyes. In summary, our study demonstrates the therapeutic efficacy of ExspeU1 in the correction of ELP1 splicing defect and improvement of visual function in a humanized mouse model of FD. This study highlights the potential of ExspeU1 as a promising therapeutic approach to rescue vision loss in FD patients.

Session Title: Genetic Therapies Poster Thursday Session

Board 8088T: Expanding the Landscape of Drug Targets for 112 Chronic Diseases Using a Machine Learning-Assisted Genetic Priority Score

Authors:

R. Chen[1], Á. Duffy[1], B. Petrazzini[1], H. Vy[1], D. Stein[1], M. Mort[2], J. K. Park[1], A. Schlessinger[1], Y. Itan[1], D. N. Cooper[2], D. M. Jordan[3], G. Rocheleau[1], R. Do[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Cardiff Univ., Cardiff, United Kingdom, [3]Icahn Sch. of Med. at Mt Sinai, New York, NY

Abstract:

Background/Aims: Identifying genetic drivers of chronic diseases is crucial for drug discovery. We developed a Machine Learning-assisted Genetic Priority Score (ML-GPS) that incorporates genetic associations with predicted disease phenotypes to enhance target discovery.
Methods: We constructed gradient boosting models to predict 112 chronic disease phecodes in the UK Biobank and analyzed associations of predicted and observed phenotypes with common, rare, and ultra-rare variants to model the allelic series. We integrated these associations with existing evidence using gradient boosting with continuous feature encoding to construct ML-GPS, training it to predict drug indications in Open Targets and externally testing it in SIDER. We then generated ML-GPS predictions for 2,362,636 gene-phecode pairs, which are interactively viewable online.
Results: Machine learning models accurately predicted phecode diagnoses, with a median area under the receiver operating characteristic curve of 0.85 [interquartile range 0.08]. Predicted phenotypes identified substantially more common, rare, and ultra-rare variant associations than observed phenotypes. Compared to Locus-to-gene and an approximated version of the GPS, ML-GPS increased the area under the precision-recall curve for drug target identification by 47.5% and 28.6% in Open Targets and 70.7% and 32.0% in SIDER, respectively. Gene-phecode pairs in the top one percent of ML-GPS in Open Targets and SIDER had odds ratios of 6.49 (95% CI 5.60-7.53) and 7.38 (95% CI 6.02-9.03) for drug indication, respectively. ML-GPS also

increased coverage of drug targets, with the top one percent of all scores providing support for 15,077 gene-phecode pairs (including 409 with drug indications) that had no support from an existing approach. ML-GPS also identified well-known target-disease relationships, promising targets without indicated drugs, and targets for several drugs currently in clinical trials, including *LRRK2* inhibitors for Parkinson's disease and olpasiran for cardiovascular disease. **Conclusion:** ML-GPS leverages machine learning to expand the landscape of potential drug targets for chronic diseases, providing supporting evidence for many targets already under investigation as well as new drug targets. ML-GPS offers substantially better prediction performance and target coverage compared to existing approaches.

Session Title: Genetic Therapies Poster Thursday Session

Board 8089T: G-quadruplex-based splice switching as a therapeutic approach in Duchenne muscular dystrophy

Authors:

T. Ishiguro[1,2], R. Iwase[1,2], R. Hara[1,2], T. Nagata[1,2], T. Yokota[1,2]; [1]Tokyo Med. and Dental Univ., Tokyo, Japan, [2]Insititute of Sci. Tokyo, Tokyo, Japan

Abstract:

RNA G-quadruplexes, which are unique noncanonical structures composed of stacked guanine quadruplexes, play diverse roles in the regulation of gene expression, from transcription to protein synthesis. Herein, we introduce a novel approach to splice-switching therapy that employs G-quadruplex-inducing antisense oligonucleotides (G-ASOs) to restore dystrophin expression in Duchenne muscular dystrophy (DMD) models. G-ASO consists of two functionally independent domains that facilitate formation of the DNA/RNA hetero-G-quadruplex (hGq) structure. The antisense domain can bind to complementary target RNA sequences, whereas the G-rich domain has a continuous G-tract that forms an hGq structure with the guanine-rich region of the target RNA. This precise binding results in the formation of an hGq structure that effectively interrupts the alternative splicing process. Unlike conventional steric-blocking inhibition of alternative splicing, this technique relies on steric hindrance through hGq formation. Notably, hGq structures can even form when the distance between the G-rich region and antisense sequence on the target RNA is approximately 70 nucleotides apart. In addition, we explored the development of bulge-inserted G-ASOs to overcome the issues associated with G-quadruplex formation through G-ASO self-assembly, enhance the efficiency of G-quadruplex formation, and improve the induction and efficiency of exon skipping therapy. In summary, this study underscores the promising role of G-ASOs in gene therapy, particularly DMD, and

contributes to the advancement of novel therapeutic strategies. These advances highlight the effectiveness of G-ASO in exon-skipping therapy, thereby marking a significant progress in RNA structural manipulation.

Session Title: Genetic Therapies Poster Thursday Session

Board 8090T: High-dose pyridoxine prevents glomerulosclerosis and other disease phenotypes in a patient and knock-in mouse model of SGPL1^{R222Q} variant sphingosine phosphate lyase insufficiency syndrome

Authors:

R. Khan[1], M. L. Allende[2], R. L. Proia[3], J. D. Saba[1]; [1]Univ. of California, San Francisco, Oakland, CA, [2]NIH, Bethesda, MD, [3]NIH, Bethesda, MD 20892, MD

Abstract:

Purpose: Sphingosine phosphate lyase insufficiency syndrome (SPLIS) is a monogenic cause of steroid-resistant nephrotic syndrome caused by loss-of-function of sphingosine-1-phosphate lyase (SPL), a pyridoxal-5'phosphate (PLP)-dependent enzyme that degrades the bioactive lipid sphingosine-1-phosphate (S1P) in the last step of sphingolipid metabolism. Supplementation with pyridoxine, a metabolic precursor to PLP, has therapeutic activity in some inborn errors of metabolism involving PLP-dependent enzymes. Observing clinical responses to pyridoxine in several SPLIS patients, we investigated the potential therapeutic effects of pyridoxine in fibroblasts from a patient harboring the most common SPLIS-associated SPL variant, R222Q, and a novel mouse model of R222Q SPLIS. **Methods:** SPL activity was analyzed by mass spectrometry assay. SPL^{R222Q} mice were generated by gene editing. SPL$^{wild\ type}$ and SPL^{R222Q} mice were fed pyridoxine-low or pyridoxine-high chow. Tissue and plasma S1P was quantified by mass spectrometry. Kidney function was measured by serum albumin and creatinine, urine albumin/creatinine ratio (ACR), histology, immunohistochemistry, and RT-qPCR of kidney injury markers. **Results:** Clinical responsiveness to pyridoxine supplementation and reduction of plasma S1P levels were observed in a SPLIS patient homozygous for the R222Q SPL variant (SPL^{R222Q}). SPL activity increased in SPL^{R222Q} fibroblasts after pyridoxine treatment. SPL insufficiency, S1P accumulation, proteinuria, glomerulosclerosis, molecular indicators of podocyte injury, and other SPLIS phenotypes developed in SPL^{R222Q} but not SPL^{WT} mice maintained on pyridoxine-deficient chow, whereas excess pyridoxine in the chow prevented these phenotypes in SPL^{R222Q} mice.

Session Title: Genetic Therapies Poster Thursday Session

Board 8091T: Mitochondrial dysfunction heightens the integrated stress response to drive ALS pathogenesis

Authors:

H. Miranda[1], C. Landry[1], J. Costanzo[1], M. Mitne-Neto[2], M. Zatz[3], A. Schaffer[4], M. Hatzoglou[1], A. Muotri[5]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Paulo Gontijo Inst. (IPG), Sao Paulo, Brazil, [3]Univ.e de Sao Paulo, Sao Paulo, Brazil, [4]Case Western Reserve Univ., Moreland Hills, OH, [5]Univ. of California San Diego, San Diego, CA

Abstract:

Vesicle-associated membrane protein-associated protein-B (VAPB) is an ER membrane bound protein. VAPB P56S causes a dominant, familial form of amyotrophic lateral sclerosis (ALS), however, the mechanism through which this mutation causes motor neuron (MN) disease remains unknown. Using inducible wild type (WT) and VAPB P56S expressing iPSC-derived MNs we show that VAPB P56S, but not WT, protein decreased neuronal firing and mitochondrial-ER contact (MERC) with an associated age-dependent decrease in mitochondrial membrane potential (MMP); all typical characteristics of MN-disease. We further show that VAPB P56S expressing iPSC-derived MNs have enhanced age-dependent sensitivity to ER stress. We identified elevated expression of the master regulator of the Integrated Stress Response (ISR) marker ATF4 and decreased protein synthesis in the VAPB P56S iPSC-derived MNs. Chemical inhibition of ISR with the compound, ISRIB, rescued all MN disease phenotype in VAPB P56S MNs. Thus, our results not only support ISR inhibition as a potential therapeutic target for ALS patients, but also provides evidence to pathogenesis.

Session Title: Genetic Therapies Poster Thursday Session

Board 8092T: Molecular mechanism and gene augmentation therapy for inherited retinal dystrophy caused by a novel mutation *PROM1* c.619G>T ⭐

Authors:

X. Zhong, P. Xu, F. Guo, Y. Wang, D. Zheng, X. Song; Zhongshan Ophthalmic Ctr., Sun Yat-sen Univ., GuangZhou, China

Abstract:

Purpose: PROM1-related inherited retinal dystrophy (PROM1-IRD) is characterized by the degeneration of photoreceptors, leading to blindness without treatment. Especially, the correlation between phenotype and genotype of hundreds of *PROM1* mutations remains largely unknown. This study aims to evaluate the pathogenicity of a novel mutation *PROM1* c.619G>T we identified in an IRD patient. Gene therapy was also tested.
Methods: All studies were approved by ethics committees. An IRD patient with a novel homozygous nonsense mutation *PROM1* c.619G>T (p.E207X) and his heterozygous mutated mother underwent detailed ophthalmic examinations. Induced pluripotent stem cells were established from their urine cells, and induced into retinal organoids (ROs). A plasmid with normal human *PROM1* was constructed and delivered by adeno-associated virus (AAV) to both ROs and PROM1-KO mice to rescue the mutated *PROM1*.
Results: The IRD patient had low vision, impaired retinal structure with loss of photoreceptors, and extinguished ffERG in both eyes. His mother displayed no significant abnormalities in her eyes. Patient-specific ($PROM1^{Mut/Mut}$) ROs cultured for more than 120 days (D) lacked detectable PROM1 and exhibited significant shorter outer segments compared to his mother's and HC's ROs. However, infection of $PROM1^{Mut/Mut}$ ROs with AAV-hPROM1 at D90 and afterwards recovered the expression and localization of the normal PROM1 protein, enhancing the outer segment structure in photoreceptors one-month post-treatment. This therapeutic effect was confirmed in *Prom1*$^{-/-}$ mice, where subretinal injection of AAV-hPROM1 led to specific expression of human PROM1 in mouse photoreceptor outer segments and improvement of the a- and b-wave amplitudes in ffERG.
Conclusion: Our study firstly confirmed the relationship of phenotype and genotype of this novel mutation PROM1 c.619G>T through patient specific ROs model. The gene therapy not only restored the expression of normal PROM1 but also recovered diseased retinal structure and function, holding a promise as a potential treatment for PROM1-IRD.
Funding: This research was supported by grants from the National Natural Science Foundation of China (82172957, 81970842); the National Key R&D Program Project of the Ministry of Science and Technology (2023YFC2506100); the Science & Technology Project of Guangdong Province (2017B020230003); the National Key Research and Development Program of the Ministry of Science and Technology (2017YFA0104101); a joint grant from Science & Technology Project of Guangzhou and Zhongshan Ophthalmic Center, Sun Yat-sen University (202201020312).

Session Title: Genetic Therapies Poster Thursday Session

Board 8093T: Prediction of *P.falciparum adenylosuccinate lyase* inhibitors from african natural product databases

Authors:

M. Sangare; African Ctr. of Excellence in Bioinformatics, Bamako, Mali

Abstract:

Malaria is one of the world's most deadly infectious diseases. According to the world malaria report 2023, there were an estimated 249 million cases of malaria worldwide . Sub saharan Africa remains the most affected region, with more than 90% of the global burden of P. falciparum. The emergence and spread of malaria parasites resistant to anti-malarial drugs are major obstacles in the fight against this disease. The discovery of new drugs is always a necessity, which requires the identification of new therapeutic targets. In this study, we have presented Plasmodium falciparum adenylosuccinate lyase (Pf ADSL) as a highly promising therapeutic target. It is an important enzyme involved in the de novo pathway of purine biosynthesis, unique in P. falciparum, which result to the formation of Adenosine Monophosphate (AMP), catalyzed by ADSL.The aim of this study is to model the three-dimensional (3d) structure of Pf ADSL and to predict its potential inhibitors from african natural products.MethodsThe 3D structure of Pf ADSL was modeled using AlphaFold 2. The protein was then subjected to a 50 ns molecular dynamics simulation (MDS) with GROMACS (version 2020.2), to determine the different conformations of the main pocket. Three (03) conformations of the main pocket were identified for virtual screening. Natural compounds were extracted from Afrodb, Sancdb, ANPDB. Virtual screening was performed with autodock-vina, ADMET predictions with SwissADME and protein-ligand interaction studies with Discovery studio.MDS was performed at 100ns with NAMD for the protein-ligand complex.Result After virtual screening, the ten (10) compounds with the best binding energies and ADMET properties were selected for each database and conformation. binding energies ranged from -9.8 to -8 kcal/mol. Molecular dynamics simulation of the Pf ADSL SANC00584 complex showed very high stability.Conclusion This study allowed us to build the 3D model of PFADSL. Virtual screening identified eight inhibitors as the most promising . Molecular dynamics was performed with one compound and showed very high stability.These results provide a solid basis for the fight against malaria. Further experimental studies will be required to assess the antimalarial activity of the identified inhibitors.

Session Title: Genetic Therapies Poster Thursday Session

Board 8094T: Selective JAK2 specific ASO based potential treatment for MDS and PV

Authors:

B. Przychodzen[1], S. Smieszek[2], C. M. Polymeropoulos[3], G. Birznieks[1], M. H. Polymeropoulos[1]; [1]Vanda Pharmaceuticals, Washington, DC, [2]Vanda Pharmaceuticals Inc., Bethesda, MD, [3]Vanda Pharmaceuticals, Bethesda, MD

Abstract:

Myeloproliferative disorders (MPD) are clonal hematopoietic stem cell malignancies with cytokine independency or hypersensitivity. Increased JAK2 kinase activity is observed in hematologic malignancies; somatic *JAK2*^{V617F} gain-of-function mutations are found in at least 95% of PV, We hypothesize that preventing JAK2 transcription by antisense oligonucleotide (ASO) mediated exon masking of the *JAK2* intron-exon junction will result in reduced *JAK2* mRNA, and thus JAK2 protein, by providing nonsense mediated decay (NMD) in the reading frame. We designed a series 19mer ASO targeting *JAK2* exon-intron junctions (*JAK2* intron-exon junction providing NMD in thereading frame, steric, non-RNAse H1) and tested these at a range of concentrations. We observed significant JAK2 protein decrease (~50%) in ASO-treated samples compared to untreated samples. JAK2 qPCR results confirmed 40-60% of target transcript. STAT5 phosphorylation status further confirmed this effect, and we report a 35% pSTAT5 reduction. JAK1 activity was not affected and the designed ASO is JAK2 specific. Our preclinical data support this ASO as a highly selective JAK2 agent, affecting direct levels of JAK2 as well as downstream STAT signaling. The current treatments target a family of kinases leading to multiple unwanted side effects. Currently humanized mice studies are being conducted testing safety and efficacy of this ASO. ASOs as a class of drugs offer the capability to directly target JAK2 mutations with high precision and effectively reduce JAK2 protein production, without off-target kinase effects.

Session Title: Genetic Therapies Poster Thursday Session

Board 8096T: The RNAs contained in extracellular vesicles from the retinal pigment epithelium show potential as circulating biomarkers of retinal degeneration

Authors:

E. Pierce[1], H. Getachew[2], T. Kaur[1], K. Valestil[1], S. Mehrotra[1], M. Garita-Hernandez[1]; [1]Massachusetts Eye & Ear Infirmary, Boston, MA, [2]Prime Med.s, Cambridge, MA

Abstract:

Background: Inherited retinal degenerations (IRDs) are important causes of vision loss. Genetic therapies show great promise for treating IRDs. A challenge with clinical translation of these therapies is the slowly progressive nature of many IRDs; it can take multiple years for the beneficial effects of stopping disease progression to be detected by measurements of visual function. This duration is impractical for clinical trials, hence great interest in identifying surrogate biomarkers for quicker detection of therapeutic responses. The RNAs in extracellular vesicles (EVs) are promising accessible biomarkers of retinal health. EVs are cell-derived membranous structures produced by all cells that carry diverse cargo including proteins, RNA and DNA. Liquid biopsies of circulating EV contents are being used to monitor tumor burden in multiple types of cancer. We hypothesize that the RNA contents of circulating EVs derived from the neural retina and retinal pigment epithelium (RPE) can be used as biomarkers of retinal health and disease. **Methods:** To test this hypothesis for the RPE, we used pre-mRNA processing factor 31 (PRPF31) mutant (*PRPF31*$^{+/-}$) and isogenic control (*PRPF31*$^{+/+}$) human induced pluripotent stem cell (hiPSC)-derived RPE (hiPSC-RPE) to investigate the RNAs contained in RPE-derived EVs, and how they change in disease. Mutations in PRPF31 are associated with autosomal dominant RP, and cause cell autonomous defects in RPE function. We isolated EVs secreted by the *PRPF31*$^{+/-}$ and *PRPF31*$^{+/+}$ hiPSC-RPE cells into the culture media, and sequenced the miRNA and poly-A RNA contents of the EVs. **Results:** A total of 135 miRNAs were identified in the RPE EVs, with 12 being enriched and 6 being depleted in the EVs from the mutant *PRPF31*$^{+/-}$ RPE cells. A total of 31,370 poly-A RNAs were also identified in the RPE EVs, with 551 enriched and 314 depleted in the mutant *PRPF31*$^{+/-}$ RPE cells. The EVs contained 89% of the RNAs detected in the RPE cells. Functional over-representation analysis showed that 11 of the 18 differentially enriched miRNAs were associated with RPE cell function or cell de-differentiation and death. Analysis of RNA biotypes of RPE cells and their EVs revealed a higher abundance of long non-coding RNAs (lncRNAs) in the EVs. LncRNAs have regulatory functions, and some are associated with ocular diseases. **Conclusions:** The results show that EVs from mutant *PRPF31*$^{+/-}$ hiPSC-RPE cells contain distinct RNAs compared to EV from the control *PRPF31*$^{+/+}$ hiPSC-RPE cells. This suggests that the composition of EV cargo, including miRNA and lncRNAs, reflects the physiological and pathological state of their parent cells and can be used as biomarkers of retinal degeneration.

Session Title: Genetic Therapies Poster Thursday Session

Board 8097T: Using the zebrafish to model*KRAS*-related complex lymphatic anomaly

Authors:

S. Sheppard[1], B. Sempowski[2], D. Shukla[2], S. Paulissen[2], C. Marshall[2], G. Margolin[2], R. Dale[2]; [1]NIH, Rockville, MD, [2]NIH, Bethesda, MD

Abstract:

Complex lymphatic anomalies (CLAs) are characterized by lymphatic malformations in multiple organ systems leading to significant disability such as chylothorax or chylous ascites. There is an urgent need to develop therapies for this rare disease population. Mosaic pathogenic variants in *KRAS* cause three subtypes of CLAs (Sheppard et al 2023, Li*,Sheppard* et al 2023). MEK inhibition ameliorates abberant lymphatic phenotypes for the p.G12D and p.G13D variants both *in vitro* and *in vivo* (Sheppard et al 2023). The purpose of this study was to determine efficacious therapies for multiple *KRAS* variants using sequencing and the zebrafish model. Zebrafish have well conserved genetics with humans, develop rapidly and externally allowing visualization of development, and have large clutches which are amenable to drug screening. Using the *mrca1* promoter, wild type or mutant human KRAS was mosaicly expressed in the lymphatic and venous endothelial cells of the zebrafish. Mosaic expression of p.G12C, p.G12D, and p.A146T increased pericardial edema to 53% (n=90), 69% (n=65), and 25% (n=203), respectively. Cystic malformations were observed in 38% (n=90) of the p.G12C expressing larvae and 66% (n=65) of the p.G12D expressing larvae in the caudal hemopoietic tissue. All three variants caused an increase in the diameter of the intersegmental lymphatic vessels (KRAS WT, 7.448 uM (ns); p.G12C, 22.76 uM (p<0.0001); p.G12D, 14.44 uM (p<0.0001); p.A146T 21.12 uM (p<0.0001)) and in the posterior cardinal vein (KRAS WT, 40.50 uM (ns); p.G12C, 100.90 uM (p<0.0001); p.G12D, 56.56 uM (p<0.0001); p.A146T 88.23 uM (p<0.0001)), demonstrating they are excellent models for the human disease. KRAS p.G12C also caused dilations in the parachordal line (KRAS WT, 3.389 uM (ns); p.G12C, 12.73 uM (p<0.0001)) and lateral facial lymphatic KRAS WT, 16.64 uM (ns); p.G12C, 30.25 uM (p<0.0001)). Bulk RNA-sequencing analysis for p.G12D demonstrated an upregulation of lymphangiogenesis genes and identified additional pathways for targeting. Studies are ongoing to compare the efficacy of previously FDA approved MEK inhibitors, mTOR inhibitors and other RAS pathway inhibitors- essentially all which could eventually be used off label or in clinical trials for the treatment of these devastating diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1042T: A compendium of somatic mutations across genomes, tissues, and individuals

Authors:

H. Xu, R. Bierman, D. Akey, J. Akey; Princeton Univ., Princeton, NJ

Abstract:

A human zygote undergoes trillions of cell divisions, resulting in a fully differentiated multicellular organism. This intricate process creates numerous opportunities for somatic mutations to occur. Even terminally differentiated cells continue to acquire somatic lesions in their genomes through non-replicative processes, such as the spontaneous deamination of methylated cytosines in the context of CpG sites. Consequently, humans are expected to harbor a significant burden of somatic variation, which contributes to cancer, aging, neurodegeneration, and an expanding list of additional human diseases. Although somatic mutations are fundamentally important to human biology and disease, many outstanding questions remain about their rates, spectrum, and determinants in healthy individuals. The two main impediments to a more comprehensive understanding of human somatic mosaicism are the lack of appropriate datasets where multiple tissues from multiple individuals have been sampled and powerful statistical methods tailored to the unique structure of this data. In this study, we describe the analysis of somatic mutations in a dataset comprised of high-coverage (~160X) whole-genome sequencing of 300 GTEx tissue samples, spanning seven tissues from 55 donors. Using a novel probabilistic approach, we identify a compendium of somatic variants that we leverage to quantify the rates and mutational spectra of somatic variants across tissues and individuals. We show that the burden of deleterious somatic mosaicism varies widely across individuals and tissues. We also show that age is significantly related to the burden of somatic mosaicism, but only in a subset of tissues. Interestingly, we find that non-coding somatic variants are enriched compared to codoing somatic variants in all tissues except sun-exposed skin. We hypothesize this observation is driven by UV-light-induced somatic mutations in the coding regions, which we show is consistent with the mutational spectrum in sun-exposed skin. We also observe that the spectrum of somatic mutations is significantly different between coding and noncoding regions, and explore how chromatin structure contributes to this difference. In summary, our data provides a high-resolution compendium of somatic variants across genomes, tissues, and individuals, and offers novel biological insights into mutational mechanisms.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1043T: A genome-wide association study of multiple epigenetic age accelerations identifies novel genetic variants within *ENO2* in an East Asian population ⭐

Authors:

K. Oh[1], M. Yuk[1], S. Yang[1], J. Youn[2], D. Yun[1], Q. Dong[3], Z. Wang[3], N. Song[1]; [1]Chungbuk Natl. Univ., Cheongju, Korea, Republic of, [2]Coll. of Pharmacy, Chungbuk Natl. Univ., Cheongju, Korea, Republic of, [3]St. Jude Children's Res. Hosp., Memphis, TN

Abstract:

[Introduction] Since the inception of epigenetic clock about a decade ago, investigations of its use as an aging biomarker including studies of its genetic and non-genetic risk factors as well as its associations with morbidity and mortality have made rapid progress. To understand the genetic underpinning of the epigenetic age acceleration (EAA) (i.e., the difference between epigenetic age and chronological age), several GWASs have been performed based on multiple epigenetic clocks. However, there is no such study in Asian populations. To address the gap, we conducted an EAA GWAS based on five epigenetic clocks in Korean population. Considering differences in genetic ancestry and life expectancy across ethnic groups (e.g., East Asians live longer than other ethnic groups), our research may lead to novel and ethnic-specific findings. **[Method]** Genome-wide SNP and methylation data were generated with blood-derived DNA from 1,962 Korean individuals. Epigenetic age (EA) was estimated using the online DNA Methylation Age Calculator. EAA was calculated as the residual from a linear regression of the EA against the chronological age. Five epigenetic clocks were considered including Hannum, PhenoAge, GrimAge, Zhang's clock, and Horvath-Skin & Blood. Gene-based and pathway enrichment analyses were performed to further explore biological mechanisms. To explore potential clinical implications, EAA-associated SNPs were further assessed for associations with 81 health-related traits. **[Result]** GWAS analysis found that rs3829957 (*ASPA*) and rs74333895 (*CDH4*) were associated with PhenoAge-EAA, rs76713297 (*RP3-340N1.6*) and rs74072052 (Intergenic) were associated with GrimAge-EAA, and rs3741922 (*ENO2*) was associated with Zhang-EAA (all $P<5\times10^{-8}$). Gene-based and pathway enrichment analyses found the association of PhenoAge-EAA with *ASPA* and *SPATA22* ($P_{Bonferroni<}2.66\times10^{-6}$), involved in Canavan disease and apoptotic signaling, and the association of Zhang-EAA with *ENO2 GNB3*, *USP5*, *TPI1*, and *CDCA3* ($P_{Bonferroni}<2.66\times10^{-6}$), implicated in the monosaccharide catabolism pathway in

hepatocytes. Notably, rs3741922 (*ENO2*), a Zhang-EAA associated SNP, showed pleiotropy in plasma iron and bilirubin, biomarkers of liver dysfunction. **[Conclusion]** We report five genetic variants associated with EAA in East Asians, of which four were novel (rs74333895, rs76713297, rs74072052, and rs3741922). Especially, *ENO2* and nearby genes were intriguing with respect to their potential roles in the monosaccharide catabolism process as well as molecular biomarkers of liver dysfunction.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1044T: A global reference map of STR and VNTR composition using vamos

Authors:

B. Gu, M. Chaisson; Univ. of Southern California, Los Angeles, CA

Abstract:

Tandem repeats (TRs), including short tandem repeats (STRs) and variable-number tandem repeats (VNTRs), are hypermutable genetic elements consisting of tandem arrays of repeated motifs. TR variation can modify gene expression and has been implicated in over 50 diseases through repeat mutation and pathogenic expansion. Both classes of TRs exhibit mutation rates orders of magnitude higher than single nucleotide variants and can vary in both motif count and composition. Our understanding of the global diversity of TR sequences has been limited by the inability of short-read sequencing to accurately resolve these highly polymorphic regions. However, recent advances in long-read sequencing (LRS) now enable the comprehensive profiling of TR variation in large-scale cohort studies.

We previously developed vamos, a tool for annotating motif count and composition in LRS samples. Here, we applied vamos to 360 publicly available LRS assemblies spanning globally diverse ancestries to construct TRcompDB, a reference database of tandem repeat variation across 805,509 STR and 370,374 VNTR loci. Despite their hypermutability, TR sequences in TRcompDB cluster by continental population group, concordant with patterns of single nucleotide variation. Furthermore, we identified 10,068 loci exhibiting strong signatures of population divergence in TR composition.

With phased single nucleotide variants from the assemblies, we reconstructed ancestral relationships between samples at each TR locus using ancestral recombination graphs. This enabled estimation of locus-specific mutation rates that account for both length and sequence composition changes annotated by vamos. The mutation rates loosely correlate with mean

locus length (r2=0.2), and span five orders of magnitude (10^-5 to 10^-10 mutations/generation) after controlling for length. Strikingly, our mutation rate map revealed that 40.6% of TR loci are mutationally constrained across all populations, suggesting widespread functional importance.
A key challenge for future LRS-based disease studies will be distinguishing pathogenic repeat expansions from naturally occurring variation. TRcompDB provides a critical resource for modeling the expected spectrum of TR polymorphism across diverse ancestries, facilitating the identification of potentially deleterious repeat mutations in disease cohorts.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1045T: A population-specific pangenome for Saudi Arabia

Authors:

M. Kulmanov[1], M. Abdelhakim[1], M. Alarawi[1], H. Aldakhil[1], D. Alhattab[1], P. Driguez[1], Y. Liu[1], R. Tawfiq[1], M. Abedalthagafi[2], R. Hoehndorf[1]; [1]King Abdullah Univ. of Sci. and Technology, Thuwal, Saudi Arabia, [2]Emory Sch. of Med., Atlanta, GA

Abstract:

Pangenomes represent the full set of genes in a species or population, and pangenome graphs are data structures designed to represent these pangenomes. Recently, pangenome graphs have been used as references for human genomics instead of linear reference genomes. However, the existing pangenome graphs do not include individuals from the Arab World. To address this, we developed a population-specific pangenome for Saudi Arabia based on the genomes of nine individuals from five regions within the country. We sequenced each individual and generated haplotype-resolved assemblies using two different assembly strategies.
We then employed Minigraph-Cactus and the Pan-Genome Graph Builder (PGGB) to generate two pangenome graphs from the 18 haplotypes, the GRCh38, and CHM13 assemblies. We developed a variant calling workflow using Giraffe and DeepVariant tools to enable the Saudi pangenome to be used as a reference for short reads. Our results demonstrate that the pangenome improves the accuracy of variant calls compared to using single reference genomes such as GRCh38 and CHM13, enhancing the potential for identifying pathogenic variants.
The pangenome includes individuals from various regions in Saudi Arabia, capturing aspects of the historic tribal structure of the Saudi population. To promote its use as a reference and encourage widespread adoption, we have made all data, including the sequencing reads,

assembled genomes, pangenome graph, and variant calling workflows, freely accessible to the scientific community.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1046T: A Public Resource of gnomAD Local Ancestry Inference Data for Diverse Human Genomes

Authors:

P. Kore[1], M. Wilson[2], Z. Koenig[3], G. Tiao[4], K. Chao[2], gnomAD Consortium, A. Martin[5], M. Daly[5], K. Karczewski[6], E. Atkinson[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Broad Inst., Cambridge, MA, [4]Broad Inst MIT & Harvard, Somerville, MA, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Massachusetts Gen. Hosp., Medford, MA

Abstract:

The Genome Aggregation Database (gnomAD) has been an invaluable resource for the genomics community, quantifying allele frequencies across diverse genetic ancestry groups. However, current gnomAD allele frequencies aggregate across largely continental-based groups without accounting for recent admixture, potentially conflating fine-scale genetic differences. This can lead to clinical misinterpretations in underrepresented groups and propagation of misinformed and harmful ideas that discrete continental genetic ancestry groups exist. To address this, we have conducted local ancestry inference (LAI) to derive more precise genetic ancestry-specific allele frequencies, enhancing the granularity and accuracy of genomic data - particularly for more recently admixed individuals with genetic similarity spanning multiple genetic ancestry groups.

Here we extend the local ancestry work from the inferred Admixed American genetic ancestry group (n=7,612) to the inferred African/African American group (n=20,805) within gnomAD. Using a curated reference panel composed of continental AFR and EUR populations from the 1000 Genomes Project and the Human Genome Diversity Project, we conducted joint phasing and LAI. This allowed us to assign more precise genetic ancestry to each chromosomal segment in African/African American individuals. We then used the Tractor tool to calculate local ancestry dosages at each site, enabling us to compute ancestry-specific allele frequencies in this massive-scale database. The widely-used gnomAD browser will serve as a key resource to advance research involving more globally representative groups. The computational pipeline, written in Hail Batch, is publicly available via GitHub. The allele frequencies and quality metrics can be downloaded or interactively explored on the gnomAD website.

LAI enables identification of disease-risk variants that occur at higher frequencies within specific genetic ancestry components due to genetic drift. For example, the gnomAD variant 17-7043011-C-T in *SLC16A11*, associated with type 2 diabetes risk, shows a 45% allele frequency in Amerindigenous regions of the Admixed American genetic ancestry group, compared to a 2% gnomAD-wide ("global") frequency. This variant example highlights the importance of more precise allele frequencies in accurately identifying the distribution of genetic risk factors within diverse genetic ancestry groups.
By calculating and releasing LAI data, we provide a higher-resolution view of allele frequency, which is obscured when assessing frequency at the genetic ancestry group-level, enabling more equitable and precise genomic research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1048T: Accelerating Medicines Partnership® Common Metabolic Diseases (AMP CMD): Building resources to generate prioritized targets for all metabolic diseases

Authors:

R. Fischer; Fndn. for the NIH, Washington, DC

Abstract:

The increasing burden and prevalence of common metabolic diseases (CMDs), including obesity, atherosclerotic cardiovascular disease (ASCVD), heart failure (HF), type 2 diabetes (T2D), pre-diabetes, metabolic dysfunction-associated steatohepatitis (MASH), and chronic and diabetic kidney disease (CKD, DKD), represent a major challenge for global public health. Additionally, the presence of one of these disorders is associated with increased risk for an individual to develop additional CMDs. Therapeutic interventions for these disorders have been primarily targeted toward single diseases. However, recent therapeutics discovered for diabetes also reduce risk for renal and cardiovascular outcomes, suggesting common pathogenic drivers likely exist for these diseases and intercepting these common disease drivers may be tractable, with a potential to develop therapies to modify the course of many CMDs. The Accelerating Medicines Partnership® Common Metabolic Diseases (AMP CMD) is a collaborative partnership among government (NIDDK) and private-sector partners (Pfizer, Novo Nordisk, Eli Lilly, and Amgen) that harnesses their collective capabilities, scale, and resources to therapeutically address multiple CMDs that share common pathogenic drivers and overlapping molecular pathways. The AMP CMD Knowledge Portal (CMDKP) and Genome Atlas (CMDGA) are open access data portals that

aggregate, analyze, and display summary statistics of genetic, genomic, and epigenetic data from subjects with CMDs and related complications. These portals host analysis and visualization tools, freely available for use by the community. The goals of AMP CMD are to amplify these resources by: 1) Expanding the CMDKP through generation of new and leverage of existing data on CMDs; 2) Addressing data gaps across all CMDs; 3) Evolving the CMDKP by development of analytic and visualization tools to support integrative analysis; and 4) Delivering prioritized targets for CMDs with supporting mechanistic data. The CMDKP currently contains 605 datasets and 704 phenotypes, from 50 countries incorporating data from diverse populations. The CMDGA currently includes 5810 assays, 8035 annotations, 49 single cell embeddings, and 18 perturbations datasets, along with supporting information such as data standards and QC metrics, pipelines, and publications. Both portals continue to expand, gathering additional genetic, genomic, and epigenomic data from heart, liver, pancreas, skeletal muscle, adipose, brain, and kidney, across all CMDs. Data from AMP CMD-funded projects will continue to be deposited into the open-access portals, along with other community-submitted data.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1049T: Accurate and robust cell type deconvolution with multiple single cell reference panels in spatial transcriptomics studies

Authors:

P. Zhao[1], Y. Ma[2], X. Zhou[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Brown Univ., Providence, RI

Abstract:

Spatial transcriptomics studies are becoming increasingly common, providing novel biological insights across genetics domains. The majority of these studies rely on technologies with relatively low spatial resolution, requiring subsequent cell type deconvolution analysis to extract cell type composition and localization information on complex tissues. Existing deconvolution methods often rely on one single cell RNA sequencing (scRNAseq) reference data for cell type devolution. However, using one scRNAseq reference is not ideal, as it is often not straightforward to determine which scRNAseq reference data to use given the wide availability of scRNAseq datasets measured in the same tissue. Here, we present a computation method called MultiCARD (Multi-reference weighted conditional autoregressive-based deconvolution), that models multiple scRNA-seq references jointly for cell type deconvolution in spatial transcriptomics. MultiCARD can make use of scRNAseq references collected from distinct

technologies and can automatically determine the relative contribution of these scRNA-seq references towards deconvolution, thus ensuring accuracy and robustness. We demonstrate the advantages of MultiCARD through simulations and in-depth analysis of two real datasets collected from distinct tissues and technologies. In the real data, MultiCARD identified unique tissue regions in early-stage embryos with precise cell type composition estimates, uncovering cellular heterogeneity not revealed by existing approaches.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1050T: Accurate genotype calls in complex genetic regions using microarrays

Authors:

F. Lin, A. Mittal, R. Varma, B. Carsten, J. Gollub; Thermo Fisher Scientific, Santa Clara, CA

Abstract:

Genotyping in complex genetic regions is challenging due to copy number (CN) variations, particularly in gene families with high homology. For example, *RHD*, one of the two genes underlying the Rh blood group system, shares high homology to *RHCE*. The frequent deletion of *RHD* results in the D-/- type, which can impact the calling of *RHCE*, the gene underlying the Rh Cc red blood cell antigen. Similarly, *CYP2D6*, responsible for metabolizing approximately 25% of prescribed medications, shares significant homology with the pseudogene *CYP2D7*. Accurate genotyping of these genes requires precise CN assessments. Similarity between gene family members complicates this task, as potential off-target matches can occur in regions with their own CN variations or SNPs. For example, *RHCE* genotyping is complicated by frequent deletion of identical sequences in *RHD*, and *CYP2D6* genotyping is confounded by variations in the *CYP2D7*. These challenges hinder accurate genotyping of variants in these regions using short-read sequencing, other 3rd party microarray platforms and technologies such as low pass sequencing (LPS) that require imputation. The Applied BiosystemsTM AxiomTM PangenomiX array addresses these challenges using multiplex PCR amplification and CN-aware calling algorithms. Designed for broad genomic coverage, the array incorporates over 800,000 markers representing diverse populations, including European, African, American, East Asian, and South Asian groups. The array includes markers pertinent to disease risk, blood group typing from AxiomTM BloodGenomiX array, and pharmacogenomic factors. In this study, we validate the array's efficacy in genotyping complex regions for blood typing and pharmacogenomics using 89 HapMap samples of diverse ancestries. We compared the direct genotyping from the

PangenomiX array to short-read sequencing and imputation of short-read data using state-of-the-art TopMed reference panel. Preliminary results show that out of 10 *RHD-RHCE* variants analyzed, the array genotyped all samples correctly, whereas short-read sequencing caused errors in samples with gene deletions in 5 variants. Additionally, rs676785, critical for Rh Cc antigen, was not called in short-read sequencing or imputation with TopMed. Similarly, out of 129 variants analyzed in *CYP2D6*, the array genotyped all samples correctly, whereas short-read caused errors in 5 variants and failed to call 43. In conclusion, our results show that the PangenomiX array outperforms short-read sequencing and technologies such as LPS that require imputation, suggesting microarrays are well suited for predictive genomics research applications. For research use only.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1051T: An Integrated Database of Experimentally Validated Major Histocompatibility Complex Epitopes for Antigen-specific Cancer Therapy

Authors:

C. Jiang[1], S. Kawakita[1], C-c. Chao[1], S. Cheng[2], B. Li[3], Z. Wang[1]; [1]Terasaki Inst. for BioMed. Innovation, Los Angeles, CA, [2]USC, Los Angeles, CA, [3]California State Univ., Northridge, Los Angeles, CA

Abstract:

Background: Cancer immunotherapy represents a paradigm shift in oncology, offering a superior anti-tumor efficacy and the potential for durable remission. The success of personalized vaccines and cell therapies hinges on the identification of immunogenic epitopes capable of eliciting an effective immune response. Current limitations in the availability of immunogenic epitopes restrict the broader application of such therapies. A critical criterion for serving as potential cancer antigens is their ability to stably bind to the major histocompatibility complex (MHC) for presentation on the surface of tumor cells. To address this, we have developed a comprehensive database of MHC epitopes, experimentally validated for their MHC binding and cell surface presentation.Findings: Our database catalogues 451,065 MHC peptide epitopes, each with experimental evidence for MHC binding, along with detailed information on HLA allele specificity, source peptides, and references to original studies. We also first provide the data of grand average of hydropathy scores and predicted immunogenicity for the epitopes. The database (MHCepitopes) has been made

available on the web. Conclusion: By consolidating empirical data from various sources coupled with calculated immunogenicity and hydropathy values, our database offers a robust resource for selecting actionable tumor antigens and advancing the design of antigen-specific cancer immunotherapies. It streamlines the process of identifying promising immunotherapeutic targets, potentially expediting the development of effective antigen-based cancer immunotherapies. Keywords: Major Histocompatibility Complex; Epitopes; Cancer Immunotherapy; Immunogenicity

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1052T: Analysis Strategy for Whole Genome Data in a National Genome Project in Korea

Authors:

J. Kim, J. JEON, S. Joe, K-H. Cho, I-K. Lim, S. Jeon, J. Yang, H. Jeong; Korea research Inst. of bioscinece and biotechnology, Daejeon, Korea, Republic of

Abstract:

The National Genome Project (NGP) in Korea endeavors to generate data to anticipate and diagnose prevalent diseases among Koreans, engaging citizens in the process, thereby facilitating an examination of the interplay between genetics and environmental factors in disease manifestation. Notably, diseases are not solely influenced by genetic determinants but are also shaped by environmental contexts, illustrated by discordant diseases observed in identical twins raised in disparate environments. A comprehensive analysis of diverse Korean genomes can elucidate the environmental factors contributing to disease susceptibility, thereby aiding in the development of tailored preventive measures and therapeutic interventions. The accumulation of extensive genome data is fundamental for the advancement of precision medicine, fostering progress in medical technology and public health outcomes. This investigation refines and complements the large-scale whole-genome analysis framework utilized in the pilot project conducted from 2021 to 2022, with the objective of enhancing its applicability to NGP phase 1, scheduled to commence from 2024 and span five years. Within this framework, we have developed analysis strategies capable of analyzing approximately 100,000 samples annually, addressing germline and somatic variant calls, as well as joint genotype analyses. Additionally, our approach encompasses processes such as WGS-based gender analysis and familial relationship analysis. These outcomes present a strategic

foundation for planning computational resources that can support consistent analysis capacity irrespective of the scale of analysis, facilitating effective resource allocation.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1053T: Bioinformatic analysis of tuberculosis to identify new therapeutic targets through drug repurposing approaches

Authors:

A. Pratap[1], T. Abdulrehman[2], A. Titus[3], S. Khader[4], K. K. Yadav[5]; [1]Infinite Computer Solutions, Noida, India, [2]eHlth.Program, DeGroote Sch. of Business, McMaster Univ., Hamilton, ON, Canada, Hamilton, ON, Canada, [3]3Dept. of Preventive Cardiology, Houston Methodist, Houston, TX, USA;, Huston, TX, [4]Sch. of Publ. Hlth., Faculty of Med., Imperial Coll. London, South Kensington, London, UK, South Kensington, United Kingdom, [5]Sch. of Engineering Med.,Texas A&M Univ., Houston, TX, USA., Huston, TX

Abstract:

Tuberculosis (TB) is a contagious bacterial infection primarily affecting the lungs caused by Mycobacterium tuberculosis. It remains a major global health issue, causing significant morbidity and mortality, particularly in developing countries. TB manifests in multiple forms — pulmonary TB, active TB, and latent TB, each presenting distinct diagnostic and treatment challenges. Understanding the genetic basis of TB across these variations is paramount for developing targeted therapies. This study aims to analyze prevalent genes and pathways associated with all three TB types and identify drugs that could be repurposed for TB variant-based targeted therapy. Gene data for each TB variant was obtained from DisGeNET, a comprehensive database of disease-gene associations. Further analysis was performed using the functional enrichment analysis tool ENRICHR for gene ontology, pathway analysis and gene-signature-based drug discovery. We identified 101 common genes among the three TB variants that were assessed for therapeutic match. The analysis also revealed common key pathways among TB variants involved in TB pathogenesis which include “Cytokine-cytokine receptor interaction”, “Inflammatory bowel disease”, and “Leishmaniasis”. The most significant drug identified from the IDG Drug 2022 database for pulmonary, active and latent TB, respectively were Adenosine Triphosphate (Padj: 5.33684564977574E-07, Genes: P2RX6; P2RX5; P2RX4; P2RX3; P2RX2; P2RX1; P2RY2; P2RY1), Quercetin (Padj: 7.85767904462568E-10, Genes: ABCB1; GLO1; ALOX15; AKR1B1; ALOX12; PTGS2; MPO; PIK3CG) and Budesonide (Padj: 0.00013, Genes: ABCB1; IL5; IFNG; MMP9). Further research, and validations of the findings are required to

develop novel targeted treatments for the TB variants. This study provides comparative insights into the genetic landscape of the TB variants and suggests that integrating gene-disease associations with the biological pathway analysis could contribute to computational drug repurposing with a targeted approach facilitating improved management and control of TB globally

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1054T: Carrier Screening in the Middle Eastern Population: A Study of the Qatar Genome Programme ⭐

Authors:

E. Nkrumah[1], H. M. Al-Mulla[1], V. Soman[2], C. Saad[3], A. Bashar[4], J. Wang[5], U. Chandran[2], H. Mbarek[3], M. Aarabi[6], M. Aarabi[7]; [1]Sarah Lawrence Coll., Bronxville, NY, [2]Dept. of BioMed. Informatics, Univ. of Pittsburgh, Pittsburgh, PA, [3]Qatar Genome Program, Qatar Fndn. Res., Dev. and Innovation, Qatar Precision Hlth.Inst., Doha, Qatar, [4]Univ. of Colorado Boulder, Boulder, CO, [5]Div. of Genetic and Genomic Med., Dept. of Pediatrics, Pittsburgh, PA, [6]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [7]Univ Pittsburgh Med Ctr, Pittsburgh, PA

Abstract:

Carrier screening is used to identify individuals who are at-risk of having a child with a recessive genetic condition. The goal of carrier screening is to allow individuals or couples without a known personal or family history of autosomal recessive (AR) or X-linked recessive (XLR) conditions to learn about potential risks that may affect their offspring and to take the steps needed to mitigate such risk. Many countries define preconception and reproductive screening in different terms and screen for different conditions. In 2021, ACMG released its carrier screening practice resource for AR conditions and recommended offering carrier screening during pregnancy and preconception for 97 conditions with a carrier frequency of 1/200 and above in the Genome aggregation database (gnomAD) v2; however, the Middle Eastern population was not represented in the gnomAD v2. Furthermore, genome data from only 158 Middle Eastern individuals were included in the 2021 updated version of gnomAD v3. This study aimed to interrogate the carrier frequency for AR conditions based on the genome data from 14,392 individuals included in the Qatar Genome Project. Carrier frequency for pathogenic/likely pathogenic variants in 2,987 genes associated with AR conditions obtained from ClinVar were calculated in the Qatari genomes. Genes with a carrier frequency of 1/200 and above were further curated for the severity of their associated clinical presentations. We

identified 69 genes with a carrier frequency of 1/200 and above and moderate to severe presentation, of which only 16 genes were shared with the ACMG list of suggested genes for carrier screening. The study also discovered certain genetic conditions with high carrier frequency within the Qatari population, such as immunodeficiency due to CD25 deficiency, spinocerebellar ataxia, combined deafness and blindness, congenital adrenal hyperplasia, and Niemann-Pick disease. This study highlights the importance of incorporating population-specific data into the development of carrier screening programs. Our findings in the Qatari population highlight the limitations of relying on generalized, global datasets and the need for targeted, population-specific research to accurately identify the genetic diversity and disease prevalence within specific ethnic and regional populations. We propose further research on the implementation of a population-specific carrier screening panel for the Qatari population based on the results of this study, which could have a significant impact on reproductive health and care in the Middle East region.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1055T: Characterization of AZF Deletions on the Y Chromosome Using Nanopore Long-Read Sequencing: Advances and Diagnostic Applications

Authors:

A. Zrhidri, S. Kemeny, E. Page, **G. Egea**; Inovie - GenBio, Clermont Ferrand, France

Abstract:

AZF deletions are genomic deletions in the euchromatic part of the long arm of the human Y chromosome (Yq11) associated with azoospermia or severe oligozoospermia. Consequently, it can be assumed that these deletions remove Y chromosomal genes required for spermatogenesis. However, the traditional AZF deletions, AZFa, AZFb and AZFc, are just a subset of rearrangements within Yq11. With advancements in Y chromosome sequencing, additional rearrangements such as deletions, duplications, and inversions, both within and beyond the classical AZF deletion regions, have been uncovered.
Nanopore long-read sequencing technology has revolutionized the field by enabling real-time monitoring and control of individual nanopores, facilitating a process known as "adaptive sampling." This technology allows for the selective enrichment or depletion of specific sequences during DNA sequencing. The application of nanopore long-read sequencing in the study of AZF deletions offers unprecedented resolution and accuracy in detecting and

characterizing these genomic rearrangements. This approach not only improves our understanding of the genetic basis of male infertility but also provides a powerful tool for the identification of novel structural variants within the Y chromosome, potentially leading to improved diagnostic and therapeutic strategies for affected individuals.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1056T: Characterizing the natural history of *SLC6A1* disorder: Insights from the Ciitizen database

Authors:

S. Sengupta[1], N. Chourasia[2], M. Demock[2], E. Bonkowski[3], E. Almanza Fuerte[4], H. Mefford[5]; [1]St. Jude Children s Res. Hosp., Memphis, TN, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [3]Univ. of Washington / St. Jude Children's Res. Hosp., Memphis, TN, [4]St Jude Children s Res. Hosp., Memphis, TN, [5]St. Jude Children`s Res. Hosp., Memphis, TN

Abstract:

Background: *SLC6A1*-related neurodevelopmental disorder (*SLC6A1*-NDD) is a genetic disorder characterized by varying degrees of developmental delay and/or intellectual disability, typically resulting from de novo pathogenic variants in the *SLC6A1* gene. Additional features may include hypotonia, epilepsy, movement disorders, autism spectrum disorder, ADHD, aggression, anxiety, and/or sleep problems. This research aims to conduct a comprehensive digital natural history analysis of individuals with *SLC6A1*-related disorders using electronic health record data from the Ciitizen® database. The main objective is to gather information about the clinical, demographic, and genetic traits of patients and gain a better understanding of the disease progression and outcomes. **Methods:** We use data from the Ciitizen® database to perform descriptive analysis, which summarizes the clinical and demographic characteristics including age, sex, age of onset, seizure types, comorbidities, and family history. Longitudinal analysis will reveal disease trajectories over time, assessing factors such as age at onset of new symptoms, seizure types, and cognitive development. **Results:** We evaluated data for n=47 patients (30 females) at a median age of 6 years (at data collection) with P/LP *SLC6A1* variants. We observed 33 missense, 4 splice, 6 frameshift, 2 nonsense and 2 deletions (whole gene/exon) variants. The diagnosis of global developmental delay was present in 42 (83%), epilepsy in 24 (51%), and autism in 26 (55%). Intellectual disability was diagnosed in 14 (29%). Median age at seizure onset was 3.4 years (SD=1.89). Seizure types included absence in 19 (40%), atonic in 10 (21%) and myoclonic in 8 (17%). Notably, developmental regression was noted in 24 (51%). Domains

associated with developmental regression included fine motor (n=4; 8%), gross motor (n=7; 14%) and language (n=13; 27%). Both motor and language regression were noted in 5 (10%). Most common neurobehavioral diagnoses in the cohort included autism in 24 (51%), aggressive behavior in 19 (40%) and ADHD in 14 (29%). Further analysis is ongoing to better understand the full spectrum of clinical features associated with the disorder. **Discussion:** The results offer valuable insights into the natural development and progression of *SLC6A1*-related NDD. This will help clarify the natural course of the disease in order to facilitate effective treatment strategies and improve medical care. Deeper patterns and correlations analysis could reveal novel associations in the dataset. Overall, this study promotes future research and partnerships to advance the understanding and treatment of these genetic disorders.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1057T: CHIP Pleiotropy and Shared Burden in Heart Failure and CAD

Authors:

J. Kwan[1], A. Jha[2]; [1]Yale, New Haven, CT, [2]Weill Cornell, New York City, NY

Abstract:

Clonal hematopoiesis of indeterminate potential (CHIP) involves mutations in hematopoietic stem cells, leading to clonal expansion. CHIP increases the risk of coronary artery disease (CAD) and heart failure (HF). Research on the shared risk between HF and CAD through pleiotropy is limited. To address this, we used a deep learning approach called the SWAT-CNN model, involving feature correlation analysis, feature selection, and classification to assess the genetic burden in HF and CAD. By adjusting the model, we predicted clone sizes and burden outcomes, considering shared phenotypes. And we used deep learning to identify CHIP-like variants and evaluated their impact on CAD and HF. The SWAT-CNN algorithm identified SNPs in exonic and intronic regions, providing phenotype burden scores. We used these scores to pinpoint the shared burden between HF and CAD. Additionally, we employed a Graph Convolutional Neural Network to identify Clonal Hierarchy Reconstruction, adjusting the clone size with a 2-40% variant allele frequency (VAF) range. After identifying the clone construction pattern, we correlated VAF and clone construction to calculate mutation burden. Identified CHIP mutations, mostly missense variants from our clone model, impacted the combined HF-CAD outcome utilizing PrimateAI. Our study included 50 randomly chosen patients each for CAD and HF from the UKBB and All of Us platforms. Our findings reveal that CHIP-induced pleiotropy incurs the

highest outcome burden at 2.93%, as outlined by our clone identification method. The cumulative effect of mixed common variants (both CHIP and non-CHIP) contributes to 1.19% of the outcome. Consequently, we conclude that the expanded range of CHIP variants enhances accuracy in identifying shared phenotypes, highlighting pleiotropy's impact on the outcome. In summary, our study shows the significant role of extended CHIP variants in identifying shared phenotypes and underscores pleiotropy's substantial impact on outcomes. Key findings include: 1. Known CHIP genes (DNMT3A, TET2, ASXL1, TP53) showed an outcome burden of 4.03%. 2. Extended CHIP genes (41 genes) exhibited a burden of 2.93%. 3. Common variants contributed to a 1.87% burden. 4. Rare variants had a minimal impact, with a 0.004% burden. This study emphasizes the importance of considering both common and rare CHIP variants in assessing the genetic burden of HF and CAD, revealing the complex interplay of genetic factors in cardiovascular disease. Our findings suggest that incorporating a broader range of CHIP variants can enhance the precision of genetic risk assessments and improve our understanding of the shared genetic architecture between HF and CAD.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1058T: Clinical Genome Resource Neurological Disorders Clinical Domain Working Group:Delineating gene/variant perplexities in common neurological disorders

Authors:

K. Latchman[1], C. Thaxton[2], M. Harms[3], C. M. Testa[4], S. Zuchner[5]; [1]Univ. of Miami Dept. of Human Genetics, Miami, FL, [2]Univ North Carolina, Chapel Hill, NC, [3]Columbia Univ., New York, NY, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract:

Clinical Genome Resource Neurological Disorders Clinical Domain Working Group:Delineating gene/variant perplexities in common neurological disorders
Latchman,K[1]; Thaxton, C[2]; Harms, M[4]; Testa, C[5]; Zuchner, S[6]; Clinical Genome Resource Neurological Disorder Clinical Domain Working Group[3]
KXL604@med.miami.edu
The Neurological Disorders Clinical Domain Working Group (CDWG), an international collaborative effort, was established in 2023 as part of the larger Clinical Genome Resource (ClinGen). The overarching goal is to validate gene-disease relationships and variant

pathogenicity using established guidelines formulated by ClinGen, American College of Medical Genetics and Genomics, and the American Association of Medical Pathology. The Neurological Disorders CDWG currently encompasses neuromuscular, cognitive, and movement disorders, with active expansion into other areas including cerebellar ataxias, dementias, and neurovascular disorders. There are a total of 10 active gene curation expert panels (GCEP) and variant curation expert panels (VCEP), focused on neurological diseases including amyotrophic lateral sclerosis (ALS), congenital myopathies, Charcot-Marie Tooth disease (CMT), muscular dystrophies and myopathies including Limb-girdle muscular dystrophy, and Parkinson's disease (PD). There is a critical need to determine the clinical validity of gene-disease relationships and variant pathogenicity of genes targeted for upcoming genetic therapies, and to advance the practice of precision medicine. The Food and Drug Administration recognizes the variant interpretation process established by ClinGen, which is highly indicative of the rigor and transparency with which ClinGen expert panels are held for variant adjudication. A major accomplishment has been the curation of more than 180 genes among the ten expert panels with 64 genes curated by the CMT GCEP, 37 genes by the congenital myopathies GCEP, 47 genes by the ALS GCEP, 30 genes by the muscular dystrophies and myopathies GCEP, and 7 genes by the PD GCEP. We will summarize the current efforts of these expert panels, as well as the Neurological Disorder CDWG's vision of areas of importance to target for future expert panel development. The main mission of the Neurological Disorders CDWG is to validate gene-disease relationships and variant pathogenicity to better inform clinical testing, reporting, and ultimately patient care. Participating clinicians, researchers, and biocurators directly contribute to this mission, including lasting impacts on shared understanding, data, knowledge and time to accelerate precision medicine in neurological disorders.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1059T: Comparative Analysis of Antihypertensive Medications Stimulating vs. Inhibiting Angiotensin II Type 2 and 4 Receptors in Relation to Type 2 Diabetes Incidence in VUMCSD Cohort

Authors:

A. Tiwari[1], R. Chen[1], Y. Tan[2], Y. Yan[1], X. Zhong[3], B. Li[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. - Nashville, TN, Nashville, TN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Background: Prevalent use of antihypertensive medications that stimulate type 2 and 4 angiotensin II receptors, compared with those that do not stimulate these receptors, has been associated with a lower risk of type 2 diabetes (T2D), but its role in the prevention of T2D itself is unclear.**Objective:** To investigate the relationship between the usage of antihypertensive drug regimens that stimulate vs inhibit type 2 and 4 angiotensin II receptors and type 2 diabetes in Vanderbilt University Medical Center (VUMC) and MarketScan participants.**Methods:** In this project, we use electronic health records (EHR) from VUMC (>3 million de-identified individuals). We constructed Type 2 and 4 angiotensin II-stimulating (n = 33610) cohort, Type 2 and 4 angiotensin II-inhibiting (n = 45411) and nonusers (n=98592) cohort. Each cohort was matched by age, gender, body mass index (BMI), and disease comorbidities using a propensity score. We then estimated the hazard ratios (HRs) and 95% CIs using stratified Cox proportional hazard models, adjusting for age at diagnosis, gender, BMI, and disease comorbidities. For these drug users' groups, the therapeutic effects are replicated in the Market Scan database.**Results:** The sample included 177,613 VUMC SD participant data (92,010 women [51.80%]; mean [SD] age, 58.2 [14.4] years). During a 5-year follow-up, we found that usage of Type 2 and 4 angiotensin II-stimulating drug users was associated with the reduced risk of T2D (HR 0.77; 95% CI 0.75 to 0.79; P= 1.4e−71) compared to the non-user control cohort. Type 2 and 4 angiotensin II-stimulating users were also significantly associated with a decreased risk of T2D in comparison with Type 2 and 4 angiotensin II-inhibiting drug users (HR 0.83; 95% CI 0.80 to 0.85; P =5.7e−30). To confirm the VUMC SD clinical validation findings we performed the external replication studies in the Market Scan database and after a 5-year follow-up we noticed that usage of Type 2 and 4 angiotensin II-stimulating drug users is associated with the reduced risk of T2D (HR 0.75; 95% CI 0.73 to 0.77; P= 3.2e−107) compared to the non-user control cohort. Type 2 and 4 angiotensin II-stimulating users are also significantly associated with a decreased risk of T2D in comparison with Type 2 and 4 angiotensin II-inhibiting drug users (HR 0.84; 95% CI 0.81 to 0.86; P =4.8e−30). The drug users' group had the therapeutic effects in the expected direction.**Conclusions:** This study using VUMC and MarketScan data suggests that the use of antihypertensive medications that stimulate type 2 and 4 angiotensin II receptors is associated with a lower risk of T2D compared with antihypertensive medications that inhibit these receptors.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1060T: Comparative analysis of structural variant calling: Short-read alignment-based vs. long-read assembly-based methods

Authors:

C-Y. Lee[1], P. Chen[1], P-M. Chien[1], P-L. Chen[1,2,3,4,5], J. Hsu[1]; [1]Graduate Inst. of Med. Genomics and Proteomics, Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [2]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [3]Genome and Systems Biology Degree Program, Academia Sinica and Natl. Taiwan Univ., Taipei, Taiwan, [4]Graduate Inst. of Clinical Med., Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [5]Div. of Endocrinology and Metabolism, Dept. of Internal Med., Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Abstract:

Structural variants (SVs) are genomic changes involving over 50 base pairs and often affecting protein function. This study compared the SV calling performance by the truth SV sets released from the Genome in a Bottle (GIAB) consortium (NIST_SV_v0.6). This SV callset includes 6,661 large deletions and 7,595 insertions located within the confident regions for HG002, the son in an Ashkenazim Trio. We utilized short-read and long-read whole-genome sequencing data to detect SVs and evaluate the performance of the SV callers. For short-read sequencing data, we employed six alignment-based SV detecting tools—Delly, Manta, MELT, SvABA, Wham, and DRAGEN—and integrated their results to improve accuracy. For long-read sequencing data, we employed the assembly-based SV caller SVIM-asm. In HG002, after merging the results of multiple short-read alignment-based tools, we found 3,575 deletions and 3,869 insertions. Using SVIM-asm, we identified 3,384 deletions and 5,317 insertions. We calculated the performance metrics for both SV calling strategies. The F1 scores for deletions and insertions from the combined short-read alignment-based SV callers were 0.83 and 0.63, respectively. For the long-read assembly-based method, the F1 scores for deletions and insertions were 0.84 and 0.83, respectively. Around 71% of deletions and 63% of insertions identified by the long-read assembly-based method were also detected by the short-read alignment-based methods. Conversely, 75% of deletions and 45% of insertions detected by the short-read alignment-based methods were also detected by SVIM-asm. In addition to HG002, which might introduce information bias by just one sample, we incorporated a diverse set of 44 samples from the Human Pangenome Reference Consortium (HPRC) to enhance SV accuracy assessments. One of these samples is HG00438. In this sample, 88% of deletions and 76% of insertions identified by SVIM-asm were also detected by short-read alignment-based methods. Conversely, 73% of deletions and 46% of insertions identified by short-read alignment-based methods were also detected by SVIM-asm. Our study suggests that the long-read assembly-based method consistently outperforms the short-read alignment-based methods. While both methods perform similarly in detecting deletions, the long-read assembly-based method excels at detecting large insertions. Including a diverse sample set further enhances the credibility and generalizability of our findings.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1061T: Construction of a Korean pan-genome based on long-read sequences

Authors:

J-Y. Kim, Y. Cho, S-J. Park, S. Lee, J. Kim, S-Y. Lee, J-L. Park, S-Y. Kim; Korea Res. Inst. of BioSci. and Biotechnology, Daejeon, Korea, Republic of

Abstract:

Recent advancements in genomic research have revealed that a traditional single reference genome cannot properly represent the genetic diversity of the human population. Using the single reference genome results in various reference biases that can adversely affect variant or gene discovery. These recognitions led to projects constructing human pan-genomes, consisting of the integrated genomes of multiple individuals in the form of a non-linear graph, and in 2023, a draft human pan-genome and a Chinese pan-genome were announced based on 47 multi-ethnic and 58 Chinese individuals, respectively. East Asian populations are distinguished from each other to some degree by the genetic diversity, but pan-genome construction studies have not yet been reported in depth, except for China. From this perspective, we present the first phase of Korean pan-genome constructed based on long-read sequences from 19 healthy Koreans. A total of three long read sequencers, Revio, Sequel2, and PromethIon, were used, and 9, 3, and 7 Koreans were sequenced, respectively. Their short reads were also sequenced by the MGI-T7 sequencer. Long read-based genome assemblies were performed for each sample, followed by rigorous cleaning processes including short read-based genome polishing. A total of 31 high-quality haplotype-phased assemblies were then obtained, with an average total genome size of 2.96 Gb, an average contiguity N50 of 24.42 Mb, and an average of long-read sequence coverage of 24.57X. Using the 31 high-quality assemblies, the Korean pan-genome was constructed and released on our website. Additionally, the Korean pan-genome was compared with the draft human and Chinese pan-genomes, and *ALDH2* gene, which is related to alcohol consumption, was identified as one of Korean-specific genes with structural variations. Our research to construct the Korean pan-genome reference and identify Korean-specific structural variation is currently in progress. Therefore, this time, the initial results and methodology are released, and the final results will be published later. Our Korean pan-genome is expected to be able to capture more genetic variations in Koreans than the existing human reference, and contributes to the global effort of constructing a comprehensive human pan-genome.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1062T: Copy number variant associations with plasma protein levels in the UKB Biobank ⭐

Authors:

X. Zou[1], F. Hu[1], H. Lou[2], O. Burren[1], X. Li[3], S. Atanur[1], S. Deevi[1], AstraZeneca Genomics Initiative, S. MacArthur[1], S. Wasilewski[1], S. O'Dell[1], G. del Angel[4], L. Tian[2], M. Fabre[1], K. Smith[1], S. Petrovski[1], Q. Wang[5], K. Carss[1]; [1]AstraZeneca, Cambridge, United Kingdom, [2]AstraZeneca, Shanghai, China, [3]AstraZeneca, Gaithersburg, MD, [4]AstraZeneca, Boston, MA, [5]AstraZeneca, Waltham, MA

Abstract:

Copy Number Variants (CNVs), which are duplications or deletions of DNA segments, can considerably affect gene expression and disease development due to changes in gene dosage and indirectly through regulatory regions. Indeed, recent large-scale studies have linked numerous CNVs to genetic disorders. However, the phenotypic heterogeneity associated with CNVs underscores the need for multi-omics approaches to decipher the complex network of molecular interactions responsible for their effects. In our study, we investigated the impact of CNVs on plasma protein levels by using CNVs called from 50K UK Biobank (UKB) whole-genome sequences (WGS) and 2,923 plasma protein abundances measured by The UKB Pharma Plasma Proteome (UKB-PPP) consortium. We conducted CNV variant- and gene-level association studies to pinpoint both common and rare CNV-pQTLs. We identified 183 significant (p<1E-8) CNV-protein variant-level associations for rare CNVs (site frequency <1%) and 175 for common CNVs (site frequency >=1%). Notably, 94% cis-pQTLs associated with decreased protein levels are deletions and 82.5% with increased protein levels are whole-gene copy gain duplications, aligning with the expected functional consequences. Through gene-level collapsing analysis, we identified 65 gene-protein associations, which included 9 pleiotropic loci. For example, a loss of *MSR1* increased levels of the proteins encoded by *C1QTNF5* (*P*=2.61E-160, *beta*=1.69) and *LGALS3BP* (*P*=1.13E-186, *beta*=1.67), and a copy number gain of *MSR1* reduced the levels (C1QTNF5:*P*=1.01E-11, *beta*=-1.58; LGALS3BP*: P*=5.19E-9, *beta*=-1.37). Finally, we augmented gene-level tests by integrating CNV deletions with small protein-truncating variants (PTVs) to enhance our ability to identify novel gene-protein associations. This not only amplified the true signals, but also facilitated the discovery of novel associations. For instance, the deletion of *IFI35*, which encodes the interferon-induced protein 35, was significantly associated, in trans, with a reduced protein level of NMYC interactor (NMI) in only the CNV-PTV combined analysis (P_{CNV_PTV}=8.64E-12, P_{CNV}=1.01E-6,P_{PTV}=6.73E-7). Our findings offer valuable insights into the

effects of CNVs on the plasma proteome, and highlight the importance of integrating multiple variant types in collapsing analysis to better elucidate the genetic architecture of complex human traits.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1063T: CRISPR CuT Visualizer (CCTV) Discovers High Efficiency Designs for Cas9 Multi-Guide Mutagenesis Experiments

Authors:

C. Raut, R. Hussain, A. Irshaid, P. Ponnandy, E. Speliotes; Univ. of Michigan, Ann Arbor, MI

Abstract:

Background: CRISPR (clustered regularly interspaced short palindrome repeats)-associated nuclease Cas9 is used to edit DNA to study effects of genomic deletion in cells. Multi-guide mutagenesis (MGM) was proposed to improve efficiency of generating directed deletions. CRISPR MGM design tools do not presently account for factors like protospacer adjacent motif (PAM) orientation or frameshifts that may improve efficiency (PMID: 34593942). Here we introduce CRISPR CuT Visualizer (CCTV): a software tool to aid in visualizing and picking guides to improve the efficiency of MGM.
Methods: CCTV determines deletion sizes across canonical & non-canonical isoforms of a selected gene and reports PAM orientations and guide efficiencies. We used reference sequences from GenBank (PMID: 23193287). Guide on/off target efficiencies were pulled from CRISPick. We first used multi-guide sgRNA designed by a commercial source with sets of 3 guides for 11 preselected genes. We evaluated guides produced by this protocol for frameshifts and PAM orientations using CCTV. We transfected Huh-7 cells via the Cas9-sgRNA complexes and extracted DNA using QuickExtract DNA Extraction. We performed Sanger Sequencing and obtained deletion proportions using the ICE software. We identified the deletion efficiency across the different 3-guide PAM orientations. For 2/11 originally targeted genes and two new genes we generated new 3-guide designs according to the higher efficiency PAM orientation (hePo). We used the same transfection and sequencing protocols to obtain deletion proportions. Our final statistical comparison compared deletion proportions between 7 genes with hePo designs and 8 genes with non-hePo MGM designs using permutation testing.
Results: We found 9/11 (~81.8%) commercial MGM designs did not generate frameshift despite generating DNA cuts. Only 3/11 (~27.2%) genes contained guides in the hePo configuration and the two most efficient deletions (>90%) were in the hePo configuration. After designing hePo

guides for an additional 4 genes we found that average deletion efficiency was higher for hePo guides (16.1% non-hePo vs 56.5% hePo, p=0.026).
Conclusions: Current MGM design tools do not optimize for generating exonic frameshift mutations nor do they take into account PAM orientation which we show increases deletion efficiency. Using CCTV, we identified the hePo design leading to a 3.5-fold increase in average deletion efficiency compared to non-hePo MGM designs in human cell-lines. CCTV will be made available at https://github.com/Speliotes-Lab/CCTV.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1065T: deciphEHR: Towards Genomic Medicine at NYU Langone Health

Authors:

K. Ayers[1], J. McCafferty[1], E. Iturrate[2], A. Chakravarti[2], O. Yaacov[2]; [1]NYU Langone, New York, NY, [2]NYU Grossman Sch. of Med., New York, NY

Abstract:

Human genome sequence variation provides valuable insights into disease susceptibility and etiology, enabling personalized treatment strategies. Today, biobanks are driving gene discovery and increasing the potential for such individualized genomic medicine. However, translation of these findings into clinical decision-making is still challenging. deciphEHR is a NYU Langone Health (NYULH) system-wide initiative for learning how best to implement genomic medicine. We use opportunistic patient ascertainment for linking their genome sequence with their electronic health record. The goal is to use these resources for gene discovery, translational science and refine implementation strategies. Its unique design allows for the integration of real-time updates from electronic health records (EHR), including longitudinal and sophisticated phenotypes and valuable insights from EHR free text. From consent, through sample collection and sequencing to computational processing and analysis, our framework was designed with scalability in mind, scaling up from 10,000 participants sequenced in the pilot year, to up to 100,000 participants per year. Participants are consented remotely or in person. Leftover blood ready for disposal is used to blot dried blood spot cards (DBS) on filter paper, which are stored at room temperature. A high-throughput, fully automated, whole genome sequencing (WGS) protocol incorporates robotic library preparation directly from DBS hole punches, skipping the DNA isolation step. For testing scaleup at low costs, we currently perform low-coverage WGS (lcWGS) at 1.5X followed by imputation but intend to transition to high-coverage WGS.

Sequence files are processed uniformly using a custom pipeline. Missing genotypes are imputed with GLIMPSE2, validated in-house for non-reference variant accuracy. Our data storage design incorporates an interface for efficiently querying the data (genomic, de-identified EHR phenotypes, annotations) using SQL syntax. Genetic ancestry, both global and local, is estimated using the allele frequencies of global populations from the publicly available HGDP+1000G resources. Estimated ancestry is highly correlated with self-reported race/ethnicity. Our work demonstrates the feasibility of utilizing existing institutional infrastructures for genomic medicine. We have successfully established an automated process for generating high-quality lcWGS from DBS, accurately impute genotypes from DBS lcWGS and estimate genetic ancestry.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1067T: Developing a natural history resource to inform decisions about when to screen for rare monogenic conditions

Authors:

J. Jackson[1], A. Miranda[2], K. Kennedy[1], J. Berg[1]; [1]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [2]Clemson Univ., Clemson, SC

Abstract:

Background:
Understanding natural history and age of onset for rare monogenic conditions is vital for decision-making about screening. We examined the literature for comprehensive and reliable data and methods for determining onset age and found a significant gap in the availability of robust, high quality natural history data for many conditions.
Methods:
We developed a literature search strategy to identify studies reporting age of onset or natural history for monogenic conditions. We extracted study design characteristics, sample sizes, methodologies, patient ascertainment, and diagnosis criteria. Manuscripts were graded for quality, accounting for potential biases: Grade A (individual case-level data from longitudinal studies), Grade B (retrospective individual case-level data from clinical records), Grade C (aggregate-level data from clinical records/population studies), and Grade D (expert opinions, gene reviews, review articles).
Results:
We piloted this work on 10 conditions considered for screening during childhood. A total of ~4,300 papers were screened for age of onset information, resulting in 49 papers representing

29,398 patients. Of these, 16 papers (26.5%) provided individual case-level data, 30 papers (67.3%) provided aggregate-level data, and 3 papers (6.2%) were based on expert opinion. Only one condition had Grade A evidence with longitudinal data and cumulative incidence curves. Seven conditions had Grade B case-level data, and four conditions had Grade C data for aggregate age of onset data. A definition for age of onset (or age of first symptom manifestation) was specified in only 9/49 (18.4%) papers; the remaining 40 (81.6%) either did not provide details about how age of onset was defined, or provided ambiguous information.
Conclusions:
Natural history data for monogenic conditions is lacking and often represented in aggregate (mean or median, often without measures of error), limiting usefulness for understanding the temporal nature of symptom manifestations. Our systematic approach to curating natural history highlights current gaps and challenges in the literature including incomplete reporting of age of onset, variability in study designs, and different sources of bias. The goal to determine optimal times for genomic screening relies on a well-curated dataset with detailed symptom onset information. Future efforts should enhance the quality and comprehensiveness of natural history studies for rare monogenic conditions.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1068T: Establishing a Community Advisory Board for a Genomic Medicine Research Program in the Somali Community of Southeast Minnesota

Authors:

M. Pasha[1], H. Bangash[2], O. Ogunsan[2], J. Petrzelka[3], A. A. Mohamed[2], I. Kullo[2]; [1]Dept. of Gen. Internal Med., Hennepin Hlth., Minneapolis, MN, [2]Dept. of Cardiovascular Med., Mayo Clinic, Rochester, MN, [3]Dept. of Cardiovascular Med., Mayo Clinic, Roch, MN

Abstract:

Purpose: The study of African genomes is an urgent need not only to promote equity but also as a scientific imperative. We established a Community Advisory Board (CAB) to provide input on a community-based participatory genomic medicine research project which includes creation of a Biobank, in the Somali community of Southeast (SE) Minnesota. We describe key themes that emerged from analysis of the initial 4 CAB meeting recordings. **Methods**: We established a Somali CAB as part of the governance structure of the Biobank to seek feedback regarding research priorities and expectations

related to genomic medicine, informed consent, data sharing, and return of actionable results. The CAB is composed of diverse members of the community and includes elders, women, youth, and community leaders. An introductory meeting was held during which the study team gave an in-depth presentation on biobanking and its benefits and risks and the need to increase representation of individuals of African ancestry in genomic studies to lessen disparities in genomic medicine. To date, a total of 4 CAB meetings have been held. Each meeting included ~10-15 CAB members, the study principal investigator and 4-5 members of the research team. Meetings were audio/video recorded and an inductive approach was used for analysis. **Results:** The CAB helped identify genomic medicine research questions that were perceived as being a priority to the community and gave feedback on the development of study materials including the informed consent document and education materials. Six key themes were identified from analysis of meeting recordings, including the need to i) promote awareness in the Somali community prior to conducting the research study - "*we're a community that goes with word of mouth…when we are starting this [study], we need one voice…it takes one person they trust to keep going.*"; ii) provide a comprehensive overview of the study (risks and benefits) and all study materials to the community; iii) address the needs, health priorities and disparities identified by the community; iv) address any mistrust that the community has towards the medical and research communities - "*mistrust is the biggest key to preventing not only participation in research but also getting treated or getting evaluated for symptoms that we are experiencing…there is a history behind that…*"; and v) ensure ownership of the research study belongs to the Somali community. **Conclusion:** We established a Somali CAB to provide input on a genomic medicine research study which includes creation of a Biobank of Somali individuals. Key themes identified will guide the research study and ensure a collaborative partnership with the community.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1069T: Evaluating Large Language Models on Medical, Lay Language, and Self-Reported Descriptions of Genetic Conditions

Authors:

K. Flaharty, P. Hu, S. Ledgister Hanchard, M. Ripper, D. Duong, R. Waikel, B. Solomon; Natl. Human Genome Res. Inst., Bethesda, MD

Abstract:

Large Language Models (LLMs) are generating interest in many medical settings. For example, LLMs can respond coherently to medical queries by providing plausible differential diagnoses based on clinical notes. However, there are many questions to explore, such as evaluating differences between open and closed-source LLMs as well as LLM performance on queries from both medical and non-medical users. In this study, we assessed multiple LLMs, including Llama-2-chat, Vicuna, Llamamed2, Bard/Gemini, Claude, ChatGPT3.5, and ChatGPT-4, as well as non-LLM approaches (Google search and Phenomizer) regarding their ability to identify genetic conditions from textbook-like clinician prompts and their corresponding layperson translations related to 63 genetic conditions. For open-source LLMs, larger models were more accurate than smaller LLMs: 7b, 13b and larger than 33b parameter models obtained accuracy ranges from 22-49%, 41-51%, and 56-68%, respectively. Closed-source LLMs outperformed open-source LLMs, with ChatGPT-4 performing best (85-87%). 3 of 10 LLMs and Google search had significant performance gaps between clinician and layperson questions. We also evaluated how in-context prompting and keyword removal affected open-source LLM performance. Models were provided two types of in-context prompts, list-type prompts, which improved LLM performance, and definition-type prompts, which did not. We further analyzed removal of rare terms from descriptions, which decreased accuracy for 5 of 7 evaluated LLMs. Finally, we observed much lower performance with real patient descriptions; LLMs answered these questions with a maximum 21% accuracy.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1070T: *De novo* assemblies and variant calls for 625 samples from the Million Veteran Program cohort using HiFi reads

Authors:

N. Duro[1], P. Schreiner[2], Y. Shi[3], I. Said[4], B. Gorman[2], S. Pyarajan[5]; [1]Booz Allen Hamilton, Tampa, FL, [2]Booz Allen Hamilton, Lexington, MA, [3]VA Boston Hlth.care System, Jamaica Plain, MA, [4]Booz Allen Hamilton, New York, FL, [5]VA, Jamaica Plain, MA

Abstract:

De novo genome assemblies can improve reference genomes by filling in gaps, correcting errors in the current reference, and providing a more accurate characterization of difficult to assemble regions like centromeres and telomeres, leading to more accurate mapping and variant calling. Higher quality mapping and variant calling enable robust discovery of variants associated with rare diseases, especially those caused by complex structural variants or mutations in non-coding

regions. Here, we present 625 *de novo* assemblies constructed from PacBio HiFi reads in samples taken from the Million Veteran Program (MVP) cohort, coupled with variant calling utilizing DeepVariant and PBSV.Assemblies were generated using Hifiasm and assembly quality was evaluated with BUSCO, QUAST, and CMOPASS. The sequenced samples have on average a unique molecular yield of 235.8 Gb, 5.7 million reads with a score higher than Q20 (mean QV = 33.7). Reads had an average length of ~12 Kb with 24.1x depth of coverage and a N50 value of ~17 Kb. On average, assemblies have a total length of 3.07 Gb constructed from 964 contigs. The average L50 value is 45 (how many of the top longest contigs can construct 50% of the genome) and the average NG50 value is 40 Mb (essentially the shortest contig from the 45 that can construct 50% of the genome). In parallel, we employed DeepVariant, PBSV, hificnv, TRGT, and cpg-tools for comprehensive variant discovery. Deepvariant is a machine learning algorithm that was utilized to identify SNPs and small indels, while PBSV is specifically tailored to call large structural variants from PacBio HiFi reads. To validate our variant calling results we benchmarked HG002 variant calls against the Genome in a Bottle call set and achieved an F1 score of 99.96. The diversity of the MVP population along with the scale of this project offers the foundations for the construction of pangenomes and ancestry specific reference panels that capture population diversity better than any efforts currently available.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1071T: Exploring Gut-Brain Axis in Alzheimer's Disease using Computational Host-Microbiome Interaction Analyses

Authors:

S. Soorej[1], A. Anilal[2], T. Abdulrehman[3], A. Titus[4], S. Khader[5], K. Yadav[6]; [1]Acton-Boxborough Regional Highschool, Acton, MA, [2]Univ. of Maryland, Baltimore, MD, [3]eHlth.Program, DeGroote Sch. of Buisness, McMaster Univ., Hamilton, ON, Canada, [4]Dept. of Preventive Cardiology, Houston Methodist, Houston, TX, [5]Sch. of Publ. Hlth., Faculty of Med., Imperial Coll. London, South Kensington, London, United Kingdom, [6]Texas A&M Univ., Sch. of Engineering Med., Houston, TX

Abstract:

Alzheimer’s Disease is a neurodegenerative condition affecting individuals aged 65 and older, currently impacting an estimated 6.2 million people and projected to rise to 13.8 million by 2060. The pathogenesis of this disease is linked to neurofibrillary tangles and high levels of Abeta aggregation, which disrupt neuronal synapses. However, recent studies highlight that the

gut microbiome could play a significant role in Alzheimer's progression; patients often exhibit gut dysbiosis that may compromise the intestinal barrier and the central nervous system through the blood brain barrier. In this study, we explored the association between the microbial signatures and Alzheimer's Disease through meta-analysis and cross-species comparison. In AD patients, there are notably many elevated bacterial species, with Shigella being among them. Using a gene list of early-onset AD from DisGeNET, we conducted KEGG human pathway enrichment analysis on the Enrichr database to identify shared pathways with AD. Shigella also appears to participate in one of these shared pathways listed in the enrichment analysis: the shigellosis pathway, which encompasses several genes, including TLR4. Based on cross-species pathway comparison, bacterial species in Shigella could be initiating the TLR4 signaling pathway and producing proinflammatory cytokines, ultimately leading to neuroinflammation. Highlighting the link between elevated microbial species in AD and the involved pathways could reveal the potential of antibiotics to target these genes and pathways. The main goal is to synergize a few antibiotics to target a broad range of bacteria involved in these pathways, improving efficacy while causing less disruption to beneficial bacteria in the body.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1072T: Genome wide association studies of 1,728 traits in a multi-ancestry cohort from the Michigan Genomics Initiative ⭐

Authors:

E. Bertucci-Richter, B. Vanderwerff, S. Patil, M. Boehnke, X. Zhou, S. Zollner; Univ. of Michigan, Ann Arbor, MI

Abstract:

As GWAS findings are increasingly used to inform clinical practice, major healthcare inequities are created when the leveraged datasets include primarily individuals of European ancestry. Although efforts to increase diversity are becoming common in the recruitment of large studies, methods for efficiently analyzing multi-ancestry cohorts are not well established. Here, we tested one approach of analyzing a multi-ancestry cohort from the Michigan Genomics Initiative (MGI) by conducting 1,728 GWAS for Phecode phenotypes spanning 17 broad categories. The MGI contains >80,000 genotyped participants, of which, 14% are of non-European ancestry. To avoid the confounding effects of population stratification, previous large-scale analyses in MGI have been focused exclusively on the European cohort, excluding participants of African (6% of

cohort), West Asian (3%), East Asian (2%), Central/Southern Asian (2%), and Native American ancestry (1%). In an attempt to efficiently analyze the entire multi-ancestry cohort, we used SAIGE to run logistic mixed models with saddle point approximation and corrected for relatedness, age, sex, genotyping array, and the first 20 genetic principal components. We report results for >52 million TOPMed imputed genetic variants and assessed replication of top hits using the Phenotype Genotype Reference Map (PGRM). We identified 1,516 independent top hits across 766 Phecode phenotypes. The mean (±SD) genomic control factor across all traits was 1.011±0.013, indicating population stratification was well controlled. Compared to previously reported associations from the European and all-ancestry PGRM GWAS Catalogs, we were well-powered (power>80%) to replicate 1,509 and 2,157 findings, respectively. Of these, we successfully replicated 70% of European associations and 62% of all-ancestry associations. In addition, we identified multiple novel associations including rs7190835 in *MBTPS1* associated with bone cancer and rs3889536 near *EDEM1* associated with amblyopia. Overall, we demonstrate that a mega-analysis approach to analyzing the multi-ancestry MGI cohort adequately controlled for population stratification, successfully replicated associations previously identified in single ancestry GWAS, and detected new associations. This approach is more streamlined than ancestry stratified analyses and subsequent meta-analysis and retained admixed individuals that might have been excluded in ancestry stratified analyses. While additional work is required to develop best practices for conducting multi-ancestry GWAS, we demonstrate one approach which is feasible to conduct at the phenome-wide scale.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1073T: Hybrid Cloud and Local Data Lifecycle Management at a Medium-Scale Genomics Center

Authors:

S. Griffith, B. Myers, A. Sanchez, M. Mawhinney, J. Paschall, B. Marosy, D. Mohr, E. Hsu, K. Doheny; Johns Hopkins Univ., Baltimore, MD

Abstract:

As preferred genomics technologies evolve and new technologies emerge, continual pressure is placed upon local data storage capacity at genomics centers. The problem of scaling storage capacity has been particularly acute for facilities with substantial sequencing work, with pressure both from a large recent increase in the rate at which short-read sequencers produce data and from the increased need to perform long-read sequencing. Aside from data capacity,

data centers must also be increasingly concerned with data recovery capabilities—all organizations are wise to have plans to recover from data loss, whether it is caused by some local disaster or from a cyber-attack, such as a ransomware incursion. The naive solutions to the dual pressures of capacity and recovery are readily apparent and alarmingly expensive: increase storage capacity and replicate all data to a secondary location. Here we discuss an alternative approach employed at one medium-scale academic genomics center. This center formed around the confluence of a broad array of genomics work being completed across its parent institution, including ongoing research projects at multiple scales, genomics core facilities, and diagnostic testing for germline and cancer genomics. The groups that constitute this center share laboratory space and IT infrastructure, including local data storage. This alternative approach employs HIPAA-compliant archival cloud storage and clearly defined data lifecycles to relieve pressure on local data storage and provide a clear path to recovery from data loss. Newly generated data, and the downstream data from active research projects and clinical tests are kept on local storage and selectively replicated to the cloud, while non-current data, such as completed research projects, are moved from local storage into the cloud, to be retrieved as needed. Data lifecycle requirements are empirically defined via close collaboration between laboratory staff, software developers, and bioinformaticians. Implementation of this hybrid storage strategy has focused on automation of data transit between local and cloud storage, with the creation of a flexible software layer to connect cloud and local infrastructure. This hybrid strategy has allowed this genomics center to increase data generation rates by 400% since June 2023, simultaneously creating a path to data recovery, while increasing data storage costs by less than 20%.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1074T: Identifying Somatic Mutations Utilizing GPU Accelerated DeepSomatic

Authors:

P. Vats[1], B. Zamirai[1], T. Zhu[1], H. Clifford[1], G. VACEK[2]; [1]NVIDIA, Santa Clara, CA, [2]NVIDIA, Duluth, MN

Abstract:

The field of human genetics has been revolutionized by Next Generation Sequencing (NGS), which is now essential for research and clinical applications, especially in cancer diagnosis and treatment. NGS technologies are crucial for identifying and characterizing somatic and germline

genomic alterations from Whole Genome Sequencing (WGS) data, providing critical insights into the genetic underpinnings of cancer, guiding personalized treatment strategies and improving patient outcomes. As genomics advances, optimizing somatic variant calling processes to enhance the speed and accuracy of genomic analyses is crucial.To address these challenges, we employed NVIDIA Clara Parabricks, a GPU-accelerated framework. This workflow includes GPU-accelerated BWA-MEM for alignment and DeepSomatic for somatic variant calling. GPUs enable parallel data processing, significantly reducing computational burden without compromising accuracy. The accelerated DeepSomatic algorithm was rigorously evaluated and benchmarked using three publicly available WGS datasets: SEQC2, COLO0829, and Lineage derived Somatic Truth (LinST).We benchmarked both baseline and accelerated DeepSomatic algorithms on these datasets. For the SEQC2 dataset (~48x WGS coverage), the accelerated version ran in 217 minutes and 56 seconds, compared to 1151 minutes and 10 seconds for the baseline. On the LinST dataset (~37x WGS coverage), the accelerated algorithm completed in 112 minutes and 23 seconds, while the baseline took 543 minutes and 48 seconds. For the COLO0829 dataset (~40x coverage), the accelerated version ran in 151 minutes and 39 seconds, versus 823 minutes and 36 seconds for the baseline. All computations were conducted on AWS cloud instances, using g4dn.12xlarge (4 Tesla T4 GPUs) for the accelerated version and m5.24xlarge (96 vCPUs) for the baseline.Achieving identical results to the baseline while significantly reducing processing time and lowering analysis costs has profound implications for clinical and research applications. In a clinical setting, faster processing times can lead to more timely diagnosis and treatment decisions, improving patient outcomes. For researchers, the ability to process large volumes of genomic data more efficiently can accelerate the pace of discovery and innovation in genomics. The integration of GPU acceleration with advanced deep learning algorithms represents a promising direction for the future of genomics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1075T: Investigation of Genetic Determinants of Cystatin C serum levels in the Korean Cohort

Authors:

J. Han, H. Jo, H. Kang, T. Jung, S. Jang, K. Lee; Chungbuk Natl. Univ., Cheongju, Korea, Republic of

Abstract:

Background: Cystatin C has potential as a biomarker as it has been reported to be associated with various diseases such as chronic kidney disease, thyroid disease, and cardiovascular

disease. This study aims to investigate single-nucleotide polymorphisms (SNPs) associated with cystatin C serum levels in the Korea cohort.Methods: Genome-wide association Study (GWAS) analyses were performed on the two Korean population-based cohorts, the health examinees study (HEXA) and the cardiovascular disease association study (CAVAS) of the Korean Genome and Epidemiology Study (KOGES). The GWAS studies of the HEXA and CAVAS cohorts were combined into a meta-analysis to perform lead SNP identification, gene-based analysis, gene-set analysis, and gene-gene interaction analysis. Lead SNP information from the combined cohort was applied to individuals in the HEXA cohort to calculate a polygenic risk score (PRS). Result: Based on the lead SNPs identified in the meta-analysis, rs2390793(LRP2, BETA: -0.0096, SE: 0.0015, P: 3.13E-10), rs2897119(intergenic, BETA: -0.0393, SE: 0.0022, P: 2.89E-73), and rs117165099(CST4, BETA: -0.0401, SE: 0.0045, P: 2.36E-19) were validated as suggestive line (P: 1E-05) in the discovery and replication stage HEXA, CAVAS cohort. A 1 SD increase in standardized PRS was associated with a 1.21 odds ratio (95% CI: 1.17-1.26) for chronic kidney disease (CKD). Conclusions: This study may provide a genetic basis for SNPs and related genes of cystatin C levels in the Korean population.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1076T: Landscape of 25,000 Korean whole-genomes in the Korea National Genome Project

Authors:

S. Kim[1], T. Hwang[1], D. Jang[1], Y. jeong[1], J-h. park[1], S-Y. Kim[1], **B. Lee**[2], S. Hwang[1], H. jeong[1]; [1]Korea Res. Inst. of BioSci. & Biotechnology, Daejeon, Korea, Republic of, [2]Korea Bioinformation Ctr., Daejeon, Korea, Republic of

Abstract:

Recent advance in whole-genome sequencing technologies enables a genome-wide exploration of a large cohort, providing compelling insights into the genetic features associated with a variety of phenotypes as well as the disease predisposition of the participants. Here, we produced whole-genome sequences of about 25,000 individuals from rare disease, alzheimer's disease, autism spectrum disorder, colon/lung cancer, and healthy cohorts in the Korea National Genome Project (NGP). By conducting a pilot study with the qualified whole genome sequences of autism spectrum disorder, and two healthy cohorts, we identified a total of approximately 10 million variants from all the individuals, consisting of 8.2 million SNVs and 1.7 million INDELs. A comparative genome analysis between the disease and normal/healthy individuals suggested a

subset of the variants as disease-associated. Next, we compared the genome profiles between other ethnic groups by using publicly accessible datasets, thereby elucidating distinct genetic characteristics of Korean or East Asian population from other populations. We believe that all these results not only provide us a comprehensive understanding of genetic architecture of Korean population, but also facilitate precision medicine studies for the population.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1077T: Large scale characterization of disease expressivity in carriers of inherited cardiovascular disease pathogenic variants in the UK Biobank.

Authors:

M. Ramaker[1], K. M. Corey[1], J. W. Abdulrahim[1], L. C. Kwee[1], K. Kottilil[1], S. Coles[1], N. Nafissi[1], K. Amos[1], S. H. Shah[2]; [1]Duke Univ., Durham, NC, [2]Duke Univ, Durham, NC

Abstract:

The goal of clinical genetic sequencing is to identify pathogenic, disease-causing variants to confirm diagnosis, support risk assessment and inform clinical decision making; however, most reported pathogenic variants are not well described. Increasing sizes of genetic-cohort studies coupled with extensive electronic health record (EHR) data have made it more feasible to characterize disease expressivity in carriers of pathogenic variants. Using whole exome sequencing (WES) and EHR data available in the UK Biobank (UKB) for >500k individuals, we identified disease expressivity in carriers of likely pathogenic (LP) and pathogenic (P) variants (as defined by of American College of Medical Genetics and the Association for Molecular Pathology) in 47 genes associated with inherited cardiovascular disease (CVDs). We identified 697 P/LPs present in 37 of these genes in UKB. We show that 3,722 participants (0.74%) carried one of these P/LP variants, 15 of whom carried two different P/LP variants. We determined disease expressivity in P/LP carriers by integrating extensive EHR and imaging data available for the following conditions: cardiomyopathies (hypertrophic and dilated), electrophysiology disorders (long QT syndrome, Brugada syndrome, arrhythmogenic right ventricular cardiomyopathy), connective tissue disorders (Ehlers Danlos syndrome, Marfan's syndrome, Loeys-Dietz syndrome, familial thoracic aortic aneurysm and dissection), hereditary transthyretin amyloidosis, and familial hypercholesterolemia. Through this, we show that of UKB participants carrying a P/LP variant in one of these CVD genes, 1,426 (3.83%) carry diagnostic codes and/or symptoms consistent with the corresponding disease. We highlight that disease

expressivity was heterogeneous, even in genes thought to be highly penetrant. LDLR had the highest percentage of P/LP carriers expressing the corresponding phenotype, familial hypercholesterolemia-related traits (n=390, 73%), followed by FBN1 P/LP carriers expressing Marfan's syndrome-related traits (n=49, 68.1%,). Conversely, for several genes, no P/LP carriers expressed disease-related traits (DSC2 [n=13], KCNE1 [n=4], KCNE2 [n=1], MYH11 [n=4], TPM1 [n=7]). In sum, this work helps better elucidate disease penetrance in inherited CVDs, can serve as a broader resource for the genomics community and can support clinical decision making.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1078T: Leveraging Explainable AI for prioritizing genomic variants across monogenic, digenic, and oligogenic frameworks

Authors:

F. De Paoli[1], S. Berardelli[2,1], E. Rizzo[1], **I. Limongelli**[1], S. Zucca[1]; [1]enGenome srl, Pavia, Italy, [2]Univ. of Pavia, Pavia, Italy

Abstract:

Background
Identifying causative variants is a major challenge in genomics, crucial for improving the diagnostic yield of rare diseases. To address this, we've developed a suite of tools that leverages Explainable AI-driven algorithms to analyze genetic data, identify potential disease-causing variants, and provide actionable insights. By enhancing precision and efficiency in genomic analysis, our solution aims to accelerate diagnosis and increase the diagnostic yield by prioritizing both monogenic, digenic and oligogenic candidates with an hypothesis-driven approach.
Methods
Firstly, we developed the "Suggested Diagnosis" AI system that leverages pathogenicity, phenotypes, and familial inheritance to prioritize causative variants under a Mendelian hypothesis. A layer of Explainable-AI (XAI) elucidates the rationale behind results, overcoming the challenges of the "black box" concept that characterize the Machine Learning based approaches. Furthermore, we developed an HPO-based algorithm to determine how well each patient's phenotype is explained by the prioritized variant.An XAI-based layer identifies potential incidental findings (IF) to reduce the number of candidates that need to be reviewed by geneticists. Secondly, beyond monogenic analysis, we integrated an additional AI-based layer to identify digenic and oligogenic variants, leveraging also gene-gene interaction evidence.

Results

The "Suggested Diagnosis", awarded as "Best performing Genome Interpretation model" within the CAGI6 challenge (PMID:38520562, PMID:37577678), was benchmarked on 6 different cohorts of WES and WGS data with both causative SNV/INDEL and CNV. On the 4 WES datasets, our solution was able to identify in the 85.7% of cases the causative variants in the top 20 positions, while on WGS data in the 100% of the cases in the top 15. Furthermore, the IF detection approach reduces the number of variants to be reviewed by geneticists by 80%.The digenic module was validated on 11 solved samples (sensitivity 73%, median rank of the causative combination 3). The tool reaches a sensitivity of 81% on a larger manually curated dataset: the digenic predictions and their explainability are freely available on oliver.engenome.com (10.1101/2023.10.02.560464).

Conclusions

Based on extensive benchmarking and collaborative efforts, our tools emphasize transparency and trust in AI predictions. The monogenic and digenic modules are natively integrated in the eVai software (evai.engenome.com), providing advanced resources for variant interpretation.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1079T: Long read analysis predicts RNA binding proteins with isoform-specific functions ⭐

Authors:

S. Park[1], M. Schertzer[1], E. Jeffery[2], G. Sheynkman[2], D. Knowles[3]; [1]New York Genome Ctr., New York, NY, [2]Univ. of Virginia, Charlottesville, VA, [3]New York Genome Ctr. & Columbia Univ., New York, NY

Abstract:

Alternative splicing (AS) plays a critical role in gene regulation and protein expression by generating different messenger RNA (mRNA) transcripts from the same pre-mRNA. AS is regulated by trans-acting RNA binding proteins (RBPs), which are often alternatively spliced themselves. Because alternatively spliced RNAs often produce different protein products, AS of RBPs is very likely to affect the downstream functions of RBPs, but studies of isoform-specific RBP functions are limited. In this study, we develop both a computational pipeline to identify RBP isoforms and predict downstream functional effects and an experimental framework to validate these predictions. First, to obtain a set of high-confidence isoforms, we analyzed 96 PacBio long-read RNA-seq datasets from the ENCODE consortium using a customized FLAIR

pipeline. We obtained a total of 22,723 isoforms, including both novel and annotated transcripts, across 1,542 known RBPs with an average of 14.7 per RBP while transcription factors have an average of 9.8. Next, we predicted variability in isoform-specific function considering three features of each isoform: protein localization, RNA-binding, and protein-protein interactions. To determine if protein isoforms of the same RBP are localized differently within a cell, we used NucImportv2 to obtain import probabilities. To assess the potential for differential RNA binding, we calculated a similarity score by comparing tappAS-identified RNA binding domains to the corresponding domain in the ENSEMBL canonical isoform. To predict potential differences in protein-protein interactions, we used IUPRED2 to predict intrinsically disordered regions and compared these along with the terminal 10 amino acids of the sequence to the canonical isoforms. To summarize these we compiled our isoform-specific predictions into a master table, providing a detailed characterization of each RBP isoform to facilitate further downstream analysis. Finally, to validate these isoform-specific functional predictions, we are expressing selected FLAG-tagged isoforms in human cell lines. Using a FLAG antibody, we perform isoform-specific immunofluorescence, RNA-immunoprecipitation, and IP- mass spectrometry to investigate differences in protein localization, RNA binding, and co-factor binding, respectively. Overall, our method provides a reliable strategy to identify and mechanistically investigate isoform-specific functions of RBPs.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1080T: Long-read sequencing of hundreds of ancestrally diverse brains provides insight into the impact of structural variation on gene expression and methylation ⭐

Authors:

K. Billingsley[1], M. Meredith[2], K. Daida[1], P. Alvarez Jerez[1], R. Genner[3], M. Mastora[4], L. Malik[5], B. Baker[1], m. makarious[1], L. Paulin[6], P. Carnevali[7], K. Miga[2], A. Singleton[8], M. Nalls[9], W. Timp[10], A. Phillippy[11], M. Chaisson[12], M. Kolmogorov[1], M. Jain[7], F. Sedlazeck[13], B. Paten[2], C. Blauwendraat[1]; [1]NIH, Bethesda, MD, [2]UCSC, Santa Cruz, CA, [3]Johns Hopkins Univ. / NIH, Bethesda, MD, [4]UCSC, santa cruz, CA, [5]NIH, Bethesda, MD, [6]BCM, Houston, TX, [7]Northeastern, Boston, MA, [8]NIA, Bethesda, MD, [9]DTI, DC, DC, [10]Johns Hopkins Univ., Baltimore, MD, [11]NIH/NHGRI, Bethesda, MD, [12]Univ. of Southern California, Los Angeles, CA, [13]Baylor Coll. Med., Houston, TX

Abstract:

Structural variants drive gene expression in the human brain and are causative of many neurological conditions. However, most existing genetic studies have been based on short-read sequencing methods which poorly capture these regions of the genome. Long-read sequencing significantly increases our power to detect disease-associated and functionally relevant structural variants but has not been a feasible option for large-scale genomic projects. Here, we leverage a new scalable wet-lab protocol and computational pipeline for whole-genome Oxford Nanopore Technologies (ONT) sequencing and apply it to neurologically normal control samples from the North American Brain Expression Consortium (NABEC) (European ancestry) and Human Brain Collection Core (HBCC) (African admixed ancestry) cohorts. Through this work, we present a publicly available long-read resource from the frontal cortex of 361 human brain samples (average N50 ~30kb and 40X genome coverage). We discover over half a million structural variants, consisting mainly of insertions (n=376,787) and deletions (n=178,247). Utilizing matched expression datasets for these samples we apply quantitative trait locus (QTL) analyses and identify structural variants that impact gene expression in post-mortem frontal cortex brain tissue. Further, we determine haplotype-specific methylation rates of millions of CpGs and with this data identify cis-acting structural variants. In summary, these results highlight that long-read sequencing at population scale can identify disease-relevant and ancestry-specific regulatory loci that were inaccessible using previous technologies. We believe this new resource will provide a critical step toward understanding the biological effects of genetic variation in the human brain. Expanding on this, as part of the NIH's Center of Alzheimer's and Related Dementias (CARD) long-read sequencing initiative, we are currently applying this framework to sequence thousands of ancestrally diverse case and control brain samples to comprehensively catalog structural variants in Neurodegenerative diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1081T: Loss of Y chromosome associates with lung function decline in men with a history of smoking

Authors:

W. Saw[1,2], K. Kim[1,2], P. Auer[3], A. Reiner[4], L. Raffield[5], Y. Jakubek[6], E. Silverman[1,2], M. Cho[1,2], D. Demeo[1,2]; [1]Brigham and Women's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Med. Coll. OF WISCONSIN, Milwaukee, WI, [4]Fred Hutchinson Cancer Res. Ctr., Milwaukee, WI, [5]UNC - Chapel Hill, CHAPEL HILL, NC, [6]Univ. of Kentucky, Lexington, KY

Abstract:

Despite its known influence on health outcomes, the impact of loss of Y chromosome (LOY) on respiratory health remains under-investigated. We explored the association between LOY measured from whole genome sequencing and respiratory health leveraging data from 5,000 male participants (aged 40 - 85 years) from the COPDGene Study with a median follow-up of 6.5 years. Our assessment included lung function, chronic obstructive pulmonary disease (COPD) as categorized by GOLD grades (1 to 4), quantitative measures of emphysema, and multiomic associations with DNA methylation and proteomic signatures, all examined in relation to LOY status. In the COPDGene Study, individuals with LOY displayed significantly lower lung function along with elevated risks of COPD (OR = 1.27; 95% CI 1.06,1.54; p = 0.012) and severe COPD (OR = 1.23; 95% CI: 1.00, 1.51; p = 0.049). Notably, LOY also associated with higher quantitative measures of emphysema (β = 0.18; 95% CI: 0.08, 0.28; $p = 3.43 \times 10^{-04}$). Longitudinal analysis further revealed LOY's association with accelerated lung function decline (β = -0.13; 95% CI: 0.-0.21, -0.05; $p = 9.67 \times 10^{-04}$). To investigate the potential mechanisms of LOY, we identified widespread differential DNA methylation patterns associated with LOY, including 1420 differentially methylated regions (DMRs). The top three regions included *TTTY14* on chromosome Y, and regions on chromosomes 20 and 3, including the *ZMYND8* and *GRM2* genes. Significantly enriched pathways included the Hippo signaling and glycosaminoglycan biosynthesis - heparan sulfate pathways. Our proteomic analysis revealed reduced *CSF2RA*, increased *RRM1*, *BDNF*, and *CAMP* protein levels, further revealing potential mechanisms underlying LOY's influence on respiratory health. Overall, this study elucidates the multifaceted relationship between LOY and respiratory outcomes in males, providing critical insights into the epigenetic and proteomic mechanisms to explore as drivers of respiratory disease in men.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1082T: MetaSTAARlite: An all-in-one tool for biobank-scale whole-genome sequencing meta-analysis ⭐

Authors:

Y. Kumarasinghe[1], J. Williams[2], H. Zhang[3], Z. Li[4], X. Li[5]; [1]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [2]Natl. Cancer Inst., Rockville, MD, [3]Natl. Cancer Inst., Cambridge, MA, [4]Northeast Normal Univ., Changchun, China, [5]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Introduction

Biobank-scale whole-genome sequencing (WGS) studies have enabled the analysis of coding and noncoding variants contributing to complex traits or diseases. Data sharing restrictions often limit the power of genetic association analysis by restricting analysis to individual studies, making the development of appropriate meta-analysis frameworks a necessity. However, meta-analysis of rare variants (RVs) has remained a key challenge when applied to WGS data, largely because there is no scalable, resource-efficient and robust pipeline for comprehensive RV meta-analysis, which evaluates variant sets rather than single variants.

Methods

We propose MetaSTAARlite, an all-in-one meta-analysis pipeline for biobank-scale WGS studies that features resource-efficient storage of summary statistics, powerful and scalable association meta-analysis, and comprehensive result summary and visualization. First, MetaSTAARlite automatically generates highly storage-efficient summary statistics of common and rare variants using our proposed single variant summary statistics and sparse weighted LD matrices in MetaSTAAR. It uses functional annotations to provide default RV analysis units, including 5 coding and 7 noncoding categories for protein-coding genes, and 1 category for ncRNA genes, but also supports specification of custom analysis units. Second, MetaSTAARlite performs functionally informed RV meta-analysis, which boosts power by dynamically incorporating multiple variant functional annotations through MetaSTAAR. MetaSTAARlite accounts for population structure and sample relatedness for continuous and discrete traits. It also includes options for single variant meta-analysis and analytical follow-up via conditional analysis.

Results

We partitioned data from 190,110 participants in UK Biobank WGS data into three datasets containing 31,685, 63,370, and 95,055 participants. We then applied MetaSTAARlite to perform WGS RV meta-analysis across these datasets to discover associations with total cholesterol. In total, MetaSTAARlite required 2.37, 13.64, and 0.85 GB to store the summary statistics for coding, noncoding, and ncRNA analysis respectively. The results mirrored those of analyses of pooled UK Biobank data, identifying 58 and 89 significant coding and noncoding associations. Notably, storage requirements increase as a linear function of dataset sample size, highlighting the resource efficiency of MetaSTAARlite.

Summary

In summary, MetaSTAARlite provides an all-in-one powerful, flexible, and resource-efficient pipeline for rare variant meta-analysis of biobank-scale sequencing studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1083T: Mitochondrial genetics in the Taiwanese population: Insights from the Taiwan Biobank on mtDNA variations and their phenotypic implications

Authors:

T. Chou, P-M. Chien, P. Chen, P-L. Chen, J. Hsu; Natl. Taiwan Univ., Taipei, Taiwan

Abstract:

Mitochondrial DNA (mtDNA) variations play a significant role in the pathogenesis of various diseases due to their influence on cellular energy metabolism. However, the diversity in mitochondrial genetics among non-European populations remains insufficiently explored and requires more comprehensive genetic insights. This study leverages the extensive genomic data from the Taiwan Biobank (TWB), encompassing over 200,000 individuals, to explore mtDNA variations and their phenotypic associations. We reanalyzed whole-genome sequences from 1,484 participants by employing the GATK pipeline for mitochondrial SNP/Indel variant calling to detect homoplasmies and heteroplasmies. This analysis identified 2,361 mtDNA variants, including 77 novel variants not previously reported in the gnomAD v3.1 database. Our mitochondrial haplogroup analysis also underscored the predominance of haplogroups M, D, and F, demonstrating the distribution of Asian lineage mitochondrial haplogroups within the Taiwanese population.To further explore the potential implications of mtDNA variants on complex traits, we utilized TWB genotyping array data to conduct mitochondrial-wide association analyses with a wide range of phenotypes. First, we constructed a Taiwanese population-specific mtDNA imputation panel from our WGS results. We further assessed the feasibility of conducting mtDNA imputation and ensured the concordance of mtDNA variant profiles. Subsequently, we performed the association analyses from 120,163 participants to discover correlations between mtDNA variants and complex traits. We identified two variants in the MT-ND2 gene associated with high myopia-a leading cause of visual impairment and blindness. Additionally, we discovered 14 mtDNA variants correlated with renal function biomarkers. The renal function associated variants can be divided into two independent sets. The first set includes ancestral variants in haplogroups M and N, suggesting a risk correlation with renal dysfunction. The most significant variant is rs2853826 (m.10398A>G, p=7.7x10^(-5)). The other set contains variants in lineage B4b, which are linked to a protective association with renal function. The most significant variant is rs35134837 (m.16217T>C, p=1.1x10^(-5)). The results show that haplogroup background can modify the susceptible risk for complex traits.This study not only significantly enriches our understanding of mtDNA diversity in the Taiwanese population but also underscores the importance of including diverse genetic backgrounds in

mitochondrial research to uncover relevant genetic insights that may be masked in predominantly European studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1084T: MitoNet1.0: Bridging core and auxiliary mitochondrial processes using genetic interaction profiles

Authors:

N. Mikolajewicz[1], S. Masud[2], D. Thomson Taylor[3], K. Chan[1], K. Brown[1], O. Sizova[1], A. Habsid[1], R. Climie[3], M. Billman[4], C. Myers[5], B. Andrews[3], M. Costanzo[3], C. Boone[3], J. Moffat[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Hosp. for Sick Children, Toronto, ON, Canada, [3]Univ. of Toronto, Toronto, ON, Canada, [4]Univ. of Bonn, Bonn, Germany, [5]Univ. of Minnesota, Minneapolis, MN

Abstract:

Background: The human mitochondrion is a complex organelle involved in numerous functions beyond its indispensable role in aerobic respiration. This complexity has resulted in challenges in defining what exactly a mitochondrial gene is. Various strategies (e.g., protein-protein interactions, protein localization, bioinformatics) have resulted in complementary, yet discordant mitochondrial catalogs, including gene ontology (GO), MitoCoP, and MitoCarta3.0, among others. ***Methods:*** We performed >300 genome-wide CRIPSR/Cas9 screens using a diverse library of loss-of-function mutant HAP1 cells. Mitochondrial genes were classified using library-side co-similarities and machine-learning-derived latent features, and functionally annotated by pathway co-similarity analysis. Genes were further classified into “core” and “auxiliary” mitochondrial processes using the CoreMito20 score, which quantifies relative gene-level involvement in mitochondrial- and non-mitochondrial processes based on profile co-similarity profiles. ***Results:*** Here we present MitoNet1.0, a genetic interaction-based network of human mitochondrial genes. We identified 1743 known and novel mitochondrial genes and recovered known MitoCarta3.0 genes with an AUPRC of 0.663. Using the MitoPathway reference standard (MitoCarta3.0), we show that genetic interaction profiles were enriched for OXPHOS (Complex I-V; AUPRC = 0.49), mtDNA metabolism/maintenance (AUPRC = 0.42), and protein sorting/import/homeostasis (AUPRC = 0.13), but not small molecular transport (AUPRC = 0.0071) or signaling (AUPRC = 0.0019) processes. MitoNet1.0 was comprised of 732 core and 1011 auxiliary mitochondrial genes. Auxiliary mitochondrial processes included Golgi and vesicle-mediated transport, the Fanconi anemia pathway, and DNA repair, among

others. ***Conclusions:*** MitoNet1.0 is a relational mitochondrial network defined using genetic interaction co-similarities profiles. It represents an invaluable resource that can be used to query the function and relationship of core and auxiliary mitochondrial genes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1085T: Native Hawaiian and Pacific Islander reference panel improves genome-wide imputation quality and accuracy in the population

Authors:

M. Lansford[1], J. M. Lazaro Guevara[2], C. Littlefield[1], D. Stucki[1], M. H. Pezzolesi[1], E. J. Taylor[1], J. Taloa[1], K. Lao[1], J. P. Tavana[3], W. L. Holland[1], K. L. Raphael[1], M. G. Pezzolesi[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]McGill Univ., Montreal, QC, Canada, [3]Brigham Young Univ., Provo, UT

Abstract:

Genotype imputation allows for effectively leveraging array data in genetic studies through inferring genetic variants not present on genotyping arrays. However, existing publicly available genotype imputation reference panels face challenges in imputing genotype data from underrepresented human subpopulations. One such group, Native Hawaiians and Pacific Islanders (NHPI), are currently not included in major reference panels such as those from the 1000 Genomes Project (1KGP), Haplotype Reference Consortium, and TOPMed. Previous work centered on specific variants (Lin M., et al., 2020) or a subset of chromosomes (Carlson JC, et al., 2023) has shown that the inclusion of even small numbers of individuals of NHPI ancestry improves imputation of variants specific to this population. Here, we incorporate NHPI individuals into an imputation reference panel to measure the improvement in genome-wide imputation quality for NHPI genotypes.

We used whole genome sequences from 100 NHPI individuals recruited through the Utah Pacific Islander Diabetes Study (UPIDS) to build on the existing 3,202 samples in the 1KGP reference panel. We compared the performance of 1KGP with the combined panel of 1KGP and UPIDS samples in imputing genotypes in an independent cohort of 100 NHPI individuals across all 22 autosomes.

We found, compared to the 1KGP panel, that the UPIDS and 1KGP combined reference panel increased the overall mean imputation quality (r^2) from 0.15 to 0.34. Among variants with r^2 of at least 0.4, the percentage of well-imputed variants (r^2 >= 0.8) increased from 71% using the 1KGP imputation to 97% using the UPIDS and 1KGP combination. Across all chromosomes, the

concordance between imputed and genotyped variants for the 100 NHPI individuals was also consistently higher for the data imputed with UPIDS and 1KGP than for those imputed with only 1KGP, demonstrating the power of including NHPI individuals in inferring variants incorrectly imputed by existing reference panels.
The improvements in imputation quality and concordance in Native Hawaiians and Pacific Islanders with the inclusion of the UPIDS whole genomes in the reference panel illustrates the importance of accounting for population-specific variation in genotype imputation and lays the foundation for future work in genotype imputation of underrepresented populations.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1086T: Quality management strategy for genomic data of a National Genome Project in Korea

Authors:

J. Jeon, S. Joe, J. Kim, i-k. Lim, K. CHO, S. Jeon, J. Yang, J. Haeyoung; Korea Res. Inst. of BioSci. and Biotechnology, Daejeon, Korea, Republic of

Abstract:

A new wave of precision medicine has become an inevitable consequence of a drastic improvement in related technologies, including high-throughput sequencing and data analysis and sharing. Here, we introduce a National Genome Project (NGP) in Korea, that aims to collect high-quality genomic data for predicting and diagnosing diseases among Koreans. The NGP has been designed to encourage the scientific community to perform integrative analyses with large-scale and high-quality genome and clinical data while protecting the participants' privacy and securing data confidentiality. For this reason, we have been improving the analysis and quality control pipelines for the WGS data to call germline or somatic variants with consistency and confidence. In this study, we report on sample contamination and swapping during the pilot phase of the NGP, which involved 15,000 individuals. We also discuss unexpected cross-sample relatedness such as bone marrow transplants and twins. Additionally, we present an enhanced quality management framework to prepare for scaling up to 340,000 individuals in NGP phase I over the next five years. The framework comprises protocols and automated pipelines designed to ensure sample and analyzed data quality and reliability across the entire process of generating high-quality genomic data. These protocols could be suggested as one of the standards for conducting large-scale population genomics research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1087T: Race and ethnicity biases introduced by filtering genomics-connected electronic health records for patients with complete data

Authors:

N. Tatonetti[1], J. M. A. Cortina[1], Y. Fatapour[1], U. Gisladottir[1], K. Brown[1], M. Zietz[2]; [1]Cedars-Sinai Med. Ctr., West Hollywood, CA, [2]Columbia Univ., Franklin, TN

Abstract:

Objective: Integrated genomic and clinical databases from national biobanks have advanced the capacity for disease genomics research. Data quality and completeness filters are common when building clinical cohorts to address errors and limitations of data missingness and care fragmentation. However, these filters may unintentionally introduce systemic biases when they are correlated with race and ethnicity. These biases may invalidate any genetic associations identified. In this study, we examined the race/ethnicity biases introduced by applying common filters to four clinical records databases, including All of Us, UK Biobank, and two geographically distinct academic medical centers.

Materials and Methods: We used 16 filters commonly used in electronic health records research on the availability of demographics, medication records, visit details, observation periods, vital signs, and other data types. We evaluated the effect of applying these filters on self-reported race and ethnicity. This assessment was performed across four distinct data sets comprising approximately 12 million patients. We then evaluated the filters with strong race/ethnicity bias for their downstream effects on genetic associations with stroke, and its subtypes, using data from All of Us.

Results: Applying the observation period filter led to a substantial reduction in data availability across all races and ethnicities in all four datasets. However, among those examined, the availability of data in the white subgroup remained consistently higher compared to other racial groups after the application of each filter. Conversely, the Black/African American group was the most impacted by each filter three datasets. This highlights the potential biases that may arise in studies involving underrepresented groups. We found that these cohort differences result in variance in association statistics for top stroke genes that could have impacts on clinical care and decision making.

Discussion and Conclusion: Our findings underscore the importance of using only necessary filters as they might have consequences on the diversity and completeness of population data which particularly affects underrepresented communities. Researchers must consider these

unintentional biases when performing data-driven research and the variances introduced into downstream genomics analysis and explore techniques to minimize the impact of these filters, such as probabilistic methods or the use of machine learning and artificial intelligence.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1088T: STimage-1K4M: A histopathology image-gene expression dataset for spatial transcriptomics

Authors:

J. Chen, M. Zhou, W. Wu, J. Zhang, Y. Li, D. Li; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Recent advances in multi-modal algorithms have driven and been driven by the increasing availability of large image-text datasets, leading to significant strides in various fields, including computational pathology. However, in most existing medical image-text datasets, the text typically provides high-level summaries that may not sufficiently describe sub-tile regions within a large pathology image. For example, an image might cover an extensive tissue area containing cancerous and healthy regions, but the accompanying text might only specify that this image is a cancer slide, lacking the nuanced details needed for in-depth analysis. In this study, we introduce STimage-1K4M, a novel dataset designed to bridge this gap by providing genomic features for sub-tile images. STimage-1K4M contains 1,149 images derived from spatial transcriptomics data, which captures gene expression information at the level of individual spatial spots within a pathology image. Specifically, each image in the dataset is broken down into smaller sub-image tiles, with each tile paired with 15,000-30,000 dimensional gene expressions. With 4,293,195 pairs of sub-tile images and gene expressions, STimage-1K4M offers unprecedented granularity, paving the way for a wide range of advanced research in multi-modal data analysis an innovative applications in computational pathology, and beyond.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1089T: The ClinGen Alport Syndrome Variant Curation Expert Panel: Guiding variant classification in *COL4A3*, *COL4A4*, and *COL4A5*

Authors:

A. Byrne[1], J. Wongboonsin[2], M-B. Roberts[3], J. Savige[4], R. Lennon[5], ClinGen Alport Syndrome Variant Curation Expert Panel; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Cleveland Clinic, Cleveland, OH, [4]Univ. of Melbourne, Melbourne, Australia, [5]The Univ. of Manchester, Manchester, United Kingdom

Abstract:

Background: The ClinGen Alport Syndrome (AS) Variant Curation Expert Panel (VCEP) was established in 2023 with the purview to define variant classification criteria tailored for the genes *COL4A3*, *COL4A4*, and *COL4A5* in the context of AS and its associated phenotypes. These refined criteria are then used to determine expert classifications of variant pathogenicity and interpretations of relevance to disease causality. From the experience of ClinGen VCEPs across other disease areas, this process is demonstrated to enhance the accuracy and consistency of variant classification, and can decrease the number of variants that are of uncertain significance.
Methods: Using the 2015 ACMG/AMP Sequence Variant Classification framework as the basis, the AS VCEP is systematically reviewing the 26 defined lines of evidence to determine ways in which they can be refined to be more specific to the known biology of collagen genes and the known characteristics of AS presentations.
Results: The lines of evidence considered in variant classification fall into two main categories: human observational data and predictive and functional data. In refining the use of human observational data-related evidence, the AS VCEP have defined the phenotypes that must be present for an individual to be considered a 'case'; a concept integral to the application of multiple of these evidence types, and established thresholds to determine whether a variant's frequency in the population is consistent with disease causality. In reviewing predictive lines of evidence, the AS VCEP is paying particular attention to glycine variants affecting the Gly-X-Y repeat in the triple helical domain, developing guidance for how to appropriately weight the known damaging impact of this established, dominant negative disease mechanism in comparison to other variant types. Finally, recommendations are also being developed to guide calibrated integration of evidence provided by functional assays.
Conclusion: The work of the AS VCEP in developing tailored criteria for the classification of variants in *COL4A3*, *COL4A4*, and *COL4A5* will provide a harmonized framework that maximizes the utility of available evidence, leading to improved accuracy in variant classification and interpretation, in turn, improving clinical care for individuals with variants in these genes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1090T: The gene expression of rev erb alpha and ror alpha differs in individuals with different chronotypes when it comes to late night eating or disturbed eating schedule

Authors:

Q. Naz, K. Usman; KGMU, LUCKNOW, India

Abstract:

Background: Humans living in modern industrialized societies have easier access to light and are often forced to be awake by social norms against their preferred times independent of the master circadian rhythm. These environmental stimuli eventually lead to the disruption of the circadian system. Thus, both disruption of hormones and alteration in clock genes affect the distribution of chronotype.Aim: The aim of this study is to investigate the hormonal disruption & its relation to mRNA expression of circadian clock genes (related to metabolic homeostasis) RevErb alpha and ROR alpha in Type 2 diabetic T2DM patients with Morningness, Eveningness and mixed chronotype.Methodology: A total of 255 subjects (morningness n= 110, mixed n= 105 & eveningness n = 40) aged 18 to 60 years were recruited in Clinical OPD of General Medicine, KGMU.Serum leptin, Cortisol, Insulin IL-6 &IL1-β were tested by ELISA Kit in our biochemistry lab. 1 µg of total RNA was reverse transcribed into cDNA using Mol BioTM reverse transcriptase. RT-PCR analysis was performed by the AB Applied Biosystems in 96 well plates using the syber green according to the manufacturer's instructions. Primers were designed with the help of NCBI. GAPDH was used as an internal control. The relative quantification of mRNA expression was calculated as a ratio of target gene to GAPDH.Results: When we compared these 3 groups, Significant Different was found in various parameters. ROR α expression was reduced in eveningness (p= 0.071) and mixed chronotype (p= 0.025) when compared with morningness chronotype (p= 0.091). Also comparison of REV ERB α expression was reduced in eveningness (p= 0.039) & mixed chronotype (0.047) when compared with morningness type or early eaters.Conclusion: Results were completely different to our expectation. Involved hormones inflammatory markers & Clock gene (REV ERB α & ROR α) mRNA expression is significantly altered in eveningness and mixed chronotype as compared to morningness type in T2DM Patients. RT-PCR determination of this specific clock gene may be helpful for earliar detection of T2DM in patients with altered feeding habit & unbalanced lifestyle.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1091T: The NIH Comparative Genomics Resource (CGR): Maximizing the impact of research organisms to human health

Authors:

V. Schneider, N. O'Leary, V. Kodali, S. Rangwala, F. Thibaud-Nissen, P. Akuiyibo, A. Marchler-Bauer, T. Murphy; NIH/NLM/NCBI, Bethesda, MD

Abstract:

NCBI is developing the NIH Comparative Genomics Resource (CGR) to maximize the impact of eukaryotic research organisms and their genomic data to biomedical research. Comparative genomics involves the evaluation of genomic features from diverse organisms to systematically explore and evaluate biological relationships. Because all biological processes rely on genes, proteins, and the biochemical pathways they belong to, and these fundamental elements are shared across the tree of life, this approach can inform nearly any aspect of biology- and help identify new model organisms that may be used to address issues in human health. However, the relevant data and applications for the growing number of sequenced organisms are spread across resources, challenging their findability and use. Additionally, assembled genomes often contain contaminating sequences and lack annotations that inform comparisons.
CGR facilitates reliable comparative genomics analyses for all eukaryotic organisms through community collaboration and an NCBI genomics toolkit. The toolkit offers high-quality genomics-related resources impactful to users' research, including interconnected databases with access points enabling seamless navigation of NCBI content and interoperable data and tools that can integrate into users' workflows. We will present our recent developments on several of these resources, including a new public cloud-ready tool for annotating new eukaryotic genomes based on the NCBI RefSeq Eukaryotic Genome Annotation Pipeline (EGAP), new web-based graphical tools supporting the visual comparison of two or more genomes based on alignments, expansion of functional data available for comparative analysis including gene ontology and expression information, and NCBI Datasets, which provides web and programmatic interfaces that help users search, browse, and download genome, gene, and ortholog-related data. We will illustrate the value of CGR in the context of health-related workflows. We will also demonstrate how connecting community-supplied genome-related content to the organism-agnostic CGR can make it easier for researchers to use data from sources and organisms they may not have known about, and how community feedback has impacted CGR development. Through CGR, we can assist researchers in using genomic data available from thousands of organisms to answer questions involving human health.

This work was supported by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM), National Institutes of Health (NIH).

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1092T: The PhenoDB Ecosystem: Open-Source Web-based Tools for Genomic Discovery

Authors:

S. Jester[1], S. Griffith[2], R. Martin[1], L. Vail[1], C. Antonescu[1], A. Sanchez[2], N. Sobreira[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Univ, Baltimore, MD

Abstract:

PhenoDB, a web-based tool, provides methods to annotate and analyze the variants of patients and their family members with rare Mendelian diseases. The PhenoDB software ecosystem includes GeneMatcher and VariantMatcher, respectively developed to enable connections between patients, clinicians, and researchers from around the world who share an interest in the same genes and variants. GeneMatcher and VariantMatcher are part of the Matchmaker Exchange, with the principal goal of sharing genomic and phenotypic data to facilitate the discovery of novel disease variants and genes. VariantMatcher allows sharing of variant-level and phenotypic data from whole exome and genome sequencing. The PhenoDB ecosystem has undergone a significant overhaul over the past year, allowing for more robust and portable software tools that can be used by genetic and genomic centers of any size around the world. The code, deployment tools and instructions, and ancillary data used by PhenoDB are available freely. The current iteration of PhenoDB is written in Python 3.8 and Django 3.2.5, using MariaDB 10 and SOLR 9.4 for data storage and search capabilities. PhenoDB deployment is now completed via Docker Compose, which can be deployed to local or cloud-based environments. This simplified deployment process allows interested centers to deploy their own PhenoDB installation rapidly and to keep those deployments updated as new versions of PhenoDB are released. Over the past year, the PhenoDB software development team has assisted with several such installations and trained many groups worldwide in using the tools that PhenoDB makes available. The recent software overhaul of PhenoDB has introduced new analysis capabilities and additional database annotations, significantly enhancing variant filtering and prioritization. These include gnomAD constraints (missense Z-score and pLI score), CADD, and RAVEL scores. In addition to these improvements, the PhenoDB software team has worked to integrate a subset of the PhenoDB analysis tools into other software platforms. For instance,

PhenoDB baseline variant analysis is now available in CAVATICA. These modifications have expanded the use of the PhenoDB database and Analysis Tool, making it easier for researchers worldwide to discover novel disease variants and genes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1093T: The Single Cell Aging Library With A Genome Browser: SCALIWAG

Authors:

J. Bartz[1,2], X. Ma[1,3], L. Zhang[1,4], X. Dong[1,2]; [1]Inst. on the Biology of Aging and Metabolism, Univ. of Minnesota, Minneapolis, MN, [2]Dept. of Genetics, Cell Biology and Dev., Univ. of Minnesota, Minneapolis, MN, [3]Lillehei Heart Inst., Dept. of Med., Univ. of Minnesota, Minneapolis, MN, [4]Dept. of Biochemistry, Molecular Biology and Biophysics, Univ. of Minnesota, Minneapolis, MN

Abstract:

Despite recent advances, single-cell RNA sequencing (scRNA-seq) remains very expensive, particularly when sequencing large datasets. Consequently, many studies, especially large-scale ones, opt to use previously published scRNA-seq data for analysis. However, collecting a substantial amount of scRNA-seq data is challenging because most datasets store gene expression data in inconsistent formats. This problem is further compounded for researchers interested in aging, as many datasets lack age-related metadata. To address this significant gap, we have created the **S**ingle **C**ell **A**ging **Li**brary **W**ith **A G**enome Browser (SCALIWAG).
Research into the biology of aging necessitates the presence of age-related metadata for each sample. However, the largest repositories of scRNA-seq data, such as GEO and SRA, do not allow users to query datasets based on age metadata. Other curated databases that enable users to select datasets containing age-related information are either difficult to query, like the Human Cell Atlas, or comprise relatively small collections of data, like AgeAnno. Additionally, these databases do not require deposited data to be formatted consistently, leading to significant differences in gene notation and file formatting across datasets. These issues make combining datasets from different sources extremely difficult and time-consuming, requiring extensive preprocessing before meaningful analysis can be performed.
SCALIWAG addresses these issues by providing the scientific community with a one-stop-shop for scRNA-seq data from donors with known ages. The gene expression data in SCALIWAG has been standardized and stored in a consistent format using the same gene naming convention.

Therefore, SCALIWAG enables researchers to analyze large amounts of data from multiple studies without the laborious process of standardizing all the datasets. This will be especially useful as machine learning and artificial intelligence become more prevalent, as our database will provide a key resource for developing these models by supplying a large amount of training data. Additionally, we have analyzed differential expression over a selection of datasets in our database. This analysis can be explored using an interactive genome browser included on our website.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1094T: Too many transcripts? Managing transcript complexity with GENCODE Primary

Authors:

J. Loveland, R. Fatima, J. Gonzalez, R. Merritt, R. Bennett, C. Davidson, T. Hunt, A. Frankish; EMBL-EBI, Hinxton, United Kingdom

Abstract:

The Ensembl-Havana team produces the Ensembl/GENCODE reference geneset for the human and mouse genomes, including protein-coding genes, pseudogenes, long non-coding RNAs and small RNAs. High quality gene annotation is of fundamental importance for genome biology and clinical genomics; annotation that is incorrect or incomplete impacts downstream analysis and introduces potentially significant errors. The huge volume of new long-read transcriptomic data has the potential to drive a massive increase in the number of alternatively spliced transcripts at protein-coding genes and lncRNAs. These new transcripts are potentially useful but their large number requires that care is taken in how they are added to the geneset to avoid overwhelming those using the annotation for browsing or computational analysis.The GENCODE geneset currently has two transcript sets: Comprehensive (containing everything) and Basic (full-length coding transcripts at protein-coding genes, a subset of lncRNA transcripts, one transcript per locus for all other gene biotypes). As the set of full-length transcripts added from long transcriptomic data grows, GENCODE Basic tends towards GENCODE Comprehensive and loses its utility. So we have created a new subset - GENCODE Primary - which uses biological data, currently expression and evolutionary constraint, to evaluate the functional potential of features (exons, introns and splice sites) that are not present in the reference transcript for that locus. For protein-coding genes the reference is the MANE Select transcript (Matched Annotation from NCBI and EMBL-EBI) from our collaboration with RefSeq. We identify features

absent from the MANE Select transcripts that possess characteristics suggestive of function. We then run the Ensembl Select pipeline that was developed to identify candidate MANE Select transcripts to find the highest scoring transcript that includes the novel functional feature. This process is iterated until all novel features at a locus are represented by a full-length transcript. Expert annotators have reviewed the complete transcript set to ensure the quality of the pipeline, adding or removing transcripts when necessary. The GENCODE Primary set is not expected to be stable initially as transcripts may change while GENCODE incorporates much new long transcriptomic data and the pipeline will be re-run for each GENCODE release. We are working on adapting this method to create a GENCODE primary transcript set for the mouse genome and also using the rankings generated by the Ensembl Select pipeline to support additional filtering.This work is supported by NIH U24HG007234 and Wellcome Trust 226083/Z/22/Z.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1095T: Tracing the evolutionary path of the CCR5delta32 deletion via ancient and modern genomes

Authors:

K. Ravn[1], L. Cobuccio[1], R. Muktupavela[2], J. Meisner[1,3], L. S. Danielsen[1], M. E. Benros[3], T. S. Korneliussen[4], M. Sikora[5], E. Willerslev[4,6,7,8], M. Allentoft[4,9], E. Irving-Pease[4,2], F. Racimo[4,2], S. Rasmussen[1,10]; [1]The Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]Section for Molecular Ecology and Evolution, Globe Inst., Univ. of Copenhagen, Copenhagen, Denmark, [3]Copenhagen Res. Ctr. for Mental Hlth., Mental Hlth.Ctr. Copenhagen, Copenhagen Univ. Hosp., Copenhagen, Denmark, [4]Lundbeck Fndn. GeoGenetics Ctr., Globe Inst., Univ. of Copenhagen, Copenhagen, Denmark, [5]Lundbeck Fndn. GeoGenetics Ctr., Globe Inst., Univ. of Copenhagen, Copenhagen, Copenhagen, Denmark, [6]GeoGenetics Group, Dept. of Zoology, Univ. of Cambridge, Cambridg, United Kingdom, [7]Wellcome Sanger Inst., Wellcome Genome Campus, Cambridge, United Kingdom, [8]MARUM, Univ. of Bremen, Bremen, Germany, [9]Trace and Environmental DNA (TrEnD) Lab., Sch. of Molecular and Life Sci., Curtin Univ., Perth, Australia, [10]The Novo Nordisk Fndn. Ctr. for Genomic Mechanisms of Disease, Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

The chemokine receptor variant CCR5delta32 is linked to HIV-1 resistance and other conditions. Its evolutionary history and allele frequency (10-16%) in European populations have been

extensively debated. We provide a detailed perspective of the evolutionary history of the deletion through time and space. We discovered that the CCR5delta32 allele arose on a pre-existing haplotype consisting of 84 variants. Using this information, we developed a haplotype-aware probabilistic model to screen for this deletion across 900 low-coverage ancient genomes and traced the origin of CCR5delta32 to at least 7,000 years BP in the Western Eurasian Steppe region. Furthermore, we present evidence of positive selection acting upon the CCR5delta32 haplotype between 7,000-2,000 BP in Western Eurasia and that the presence of the haplotype in Latin America can be explained by post-Columbian genetic exchanges. Finally, we point to new complex CCR5delta32 genotype-haplotype-phenotype relationships, which demand consideration when targeting the CCR5 receptor for therapeutic strategies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1096T: Underdiagnosis and Health Disparities in Hereditary Hemorrhagic Telangiectasia

Authors:

N. Singh-Miller, C. Wilczewski, J. Johnston, L. Biesecker, C. Turner; NIH/NHGRI, Bethesda, MD

Abstract:

The American College of Medical Genetics list for reporting Secondary Findings in clinical exome and genome sequencing (ACMG SF) is a critical tool for opportunistic health care screening. Returning these findings to patients is now common in clinical genetics practice. In this study, we investigated the yield of SF for hereditary hemorrhagic telangiectasia (HHT) in the *All of Us* program v7 dataset. The dataset consists of ~245,000 short-read genomes, not selected for any specific disease or condition. Importantly, many patients with short-read genomes also have available coded electronic medical record (EMR) data (N~206,000).
HHT is an autosomal dominant (AD) condition, associated with the genes *ACVRL1* and *ENG*. The ACMG SF recommends reporting all pathogenic (P) and likely pathogenic (LP) variants in these genes. Within this context, we found a genotype-positive cohort of 56 individuals heterozygous for one of 43 variants: a rate of 1 in 4,500. We also considered a cohort of individuals who have an HHT diagnosis, 93 in 206,000: a rate of 1 in 2,200. These rates were higher than the current accepted prevalence estimates for HHT of 1 in 5,000 to 1 in 10,000.
Next, we investigated the ancestry breakdowns within these cohorts using PCA-derived genetic ancestry. For the genotype-positive cohort we find the expected ancestry breakdown matched that of the entire *All of Us* genotype cohort. This suggests that P/LP variants

in *ACVRL1* and *ENG* are proportionally distributed within the genetic-ancestry groups in All of Us. However, analysis of the genotype-phenotype overlapping cohort, revealed a likely under diagnosis of HHT within people of African descent; where individuals of this group were expected to account for 24% of diagnoses, but account for only 15% (p-value = 0.04). Additionally, we explored the genotype-phenotype relationships for these variants, recapitulating the expected HHT phenotypes through phenotype-wide association studies (PheWAS). However, our analysis also showed that these phenotypes are only present in some of the genotype-positive cohort; for example, HHT in 48%, epistaxis in 20%, and pulmonary arteriovenous malformation in 5%.
Here we have shown that genotype ascertainment identifies individuals at risk of HHT at a rate higher than the generally accepted prevalence. Our preliminary data suggest that there are health disparities in HHT diagnosis. We hypothesize that HHT is more common than currently accepted, that it is underdiagnosed, especially in populations of African descent, and that genomic ascertainment could be a tool to reduce those inequities.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1097T: Understanding Willingness to Share SBEDOH Information: Cultural Sensitivity and Community Engagement to enhance Alzheimer's disease research in the understanding of AD risk.

Authors:

L. Adams[1], F. Rajabli[2,3], M. Douglas[1], J. Sanchez[1], R. Moise[4], M. Cuccaro[1,3], M. Pericak-Vance[1,3], A. Seixas[5,4]; [1]Univ. of Miami, Hussman Inst. for Human Genomics, Miami, FL, [2]Univ. of Miami, Miami, FL, [3]Dr. John T Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami, Miami Fl, Miami, FL, [4]Dept. of Psychiatry and Behavioral Sci., Univ. of Miami, Miami, FL, Miami, FL, [5]Dept. of Informatics and Hlth.Data Sci., Univ. of Miami, Miami, FL, Miami, FL

Abstract:

Background. Understanding social, behavioral, and environmental determinants of health (SBEDOH) are key in identifying modifiable risk factors for Alzheimer’s Disease and related Dementias (ADRD), particularly the disproportionate burden seen in minorities. Identifying both genetic and SBEDOH factors contributing to AD risk and their interactions are critical in future precision medicine initiatives. However, a lack of standardized SBEDOH methods exists with current methods limited in scope and cultural sensitivity (i.e. can be used across different settings and populations and devoid of stigma). To address these gaps, we created an SBEDOH

survey and conducted focus groups in Black American (BA) and Hispanic/Latino (H/L) participants to determine the culturally sensitivity and how best to communicate the survey benefits to diverse groups.**Methods**. Focus groups were conducted, engaging 23 BA and 26 (H/L) individuals, using semi-structured interviews to elicit in-depth discussions around the cultural sensitivity of SBEDOH items (56 questions, across 13 SBEDOH domains). Sessions were transcribed verbatim, and thematic analysis performed using *NVivo* software. Two primary coders analyzed data, with a third coder resolving discrepancies.**Results**. We identified nine cultural sensitivity themes: initial reactions, participation interest, impact perception of information sharing, survey benefits communication, research engagement, tangible benefits, potential misuse of shared data, emotional reactions, and attitudes towards SBEDOH. BA's reluctance was linked to historical mistrust (e.g. slavery), while H/L's were about privacy. Both groups noted the importance of clarity on the survey's benefits to their communities and research at-large. Both groups emphasized the expectation that information be relayed by culturally responsible professional and SBEDOH should come with guidance on available social care services, highlighting the importance of a reciprocal relationship for sharing sensitive information.**Conclusion**. This study affirms that when approached with cultural sensitivity and clear communication, BA and H/L individuals are receptive to sharing SBEDOH information. The significance of culturally attuned methods and articulation of direct community benefits are critical to success, underscoring the need for an environment that fosters connectivity and trust resulting in enhanced participation in ADRD research. Moving forward we can examine environmental phenotypes in association with known risk and/or protective AD genes. Availability of SBEDOH data is critical to our understanding of the complex tapestry of AD risk.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1098T: Variant Impact Predictor database (VIPdb), version 2, with ClinGen PP3/BP4 calibrated threshold recommendations, and CAGI assessments

Authors:

S. E. Brenner[1], Y-J. Lin[1], A. S. Menon[1], Z. Hu[1], T. Bergquist[2], S. L. Stenton[3], E. A. W. Nadeau[4], A. B. Byrne[3], M. S. Greenblatt[4], S. M. Harrison[3], S. V. Tavtigian[5], A. O'Donnell-Luria[3], L. G. Biesecker[6], P. Radivojac[7], V. Pejaver[8]; [1]Univ. of California, Berkeley, CA, [2]Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Univ. of Vermont, Burlington, VT, [5]Huntsman Cancer Inst., Salt Lake City, UT, [6]NIH NHGRI, Bethesda, MD, [7]Northeastern Univ., Boston, MA, [8]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Variant interpretation is essential for identifying disease-causing variants from the millions detected in patients. Hundreds of Variant Impact Predictors (VIPs), also known as Variant Effect Predictors (VEPs), have been developed for this purpose, with a variety of methodologies and goals. To facilitate selection of VIPs, we created the Variant Impact Predictor database (VIPdb), with the new version 2 describing 392 VIPs developed over the past 25 years. Features in the database include publication details, access information, and citation patterns.
VIPdb now includes computer-parsable threshold scores from ClinGen recommendations for clinical classification using VIPs. Recommendations from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) for interpreting sequence variants specify the use of computational predictors at the lowest, "supporting," level of evidence for pathogenicity or benignity using criteria PP3 and BP4, respectively. Through rigorous independent calibration, we have shown that some computational tools could be used at higher levels of evidence strength, with many able to provide “strong” evidence, including new ones calibrated in the past year. These calibrated thresholds enhance the role of VIPs in providing evidence for clinical decisions, potentially improving genetic disease diagnosis.
To help users select amongst methods, VIPdb now also includes results from CAGI, the Critical Assessment of Genome Interpretation. CAGI aims to advance the state-of-the-art for computational prediction of genetic variant impact, particularly where relevant to disease, by running experiments in which participants make blind predictions of phenotypes from genetic data, and these are evaluated by independent assessors. This comprehensive collection in VIPdb facilitates VIP exploration for various applications of variant interpretation.
Availability: VIPdb version 2 is available at https://genomeinterpretation.org/vipdb

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Thursday Session

Board 1099T: Whole genome sequencing analysis of Alzheimer's disease in a west African cohort: preliminary results from the DAWN study

Authors:

B. Ayele[1,2], F. Rajabli[1], L. Adams[1], P. Whitehead[1], M. Coker[3], K. Akinwande,[4], S. Diala[4], M. Ogunronbi[4], K. Scott[1], R. Obiako[3], K. Hamilton-Nelson[1], K. McInerney[1], A. Zaman[1], I. Martínez[1], K. Wahab[5], P. Mena[1], A. Akpalu[6], B. Kunkle[1], F. Sarfo[7], J. Vance[1], N. Okubadejo[8], O. Baiyewu[9], M. Cuccaro[10], S. Blanton[1], W. Bush[11], J. Haines[11], R. Kalaria[12,13], G. Byrd[14,15], A. Ogunniyi[16], M. Owolabi[17], A. Griswold[1], M. Pericak- Vance,[1], R. Akinyemi[16,18]; [1]John P. Hussman Inst. for Human

Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Dept. of Neurology, CHS, Addis Ababa Univ., Addis Ababa, Ethiopia, [3]Ahmadu Bello Univ., Zaria, Nigeria, [4]Inst. for Advanced Med. Res. and Training, Coll. of Med., Univ. of Ibadan, Ibadan, Ibadan, Nigeria, [5]Univ. of Ilorin Teaching Hosp., Ilorin, Nigeria, [6]Univ. of Ghana, Accra, Ghana, [7]Komfo Anokye Teaching Hosp., Kumasi, Ghana, [8]Univ. of Lagos, Lagos, Nigeria, [9]Univ. Coll. Hosp., Ibadan, Nigeria, [10]John P. Hussman Inst. for Human Genomics, Miami, FL, [11]Case Western Reserve Univ, Cleveland, OH, [12]Newcastle Univ., Newcastle, United Kingdom, [13]Univ. of Nairobi, Nairobi, Kenya, [14]NC A&T State Univ, Greensboro, NC, [15]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Sch. of Med., Winston-Salem, NC, [16]Coll. of Med., Univ. of Ibadan, Ibadan, Nigeria, [17]H3Africa Stroke Investigative Res. & Ed.al Network (SIREN), Univ. of Ibadan, Nigeria, [18]NeuroSci. and Aging Res. Unit, Inst. of Advanced Med. Res. and Training, Coll. of Med., Univ. of Ibadan, Ibadan, Nigeria

Abstract:

Abstract: Background: Alzheimer's disease (AD) is the most common neurodegenerative disorder among the aging population with a strong genetic contribution. The DAWN study is aimed to investigate the genetics of AD in an underrepresented diverse population including nine countries in sub-Saharan Africa through the Africa Dementia Consortium (AfDC). In this analysis, we aimed to identify rare, pathogenic AD variants through a whole genome sequencing (WGS) analysis of cohort of individuals from Nigeria and Ghana. **Method:** We performed whole genome sequencing of 185 individuals from AfDC sites in Nigeria and Ghana. DNA was extracted at the University of Ibadan, Nigeria and an aliquot shipped to the University of Miami for processing. The sequencing was done using standard PCR-free Illumina whole genome sequencing protocols on the NovaSeqX Plus instrument. We applied an Alzheimer's Disease Sequencing Project (ADSP) developed bioinformatics pipeline and quality control procedures. Variants were filtered for those in 22 established AD genes and risk loci based on the ADSP Gene Verification Committee recommendations including genes with variants causing familial AD (e.g. *PSEN1, PSEN2, APP*) and those with rare susceptibility variants (e.g. *SORL1, TREM2, APOE*). The prioritized rare variants were confirmed using Sanger sequencing. **Result:** Of the 185 participants, the mean age at examination was 73.1 years and females accounted for 54.1%. These were divided into half AD cases (47.5%) and cognitively unimpaired controls (52.5%). After filtering for AD genes, we identified a total of 25,572 variants of which 437 (1.7%) were protein coding. Filtering further for population frequency, we identified a novel mutation in PSEN1 changing from alanine to threonine at position 431 (PSEN1 A431T) in a 63-year-old male AD patient from Nigeria with positive family history of AD. This variant has never been identified, though other amino acid changes have been reported in familial AD. Similarly, we identified a previously reported SORL1 variant introducing a premature stop codon at amino acid 985 (SORL1 R985X) variant in 58-year-old female patients from Nigerian with diagnosis of AD. This variant has previously been identified in two AD cases of European background and

other protein truncating variants in SORL1 have been shown to have strong effects on AD risk. **Conclusion:** We identified a novel PSEN1 variant and a rare SORL1 in this unique west African population. This finding supports the importance of the study of diverse ancestries to identify new and recurrent AD causative/risk variants.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2002T: A compound heterozygous variant in *SLC52A2* causes BROWN-VIALETTO-VAN LAERE syndrome in a Malian family

Authors:

A. Bocoum[1], K. Nimaga[2], S. Bamba[1,3], M. Goita[1], A. Maiga[1], M. Doumbia[2], O. Traore[1], M. Sangare[1], S. Diarra[1,3], A. Yalcouyé[4,1], C. Guinto[1], M. Khokha[3], S. Lakhani[3], G. Landoure[1]; [1]Univ. of Sci. Techniques and Technologies of Bamako, Bamako, Mali, [2]Clinique médicale Dinandougou, Marka-Coungo, Mali, [3]Pediatric Genomics Discovery Program, Dept. of Pediatrics, Yale Univ., New Haven, CT, [4]JHU, USA, Bamako, Mali

Abstract:

Introduction: The Brown-Vialetto-Van Laere syndrome (BVVL) is a rare neurological disorder characterized by progressive pontobulbar palsy associated with sensorineural deafness. It is caused by mutations in the *SLC52A2* gene, which encodes for a protein of the Riboflavin Transporter family. Approximately 50% of cases are familial, with a suspected autosomal recessive inheritance. To our knowledge, only one case has been described in Africa, making the data about this condition scarce in our region. Here, we describe the clinical and genetic features of BVVL syndrome in a Malian family. Aim: To clinically characterize a patient with BVVL syndrome and identify the underlying genetic defect. Method: One patient and his relatives were enrolled in our research protocol after full written consent. The patient was thoroughly examined by neurologists and an EEG was performed. DNA was extracted for Whole Exome Sequencing (WES). Segregation analysis was done in all available family members. In addition, *in silico* analysis was done to assess deleteriousness. Results: An 8-year-old boy was referred to our unit for seizures and developmental delay. The Disease started at the age of 4 with episodes of absence occurring once a week, that gradually worsened to be once a day. No history of similar cases in the family was noted. Neurological examination found a unilateral convergent strabismus on the left eye, tetraparesis, and diffuse muscle atrophy of the four limbs. He also had presented coordination disorders. EEG showed a globally slow background with generalized abnormalities. WES identified a compound heterozygous variant in the *SLC52A2* gene (c.151G>A; p.V51I and c.155C>T; p.S52F) in the affected individual, segregating with the disease

status in the family. These variants are classified of uncertain significance and pathogenic, respectively, according to ClinVar. Conclusion: This study reports the second molecularly diagnosed case of BVVL syndrome in sub-Saharan Africa, expanding the spectrum of this condition. It also opens the way to determining the spectrum of SLC52A2-related conditions and further our understanding in their pathophysiology.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2003T: A *Drosophila* model of mucopolysaccharidosis IIIB

Authors:

B. Simkhada, N. O. Nazario-Yepiz, T. F. C. Mackay, R. R. H. Anholt, P. S. Freymuth, V. Shankar, R. Lyman; Clemson Univ., Greenwood, SC

Abstract:

Mucopolysaccharidosis type IIIB (MPS IIIB) is a rare lysosomal storage disorder caused by defects in alpha-N-acetylglucosaminidase (NAGLU) and characterized by severe effects in the central nervous system. Mutations in *NAGLU* cause accumulation of partially degraded heparan sulfate in lysosomes. The consequences of these mutations on whole genome gene expression and their causal relationships to neural degeneration remain unknown. Here, we used the functional *Drosophila melanogaster* ortholog of *NAGLU*, *Naglu*, to develop a fly model for MPS IIIB induced by gene deletion ($Naglu^{KO}$), missense ($Naglu^{Y160C}$), and nonsense ($Naglu^{W422X}$) mutations. We used the Drosophila activity monitoring system to analyze activity and sleep and found sex- and age-dependent hyperactivity and sleep defects in mutant flies. Fluorescence microscopy on mutant fly brains using Lysotracker dye revealed a significant increase in acidic compartments. Differentially expressed genes determined from RNA sequencing of fly brains are involved in biological processes that affect nervous system development. A genetic interaction network constructed using known interacting partners of these genes consists of two major subnetworks, one of which is enriched in genes associated with synaptic function and the other with neurodevelopmental processes. Our data indicate that lysosomal dysfunction arising from disruption of heparan sulfate breakdown has widespread effects on the steady state of intracellular vesicle transport, including vesicles associated with synaptic transmission. Evolutionary conservation of fundamental biological processes predicts that the *Drosophila* model of MPS IIIB can serve as an *in vivo* system for the future development of therapies for MPS IIIB and related disorders.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2004T: A Novel Variant in *ARSA* Causes Unusual Phenotype of Infantile Metachromatic Leukodystrophy

Authors:

K. Sanni[1], A. Maiga[1], A. Arama[2], S. Bamba[1,3], M. Goita[1], A. Yalcouyé[4,5], M. Sangare[1], M. Albakaye[6], M. Kotioumbe[1], S. Djimde[7], S. Diarra[1,3], M. Khokha[3], S. Lakhani[3], G. Landoure[1]; [1]Univ. of Sci. Techniques and Technologies of Bamako, Bamako, Mali, [2]Ctr. de Santé Communautaire (CSCOM), Fégui, Mali, bamako, Mali, [3]Pediatric Genomics Discovery Program, Dept. of Pediatrics, Yale Univ., New Haven, CT, [4]Faculté de Médecine et d'Odontostomatologie, Université des Sci., des Techniques et des Technologies de Bamako, Bamako, Mali, Bamako, Mali, [5]McKusick-Nathans Inst. and Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, USA, Baltimore, MD, [6]Service de Médecine, Hôpital « Fousseyni Dao », Kayes, Mali, [7]Service de Neurologie, Ctr. Hosp.ier Univ.ire « Point G », Bamako, Mali

Abstract:

Introduction: Metachromatic Leukodystrophy (MLD) is a rare autosomal recessive lysosomal storage disorder, caused by variants in the Arylsulfatase A gene (*ARSA*) gene. MLD mostly affects the white matter and the peripheral nervous system, leading to serious disability and early death. Several variants in the *ARSA* gene have been reported to cause MLD across different populations, however, reported cases are scarce in sub-Saharan Africa (SSA). Here, we report a novel missense variant in *ARSA* gene causing an unusual phenotype of MLD in a Malian patient, the first in this population and further expands its genetic spectrum, globally. **Aim**: To study the clinical and genetic features of Metachromatic Leukodystrophy in one Malian family **Methods**: Four individuals from a Malian Soninke family were enrolled in our protocol after giving full and informed consent. After undergoing a thorough neurological examination, Electroencephalography (EEG) and brain imaging were performed in the proband. For genetic testing, DNA was extracted from peripheral blood from all four individuals for Whole Exome Sequencing (WES) and subsequent segregation studies. Additionally, the putative variant was checked for pathogenicity using *in silico* tools. **Result**: A 2-year-old boy, born from healthy consanguineous parents, was seen in our clinic for developmental disorders and spastic seizures. Family history is consistent with autosomal recessive inheritance. Disease began at 11 months of age with gait and speech disorders that gradually led to complete loss of ambulation and speech. Neurological examination found spastic tetraparesis, axial hypotonia, continuous spasms, bilateral reactive mydriasis, speech impairment, continuous crying and bilateral Babinski sign. EEG showed abnormal electric activity. CT scan revealed marked cortical and subcortical atrophy, associated with periventricular hypodense signals. The clinical features

were initially suggestive of West syndrome. However, WES identified a novel missense variant (c.191T>C:p.Phe64Ser) in ARSA, segregating with the phenotype, confirming the molecular diagnosis of MLD. This variant is predicted to be deleterious by several *in silico* tools including CADD=31 and affect a highly conserved domain across different species. **Conclusion**: We established the molecular diagnosis of MLD in a SSA family for the first time, expanding its genetic spectrum. Moreover, our patient presented with uncommon clinical features, suggesting that some genetic disorders might have a unique profile in the African population.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2005T: A novel variant in *CLN5* of Japanese girl with Neuronal Ceroid Lipofuscinosis

Authors:

E. Nishi[1], K. Yanagi[2], N. Okamoto[1], K. Yanagihara[3], T. Kaname[2]; [1]Dept. od Med. genetics, Osaka Women's and Children's Hosp., Osaka, Japan, [2]Dept. of Genome Med., Natl. Res. Inst. for Child Hlth.and Dev., Tokyo, Japan, [3]Dept. of child neurology, Osaka Women's and Children's Hosp., Osaka, Japan

Abstract:

Introduction:Neuronal ceroid lipofuscinoses (NCL) constitute a spectrum of inherited neurodegenerative disorders characterized by progressive psychomotor regression, refractory epilepsy, and visual impairment. Pathologically, these disorders are known for the accumulation of lipofuscin granules within lysosomes. To date, types 1 through 14 (excluding type 9) have been reported, with diverse causative genes suggesting varied pathogenic mechanisms, though many aspects remain unclear. CLN5 (MIM#256731), also known as the late-infantile type or Finnish type, typically manifests between ages 4 and 17 and presents diagnostic challenges distinct from CLN1 or CLN2 which can be enzymatically diagnosed, relies on identifying accumulations in fibroblasts or through genetic testing. In Japan, prevalence is unknown, considered very rare, with no reported individuals. We report the diagnosis of CLN5 in a Japanese individual with severe intellectual disability, regression, and progressive myoclonus epilepsy, conditions posing diagnostic challenges, through trio whole-genome sequencing (WGS) revealing a novel homozygous pathogenic variant in *CLN5*. Clinical Report:An 11-year-old Japanese girl presented with progressive myoclonic epilepsy and psychomotor regression; differential diagnosis such as GM2 gangliosidosis and CLN1/CLN2 were considered and enzyme activities were performed, but no diagnosis was made. Trio-WGS identified a novel homozygous variant (c.966T>G, p. (Tyr322*)) in *CLN5* with carrier parents. Subsequent evaluations showed

progressive cerebral and cerebellar atrophy and optic nerve thinning on brain MRI, blindness, along with increased lipofuscin granules in fibroblasts observed via electron microscopy. Discussion:Early detection, facilitated by comprehensive genetic testing such as WGS, is critical in diagnosing particularly rare conditions like CLN5. Gene therapy research for CLN5 is advancing, with phase I/II clinical trials underway, offering hope for this rare progressive disorder. Limitations and Conclusions:Limitations of this study include the single reported case and the lack of functional analysis of CLN5. Further studies, data accumulation, and the development of a diagnostic system for early detection of CLN5 are anticipated in the future.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2006T: A prevalent *MOCS2* variant in the Roma population is associated with a mild form of molybdenum cofactor deficiency

Authors:

S. Cho[1], C. Winkler[2], B. Stiburkova[3]; [1]Ajou Univeristy Coll. of Med., Suwon, Korea, Republic of, [2]Leidos BioMed. Res. Inc./NCI, Frederick, MD, [3]Faculty of Med., Charles Univ. in Prague, Prague 2, Czech Republic

Abstract:

Xanthinurias are rare inherited disorders of purine metabolism. Xanthinuria type III is caused by molybdenum cofactor deficiency (MoCD) due to pathogenic variants in *MOCS1*, *MOCS2*, or *GEPH* genes. Here we described five Roma patients from four independent families with putative xanthinuria I/II with hypouricemia, accumulation of xanthine/hypoxanthine, deficiency of xanthine oxidase activity, variable age of diagnosis and only asymptomatic or mild clinical signs and symptoms. Whole exome sequencing (WES) was performed on all probands, aged 3 to 43 years, due to lack of genetic confirmation for xanthinuria types I and II. The causality of the putative pathogenic variant was confirmed by analysis of sulfite and related metabolites and in vitro functional characterization of metal-binding pterin (MPT) synthesis and protein complex formation. Considering the rarity of the condition and recessive inheritance, 34 candidate variants were identified after filtering out allele frequency threshold in Non-Finish Europeans. An ultra-rare variant rs776441627 c.244A>T (p.Ile82Phe) in two overlapping reading frames in *MOCS2* segregated with the disease in all five patients (four homozygotes, one compound heterozygote). The variant has an allele frequency of of 3.6% in a Roma population control group. Functional characterization revealed the significantly decreased MPT synthesis activity and confirmed the causality of c.244A>T in MoCD. The c.244A>T variant in *MOCS2* is causal variant for MoCD with mild to asymptomatic clinical phenotype and fully penetrant biochemical

phenotype. Hypouricemia should be kept in differential diagnostic algorithm of pediatrics and adult patients with neurological signs and *MOCS2* should be considered for inclusion of gene panels for clinical genetic screening for patients with xanthinuria.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2007T: An effective frameworkpredicting splicing single nucleotide variants in exome sequencing ⭐

Authors:

Y. Utsuno[1], K. Hamanaka[1,2], M. Sakamoto[1,3], N. Tsuchida[1,3], Y. Uchiyama[1,3], E. Koshimizu[1], A. Fujita[1], S. Miyatake[1,4], T. Mizuguchi[1], N. Matsumoto[1]; [1]Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [2]The Inst. for the Advanced Study of Human Biology (ASHBi), Kyoto Univ., Kyoto, Japan, [3]Dept. Rare Disease Genomics, Yokohama City Univ. Hosp., Yokohama, Japan, [4]Dept. of Clinical Genetics, Yokohama City Univ. Hosp., Yokohama, Japan

Abstract:

Splicing variants play a significant role in approximately 10% of Mendelian disorders. While recent software tools, like SpliceAI, have enhanced the precision of predicting these variants, they may have limitations when used independently. In our study, we have developed a system for the effective screening of pathogenic single nucleotide variants (SNVs) related to splicing. This system aligns with the 2023 ClinGen guidelines, emphasizing a comprehensive evaluation of genes and variants by assigning a Priority Score (ranging from 0 to 9) to all SNVs specifically within open reading frames and intron regions. A higher Priority Score indicates a higher likelihood of a pathogenic splicing variant. To validate our approach, we used a set of known pathological splicing variants from HGMD Professional 2023.4 as True Positives and variants with a minor allele frequency exceeding 1% in gnomAD v2.1.1 as True Negatives. Our framework exhibited superior performance in discriminating between these variants compared to using SpliceAI in isolation (with an area under the receiver operating characteristic curve (AUC) of 0.996 versus 0.994, validated by DeLong's test with a p-value of 1.95×10^{-4}). Furthermore, when implementing this approach on 1,339 families with unidentified causative variants from exome sequencing (ES), we identified splicing variants potentially associated with diseases in genes such as *MECP2*, *JAKMIP1*, *PDHA1*, and *COL2A1* across four cases. Additionally, we identified *ZNF516*, *NFE2L1*, and *UBN1* as promising new candidate genes that may link to conditions like cerebellar vermis hypoplasia, autism spectrum disorder, and porencephaly. Our framework enhances the efficiency of detecting pathogenic splicing variants in ES.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2008T: An Oral-Facial-Digital Syndrome Patient Who Expands the Phenotypic Landscape of *KIAA0753* Related Disorders Suggesting a Continuum of Disease

Authors:

J. Tellerday[1], J. Black[2], D. Hansen[3], A. Zakrzewski[4], C. Fortner[3], N. Dosa[2], R. Lebel[2]; [1]SUNY Upstate Med. Univ., Syracuse, NY, [2]Ctr. for Dev., Behavior, and Genetics, SUNY Upstate Med. Univ., Syracuse, NY, [3]Dept. of Pediatrics, SUNY Upstate Med. Univ., Syracuse, NY, [4]Dept. of Orthopedic Surgery, SUNY Upstate Med. Univ., Syracuse, NY

Abstract:

KIAA0753 (OMIM #617112) encodes a protein complex subunit involved in cilial maintenance and ciliogenesis that has been associated with several disorders: Oral-Facial-Digital (OFD) syndrome Type XV, Joubert Syndrome Type 38, and a form of Short-Rib Thoracic Dysplasia (SRTD). These diagnoses display considerable phenotypic variability making specific diagnosis challenging. We present a patient who we believe to be the oldest documented person with OFD type XV, and *KIAA0753* related disease, who displays novel phenotypic characteristics. This 18 year old female was born at 36 weeks via C-section after abnormal ultrasound findings included ventriculomegaly and macrocephaly. She had a three week NICU stay due to apnea, seizures, and aspiration. Imaging revealed corpus callosum partial agenesis and gray matter heterotopia. A diagnosis of oral-facial-digital syndrome (presumed Type I) was made in the neonatal period due to phenotypic elements including polydactyly of the left foot and right hand, brachydactyly, and high arched palate. An Invitae Skeletal Disorders Panel (NGS) identified two heterozygous pathogenic variants in *KIAA0753*: c.1891A>T (p.Lys631*) and c.757C>T (p.Arg253*). An intronic variant in *KIAA0753* was also reported, of uncertain significance. Variants of uncertain significance (VUS) were also reported in *LONP1* and *FGF23*. Other genes associated with OFD were included and did not display variants. A whole genome chromosomal microarray was negative. A skeletal survey was undertaken; while the patient has skeletal manifestations including short stature and scoliosis, she did not display rib abnormalities (shortening, broadening) or rhizomelia typically seen in SRTD.
As the patient aged, she required a care team of numerous medical specialties. Endocrine dysregulation is uncommon in the literature and has only been seen as growth hormone deficiency, which she has. She demonstrates primary amenorrhea, of which caregivers should be aware as female patients enter teenage years. Gastrointestinal tract manifestations are rare;

our patient required gastrostomy in the newborn period and has persistent GERD and constipation. She also displays novel ocular manifestations including coloboma, cataract, and retinal detachment. Persistent seizures outside of the newborn period have not been documented; our patient had seizures throughout her first years of life controlled with oxcarbazepine. We propose that the phenotypic spectrum of *KIAA0753* related disease is broadened by these observations, and suggest a continuum among the discrete disorders linked to *KIAA0753*.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2009T: Analysis of a novel complex structural abnormality of the *FLNA* gene in a patient with periventricular nodular heterotopia and thrombocytopenia using HiFi read sequencing

Authors:

K. Yanagi[1], M. Yamamori[1], A. Sakamoto[2], T. Uchiyama[3], S. Kunishima[4], D. Hasegawa[5], D. Kawaba[6], A. Kobayashi[7], A. Tamura[8], A. Igarashi[1], Y. Matsubara[1], A. Ishiguro[9], T. Kaname[1]; [1]Dept. of Genome Med., Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [2]Div. of Hematology, Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [3]Dept. of Human Genetics, Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [4]Dept. of Med. Technology, Sch. of Hlth.Sci., Gifu Univ. of Med. Sci., Gifu, Japan, [5]Dept. of Hematology and Oncology, Kobe Children's Hosp., Hyogo, Japan, [6]Dept. of Hematology and Oncology, Hyogo Prefectural Kobe Children's Hosp., Hyogo, Japan, [7]Dept. of Pediatrics, Natl. Hosp. Organization Kobe Med. Ctr., Hyogo, Japan, [8]Dept. Pediatrics, Kobe Univ. Graduate Sch. of Med., Hyogo, Japan, [9]Ctr. for Postgraduate Ed. and Training, Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan

Abstract:

Periventricular nodular heterotopia (PNH) is an X-linked dominant disorder characterized by nodule formation in the brain tissue lining the ventricles due to the failure of neuronal migration from the ventricular zone to the cortex during development. *FLNA* (NM_001110556.2) is involved in various cellular functions. Loss-of-function variants of *FLNA* are known to cause not only PNH, but also congenital heart disease, lung disease, gastrointestinal symptoms, and macrothrombocytopenia. To date, more than 400 pathogenic or likely pathogenic variants are registered in the literature and the ClinVar database, most of which are SNVs or simple CNVs. Here we present a novel complex structural abnormality of *FLNA* revealed in a patient with PNH and thrombocytopenia using HiFi read sequencing. [Case] The proband was 17-year-old female born at 37 weeks of gestation with

weight 2,412 g. Her head control was at age 5 months and she started to walk at 1 years and 8 months. Her brain MRI showed lateral ventricle subependymal ectopic gray matter at 2 years of age. She had first febrile seizures at age 7 and thrombocytopenia with giant platelets was observed at age 12. Based on brain MRI and thrombocytopenia, filamin deficiency was suspected, but candidate variants were not detected by gene panel testing for platelet disorders. [Results] Whole exome analysis detected a copy number loss spanning intron 31 and the 3'UTR in *FLNA*. In addition to the copy number loss, at least three split reads were detected, suggesting a large inversion with complex duplications and deletions in the *FLNA* gene. Subsequent HiFi read sequencing revealed detailed structural abnormalities. [Conclusion] HiFi read sequencing was useful for analyzing complex structural abnormalities. It is important to utilize long lead sequencing to accumulate knowledge about complex structural abnormalities.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2010T: Ankylosing spine-brachydactyly syndrome: A possible novel skeletal dysplasia

Authors:

S. Eroglu[1], G. Nishimura[2], C. R. Ferreira[1], J. Hamada[3], K. Sano[4], E. Leon[5]; [1]NHGRI, Bethesda, MD, [2]Musashino-Yowakai Hosp., Tokyo, Japan, [3]Ehime Univ. Hosp., Ehime, Japan, [4]Ehime Rehabilitation Ctr. for Children, Ehime, Japan, [5]Children S Natl. Med. Ctr., Washington, DC

Abstract:

We report a novel autosomal dominant skeletal dysplasia observed in five individuals across three unrelated families. The shared phenotypic features include short stature, distinctive facial features (arched eyebrows, small palpebral fissures, convex nasal ridge, narrow mouth), various ocular abnormalities, short neck, rigid spine, brachydactyly, and limited joint mobility. The overlapping radiological features are characterized by severe brachydactyly, an ankylosing spine with coronal clefts, mild iliac hypoplasia, and broadening of long bones.
Proband 1 was born to Caucasian parents, and his phenotype was remarkable for arched eyebrows, small palpebral fissures, convex nasal ridge, narrow mouth. His radiographs showed brachycephaly, ankylosis of the cervical spine, coronal clefts, mild iliac hypoplasia, brachydactyly, and broadening of long bones. He underwent surgery for craniovertebral junction stenosis. His affected mother shared the phenotype, but also had high-frequency hearing loss and dental anomalies. His chromosome microarray and exome sequencing were unrevealing but identified maternally inherited variants in several genes not previously associated with disease. Proband 2, who is of Japanese heritage and has an affected father,

displayed a phenotype similar to that of Proband 1. She exhibited congenital restriction of joint mobility, hyperopia, and short stature, with growth improvement following GH therapy. Her radiographs revealed brachycephaly, ankylosis of the cervical spine, bamboo thoracolumbar spine, coronal clefts, iliac hypoplasia, broadening of long bones, and brachydactyly. Proband 3 is a 3-year-old girl from Argentina who had brachycephaly, Klippel-Feil-like ankylosis, bamboo spine, severe upper thoracic kyphosis, multiple coronal clefts, mild iliac hypoplasia, and severe brachydactyly.
These shared phenotypic traits suggest the presence of an autosomal dominant and distinct recognizable syndrome, with the most striking features being short stature, distinctive facial features, ankylosing spine and severe brachydactyly. We aim to identify a novel gene by collecting next-generation sequencing data from probands 2 and 3 and their affected parents, followed by performing comparative analyses. In conclusion, this research has the potential to enhance our knowledge about a possible novel skeletal dysplasia and elucidate the genetic basis of this newly reported, recognizable malformation syndrome.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2012T: Assessment of suspected candidate gene variants in hypermobility Ehlers-Danlos syndrome patients from the HEDGE study cohort

Authors:

C. Laukaitis[1,2,3], D. N. Subramanian[4], V. Janousek[5], A. Seth[6], W. He[6], M. Galindo[7], R. Handsaker[8,9], C. Francomano[10,11], W. Gandy[11], J. Hirschhorn[6,8], HEDGE Consortium; [1]Carle Illinois Coll. of Med., Urbana, IL, [2]Inst. for Genomic Biology, Urbana, IL, [3]Carle Hlth., Urbana, IL, [4]Peter MacCallum Cancer Ctr., Melbourne, Australia, [5]Charles Univ., Prague, Czech Republic, [6]Boston Children's Hosp., Boston, MA, [7]Univ. of Arizona, Tucson, AZ, [8]Broad Inst., Cambridge, MA, [9]Harvard Med. Sch., Boston, MA, [10]IU Sch. of Med., Indianapolis, IN, [11]The Ehlers-Danlos Society, New York, NY

Abstract:

Background: Hypermobile Ehlers-Danlos syndrome (hEDS) is a significant diagnostic challenge, due to its unknown genetic etiology and phenotypic overlap with other hereditary connective tissue disorders (HCTDs). The hEDS Genetic Evaluation (HEDGE) study aims to uncover its genetic basis through variant analysis of known disease-associated genes alongside discovery of novel causative genes. **Methods:** Whole genome sequencing and phenotyping was performed on 1012 individuals with a clinical diagnosis of hEDS. Variants in known EDS/HCTD genes (n = 90), plus other proposed candidate disease genes (n = 184), were filtered for rarity (GnomAD

AF ≥ 0.005) and impact (HIGH: protein-truncating and canonical splice site variants; MODERATE: predicted pathogenic missense and in-frame indel variants). Suspected disease-associated variants were further prioritised for curation and classification against the 2015 ACMG guidelines with subsequent modifications and gene-specific recommendations. **Results:** Out of 53 HIGH and MODERATE impact variants prioritised for curation (from a total of 519 post-filtering), six likely pathogenic heterozygous variants in EDS/HCTD autosomal dominant disease genes (COL1A2, COL5A1, COL5A2, FBN1, SMAD3, LOX) were identified in six individuals. Except for LOX, all were thought to explain the phenotype and confirm a diagnosis other than hEDS in these individuals. Three other likely pathogenic heterozygous variants denoting carrier status for autosomal recessive disease (COL12A1, C1S, ALDH18A1) were identified in three individuals. The remaining prioritised variants were classified as variants of uncertain significance (n = 35) or likely benign (n = 9); curation of remaining potential disease-associated EDS/HCTD variants is ongoing. Amongst the 184 candidate disease genes, no HIGH or MODERATE impact variants were present in the cohort at a high frequency (>10 unique variants per gene). **Conclusions:** Very few individuals with clinically diagnosed hEDS appear to have an alternative genetic explanation for their condition, emphasizing the importance of expert phenotyping and the utility of the HEDGE cohort for investigating the unknown etiology of hEDS. Future work will include GWAS and burden analysis of the cohort versus a matched unaffected control population to uncover common and polygenic causes and identify novel candidate genes and variants, enhancing our understanding of hEDS.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2013T: Candidate variants in regulatory and non-coding regions of coding genes identified in 4 large UK families from an NGS study - 2x HTAAD [Hereditary Thoracic Aortic Aneurysm and Dissection] and 2x hEDS [hypermobile Ehlers-Danlos syndrome]

Authors:

J. Aragon-Martin[1,2,3], J. Kathirvelan[1], R. Yazdi[1], S. La Penna[1], H-Y. Yang[1], S. Srinivasan[1], C. Veshala[1], D. Morris-rosendahl[4], C. Nienaber[5], N. Naqvi[5], F. Dijk[5], N. Ghali[5], A. Reed[5], J. Pepper[5], J. Gaer[5], U. Rosendahl[5], J. Sneddon[6], E. Fisher[7], J. Dean[8], M. Jahangiri[9], E. R. Behr[9], M. Simpson[10], A. Smith[11], M. Y. B. Salmasi[4,2], N. Vadgama[12], J. Nasir[13], A. Child[2,14], M. Moosaje[3,15], T. Athanasiou[2,16], P. Deloukas[17]; [1]Queen Mary Univ. of London, London, United Kingdom, [2]Imperial Coll., London, United Kingdom, [3]Inst. of Ophthalmology, UCL, London, United Kingdom, [4]Royal Brompton Hosp., Guy's and St. Thomas' NHS Fndn. Trust, London, UK, London, United

Kingdom, [5]Royal Brompton Hosp., Guy's and St. Thomas' NHS Fndn. Trust, London, United Kingdom, [6]East Surrey Hosp., Redhill, Surrey, United Kingdom, [7]The Park Surgery, Albion Way, Horsham, Surrey, United Kingdom, [8]Clinical Genetics, NHS Grampian, Foresterhill, Aberdeen, United Kingdom, [9]St. George's, Univ. Hosp., NHS Fndn. Trust, London, United Kingdom, [10]King S Coll. London, London, United Kingdom, [11]Kings Coll. London, London, United Kingdom, [12]Stanford Univ., Palo Alto, CA, [13]Univ. of Northampton, Northampton, United Kingdom, [14]Marfan Trust, London, United Kingdom, [15]Moorfiels Eye Hosp. NHS Fndn. Trust, London, United Kingdom, [16]Imperial Coll. Hlth.Care NHS Trust, London, United Kingdom, [17]Queen Mary Univ. London, London, United Kingdom

Abstract:

Background: Hereditary Thoracic Aortic Aneurysm & Dissection (HTAAD) can arise abruptly leading to fatal consequences. About 20% of non-syndromic TAAD is inherited with variable penetrance and severity. Whereas, Ehlers-Danlos syndrome (EDS) is a heterogeneous genetic disorder characterized by musculoskeletal problems. hEDS is one of the 13 subtypes of EDS characterised by hypermobile joints & soft tissue, but without the complications of arterial dissection or bowel rupture seen in vascular EDS. This study - 2 HTAAD and 2 hEDS families. HTAAD_Family1: The proband had aortic dissection age 48. The family of the proband was also affected with 2 of his sisters & 1 daughter having upper borderline sinus of Valsalva measurements (4, 4.2 and 3.9cm respectively). HTAAD_Family2: The proband had emergency aortic root replacement at age 52; 2/3 of his sons (age 25 [4cm] and 28 [4.3cm] on losartan medication), & his cousin (37 [4.2cm]) have an aortic aneurysm. Negative results from an “aortopathy gene panel” in 2014. hEDS_Family1: A female aged 54 was presented with chest pain diagnosed as costochondritis. On examination she is mildly myopic, but with no lens or retinal problems, with normal palate and teeth, & with no scoliosis, pectus deformity or arachnodactyly. Her Beighton Score is 4/9 & she has 90° little fingers and thumb touching forearm bilaterally. Her sister, 2 maternal aunts, her daughter & her granddaughters were also affected. hEDS_Family2: The family phenotype includes hypermobility of joints with associated pain, dislocations, scoliosis, disc degeneration, ligament weakness, long flat feet, long fingers, osteoporosis. The proband, her mother & maternal grandmother similarly affected, as well as 9 first- & second-degree relatives (8F:1M).
Methods: From each family pedigree, the 2 most distantly related affected individuals were processed for WES or WGS. Alignment to the GRCh38, variant calling & annotation were performed. The variant list was filtered using our in-house pipeline; inclusion: all 5’UTR, 3’UTR, intronic & synonymous variants in coding genes, with MAF < 0.001, CAD score > 20, & segregated with the disease; exclusion: all homozygous variants.
Results: The final list contains 7 candidate genes: 1 synonymous [*MUC19*], 1 splice site [*RNF207*], & 5 intronic [*DPYD*, *EBF2*, *OTX1*, *TCF20* & *TRPM3*]. HTAAD: *RNF207* high expression in

the ascending aorta (Bgee). hEDS: *DPYD* high expression in synovial joint, calcaneal tendon & skin of hip (Bgee).
Conclusions: Confirmation is being sought by family segregation analyses which are being performed for these 7 candidate genes to confirm their association with HTAAD or hEDS phenotype in these 4 families.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2014T: *CASZ1* gene: From a promising GWAS risk locus to a possible new monogenic etiology for primary aldosteronism

Authors:

L. Santana[1], A. G. Guimaraes[1], F. Freitas-Castro[1], A. W. Maciel[1], G. F. Fagundes[1], M. A. Pereira[1], L. F. Drager[2], L. A. Bortolotto[2], M. B. Fragoso[3], B. B. Mendonca[1], A. Latronico[1], M. Q. Almeida[1]; [1]Hosp. das Clínicas, Faculdade de Med. da Univ.e de São Paulo, Sao Paulo, Brazil, [2]Inst. do Coração (InCor), Faculdade de Med. da Univ.e de São Paulo, Sao Paulo, Brazil, [3]Inst. do Câncer do Estado de São Paulo (ICESP), Faculdade de Med. da Univ.e de São Paulo, Sao Paulo, Brazil

Abstract:

Background/Objectives: Primary aldosteronism (PA) by bilateral adrenal hyperplasia or aldosteronomas is the most common cause of endocrine hypertension. Pathogenic germline variants in *KCNJ5*, *CACNA1D*, *CACNA1H*, *CLCN2*, *PDE2A*, and *PDE3B* genes have already been documented in individuals with PA. However, most of the cases still lack a defined genetic etiology. Genomic investigation of these individuals through whole-exome sequencing (WES) would enable the identification of a new monogenic association. Methods: In our study, twenty-nine probands were selected, with a confirmed diagnosis of bilateral PA by adrenal venous sampling. All probands had negative genetic screening for glucocorticoid-remediable and *KCNJ5* PA. WES was performed using DNA nanoball sequencing (DNBSEQ). Results: After quality control, SNV/CNV calling and gene/allele prioritization workflow, 4 heterozygous SNVs were listed as candidates in 3/29 patients (10.34%). One of them (hypertension diagnosed at 36 years of age) presented a variant in a possible new candidate gene, *CASZ1*. The NM_001079843.3(*CASZ1*):c.1912G>C/p.(Asp638His) variant its absence in population genomic databases (gnomAD v.4). The zinc finger transcription factor CASZ1, an emerging GWAS risk loci for PA, has *CACNA1D* as one of its targets and participates in key biological processes for PA pathogenesis, such as: co-repression of mineralocorticoid receptor transcriptional activity; mediation of tissue response to aldosterone and differential expression in aldosterone-

producing cell clusters (APCC) and zona glomerulosa. Conclusion: A rare *CASZ1* variant was identified in a woman with bilateral PA. The functional characterization of the p.(Asp638His) variant might reinforce the role of *CASZ1* as a new monogenic cause of PA. Grants: FAPESP #2021/10363-0 and #2019/15873-6.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2015T: Clinical and Genetic Spectrum in Chinese Hereditary Spastic Paraplegias patients and Identification of a new candidate gene for HSP

Authors:

Z-Y. Wu; Zhejiang Univ., Hangzhou, China

Abstract:

Background and objectives. Hereditary spastic paraplegias (HSPs) refer to a genetically and clinically heterogeneous group of neurodegenerative disorders. To date, considerable efforts have been made to unravel the genetic etiology of HSP, but the molecular diagnostic rate ranged between 29% to 61.8%. This study aimed to describe the relationship between genotype and phenotype in Chinese HSP patients and further investigate new causative genes for HSP. **Methods.** Totally 384 HSP pedigrees from all over China recruited from May 2018 to December 2023 were included in the study. The whole exon sequencing (WES) and multiplex ligation-dependent probe amplification (MLPA) were used to screen for pathogenic genes of HSP, combined with Sanger sequencing validation, co-segregation analysis and in vitro experiments, the pathogenicity of variants was determined according to the (ACMG) guidelines. **Results.** In our cohort, 24 pathogenic genes related to HSP were detected in 187 families, respectively, with a positive rate of 48.70%. Among which, 44.9% (84/187) patients carried *SPAST* gene variants. A total of 68 *SPAST* variants were identified, of which 13 were large fragment deletion and 1 was large fragment repetition. *SPAST* was the most common pathogenic gene of HSP. Fifteen patients carried variants in *SPG11* gene, the most common autosomal recessive inherited gene. In addition, we identified variants in *PCYT2, HPDL*, *RNF170* and *COQ4* genes for the first time. We performed WES screening in remaining families. After WES screening, the candidate gene *SREBF2* was detected in two HSP families. Two homozygous variants (c.1811T>G, p. L604W and c.2950A>G, p. T984A) were rare in public or in-house database and predicted to be damaging by prediction software. **Discussion and Conclusion.** Our findings extend the genotypes and phenotype spectrum of Chinese HSP patients and identify *SREBF2* as a new candidate gene for HSP.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2016T: Compound heterozygous*TXNDC15*variants associated with Joubert syndrome phenotype and forebrain malformation

Authors:

Y. Kuroda[1], K. Nagai[1], Y. Kawai[1], R. Tachioka[1], T. Nishikawa[1], T. Naruto[1], T. Ikegawa[1], N. Aida[1], K. Kurosawa[1,2]; [1]Kanagawa Children's Med. Ctr., Yokohama, Japan, [2]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan

Abstract:

TXNDC15 encodes thioredoxin domain-containing protein 15, protein disulfide isomerase containing a thioredoxin domain. *TXNDC15* plays a role in ciliogenesis.
Biallelic *TXNDC15* variants have been reported in at least six fetuses with Meckel syndrome. Here, we described a 1-year-old female patient with compound heterozygous *TXNDC15* variants demonstrating cerebellar vermis hypoplasia with molar tooth sign, mild holoprosencephaly (HPE), cortical dysgenesis, and developmental delay. The patient was a 1-year-old female, the first child of healthy nonconsanguineous parents. She was delivered at 39 weeks gestational age after an uneventful pregnancy, with a birth weight of 3,120 g (-0.1 SD), birth length of 51.0 cm (+0.8 SD), and an occipital-frontal head circumference (OFC) of 33.0 cm (-0.3 SD). She had strabismus and hypermetropia. Head MRI at 10 months revealed widespread cortical malformation /polymicrogyric cortex, dysplastic basal ganglia, callosal dysplasia, olfactory groove hypoplasia with partial cortical continuity of the frontal base, and cerebellar vermis hypoplasia with molar tooth sign. She controlled her head position at 5 months old but was not able to roll over at 1 year and 7 months. Developmental delay was noted, and she spoke no words at 1 years and 7 months old. As electroencephalogram revealed giant spikes and polyspikes of right frontal to parietal area at the age of 18 months, levetiracetam was administered. Compound heterozygous *TXNDC15* variants (NM_024715.4:c.485A>G p.(Glu162Gly) NC_000005.10:g.134888076A>G / NM_024715.4:c.786dup p.(Pro263Serfs*2) NC_000005.10:g.134896324dup) were identified by trio exome sequencing. No other candidate variants were detected in known ciliopathy genes. p.(Glu162Gly) variants were present in gnomAD (0.00000159).*TXNDC15* have been reported in six patients with MKS. This is the first report of *TXNDC15* variants associated with JS phenotypes such as molar tooth sign. Mild HPE is rare but distinctive features of MKS, so MRI finding had overlapping characteristics of JS and MKS. JS and MKS represent the spectrum disorders, and the causative genes overlap between two syndromes. In general, complete loss-of-function variants with severely affected protein function contribute to MKS phenotypes, and lesser deleterious variants lead to JS phenotypes. It

is possible that hypomorphic *TXNDC15* variants would cause JS phenotype.
Biallelic *TXNDC15* variants might be associated with distinctive phenotypes similar to JS.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2017T: Comprehensive assessment of recessive, pathogenic *AARS1* alleles in a humanized yeast model reveals loss-of-function and dominant-negative effects

Authors:

K. Jonatzke, M. Kuo, M. Parish, A. Antonellis; Univ. of Michigan, Ann Arbor, MI

Abstract:

Alanyl-tRNA synthetase 1 (*AARS1*) encodes the enzyme that ligates tRNA molecules to alanine in the cytoplasm, which is required for protein translation. Variants in *AARS1* have been implicated in later-onset dominant peripheral neuropathy and in early-onset, multi-system recessive phenotypes; to date, no single variant has been associated with both dominant and recessive disease. *AARS1* variants associated with recessive disease have been shown to cause a loss-of-function effect in various functional assays; however, pathogenic alleles have not been assessed in a single assay in a side-by-side manner, which will allow genotype-phenotype relationships. To address this, we employed a humanized yeast model to evaluate the functional consequences of all reported missense *AARS1* variants implicated in recessive disease. The majority of variants showed variable loss-of-function effects, ranging from no growth to normal growth. Missense variants associated with no growth were then evaluated for dominant-negative effects in an assay where wild-type human *AARS1* is also expressed. Interestingly, K81T *AARS1*, a variant implicated in recessive disease, demonstrated both loss-of-function and dominant-negative effects, indicating that certain *AARS1* variants may be capable of causing both dominant and recessive disease.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2018T: Comprehensive Genetic Analysis of the Mexican Familial Hypercholesterolemia Registry Participants

Authors:

P. Salazar-Fernandez[1], **C. D. Quinto-Cortés**[1], N. Mourtzi[1], S. Moreno-Grau[1], D. M. Montserrat[1], M. Barrabes[1], D. Bonet[1], A. Lopez-Pineda[1], J. Wall[1], A. G. Ioannidis[1], C. Bustamante[1], N. Katsanis[1], A. J. Martagon[2], R. Mehta[2], G. A. Galan-Ramirez[2], R. Zubirán[2], D. Elias-Lopez[2], FH Mexican Registry Group, T. Tusié-Luna[2], C. A. Aguilar-Salinas[2]; [1]Galatea Bio Inc., Miami, FL, [2]Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico

Abstract:

Familial hypercholesterolemia (FH) is a genetic disorder influenced by three key genes related to low-density lipoproteins (LDL): LDLR, APOB, and PCSK9. Leveraging the Mexican FH registry (www.fhmexico.org.mx), established in 2017, we analyzed 302 samples from 169 families using whole exome sequencing (WES), whole genome sequencing, and a custom SNP array to capture both common and rare genetic variants associated with FH.We identified driver mutations in 144 FH patients from 68 families, with the majority (125) attributed to LDLR; considering an initial clinical diagnosis of FH, our diagnostic yield exceeds 70% and suggest that such an approach has significant clinical utility. Notably, ~ 8% of the diagnoses were driven by copy number variants, including deletions in APOB and LDLR and a whole gene duplication of PCSK9. Moreover, expanding the range of analysis to other genes in the LDL-cholesterol pathway, we identified one family with a multi-exon deletion of APOE. Third, we were intrigued to observe ten individuals from eight different families with multiple mutations in the three major LDL genes, including an individual with two mutations in LDLR in addition to a known pathogenic APOB mutation that track with more severe disease expressivity. Finally, ancestry-adjusted polygenic risk score analyses identified a strong correlation between genetic factors and reported diagnoses of coronary artery disease, hypertension, LDL, HDL, fasting glucose (FG), and Type 2 Diabetes, with a significantly higher number of individuals exhibiting elevated LDL and HDL levels compared to reference populations.Overall, our analysis has illuminated a contribution of common and rare variants to FH that can serve as a useful model for understanding the role different classes of deleterious genetic variation can play to disease manifestation and emphasize the value of integrating whole exome sequencing and polygenic risk scores to better understand and manage FH.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2019T: *CSTB* null disease demonstrates the exquisite dosage sensitivity of cystatin B in humans

Authors:

D. Calame[1], A. Rogers[2], A. Jolly[1], N. Gogate[1], M. Verma[3], K. Helbig[4], C. Skraban[5], I. Helbig[6], L. Emrick[7], M. Zaki[8], H. Shamseldin[9], M. Hashem[10], A. Peron[11], A. Gilberti[12], S. Malone[13], C. Patel[14], A. Hahn[15], P. J. Morrison[16], S. Cabet[17], N. Chatron[18], V. Desportes[17], G. Lesca[17], A. Smith[19], A. Pagnamenta[20], R. Maroofian[21], T. Pearson[22], J. Posey[1], J. Lupski[1], M. Mcnutt[23]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Louisville, Louisville, KY, [3]Univ. of Texas Southwestern, Dallas, TX, [4]Children's Hosp. of Philadelpia, Philadelphia, PA, [5]The Children S Hosp. of Philadelphia, Philadelphia, PA, [6]Children s Hosp. of Philadelphia, Philadelphia, PA, [7]Baylor Coll. Med., Houston, TX, [8]Natl. Res. Ctr., Cairo, Egypt, [9]King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [10]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [11]Università degli Studi di Firenze, Besnate, Italy, [12]Meyer Children's Hosp., Florence, Italy, [13]Lady Cilento Children's Hosp., Brisbane, Australia, [14]Royal Brisbane & Women's Hosp., Brisbane, Australia, [15]Univ. Hosp. Gießen, Gießen, Germany, [16]Queen's Univ. Belfast, Belfast, United Kingdom, [17]CHU de Lyon, Lyon, France, [18]Service de Génétique, Hospices Civils de Lyon, Bron Cedex, France, [19]Manchester Univ., Manchester, United Kingdom, [20]Univ. of Oxford, Oxford, United Kingdom, [21]Univ. Coll. London, London, United Kingdom, [22]Nationwide Children's Hosp., Columbus, OH, [23]UT Southwestern Med. Ctr., Dallas, TX

Abstract:

Background: Unverricht-Lundborg disease (ULD) is the most common progressive myoclonic epilepsy. Individuals with ULD have normal development, myoclonic epilepsy beginning in late childhood, and mild to no cognitive impairment. ULD results from biallelic genetic variants in the cystatin B gene *CSTB*. Nearly all individuals with ULD have biallelic dodecamer repeat expansions in the *CSTB* promotor which reduce CSTB mRNA expression to 5-10% of normal. Rare compound heterozygotes with dodecamer repeat expansions *in trans* with a predicted loss-of-function (LOF) or missense variant have earlier disease onset and more refractory epilepsy. Additionally, severe neurodevelopmental disorders (NDD) have been described in two families with homozygous *CSTB* null alleles.

Methods: Individuals with biallelic *CSTB* variants (LOF and missense) and severe NDD were identified through clinical and research lab databases and clinical practice. Deep phenotypic analysis of clinical data was performed.

Results: The cohort consists of 25 individuals from 20 unrelated families. 14 families had biallelic LOF variants, 3 had LOF and missense variants *in trans*, and 3 had biallelic missense variants. The median age at last assessment was 6 y (range 18 m-15 y). 6 individuals are deceased. Microcephaly is nearly uniformly seen (median Z-score -5.7) and can be congenital or acquired. Most individuals do not achieve any developmental milestones. 38% could roll over, and 12% were able to crawl or walk with assistance. These developmental milestones were often lost over time. Movement disorders including spasticity, dystonia, chorea, and myoclonus were common. Chorea was described as near continuous, began in first years of life, and

decreased over time. Epilepsy occurred in 68% and was refractory in 29%. Brain MRI demonstrates reduced brain volume, delayed myelination, and hypomyelination. Repeat MRIs showed severe progressive brain atrophy. Quantitative phenotypic analysis using Human Phenotype Ontology (HPO) did not identify genotype-phenotype correlations, but pairwise comparisons with all OMIM disease genes showed *CSTB* null disease most closely resembles developmental and epileptic encephalopathies (DEE).
Conclusions: Our data firmly establishes *CSTB* null disease as a profound NDD distinct from ULD. The *CSTB* allelic series demonstrates exquisite CSTB dosage sensitivity with therapeutic implications especially considering recent pre-clinical AAV9 gene therapy studies.
While *CSTB* null disease shows considerable overlap with DEEs, the data demonstrate *CSTB* null disease is a developmental neurodegenerative process beginning *in utero*.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2020T: *De novo* variants in *TMEM184B* cause a neurodevelopmental syndrome underscored metabolic signaling deficits ⭐

Authors:

F. Ullah[1], K. Chapman[2], Z. Yahiku[3], S. M. Kodiparthi[4], G. Kellaris[5], T. Stödberg[6], A. Accogli[7], V. Salpietro[8], P. Striano[9,10], C. Reuter[11], S. Berger[12], P. Kelvin[13], S. Sirimulla[14], E. E. Davis[15], M. Bhattacharya[16]; [1]Lurie Children's Hosp., Chicago, IL, [2]Children's Natl. Rare Disease Inst., Bethesda, MD, [3]Dept. of NeuroSci., Univ. of Arizona, Arizona, AZ, [4]AIQure LLC, El Paso, TX, USA, Texas, TX, [5]Northewestern, Chicago, IL, [6]Karolinska Inst., Stockholm, Sweden., Stockholm, Sweden, [7]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada., Montreal, QC, Canada, [8]UCL, London, United Kingdom, [9]Dept. of NeuroSci.s, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Hlth.(DINOGMI), Univ. of Genoa, Italy., Genoa, Italy, [10]IRCCS Giannina Gaslini Inst., Genoa, Italy, [11]Stanford, San Francisco, CA, [12]Childrens Natl. Res. Inst., Washington, DC, [13]Dept. of Cellular and Molecular Med., Univ. of Arizona Coll. of Med. - Tucson, Arizona, AZ, [14]AIQure LLC, El Paso, TX, USA., Texas, TX, [15]Lurie Children`s Hosp. of Chicago, Chicago, IL, [16]Dept. of NeuroSci., Univ. of Arizona, Tucson, AZ

Abstract:

Transmembrane protein 184B (TMEM184B) is an endosomal 7-pass transmembrane protein with conserved roles in synaptic structure and axon degeneration. However, the specific contributions of *TMEM184B* to neurodevelopment remain unclear, and pathogenic variants have not previously been linked to human diseases. We identified six unrelated pediatric cases

who harbor rare *de novo* heterozygous variants in *TMEM184B* (1 splice site variant and 5 missense variants). Affected individuals display neurodevelopmental manifestations including intellectual disability, corpus callosum hypoplasia, seizures, and microcephaly with concomitant short stature. TMEM184B is predicted to contain a core domain, and *in silico* studies map all variants around this domain. To test variant pathogenicity and the nature of variant effect, we used zebrafish to test hypotheses of dominant toxicity or haploinsufficiency. First, we performed ectopic expression of human *TMEM184B* mRNA harboring each variant, followed by phenotyping of head size as a proxy for neurodevelopment. We show that two of five variants tested, p.Lys184Glu and p.Gly162Arg, induce dominant toxic effects while the other three variants induced no differences. Next, we tested the other three variants using an *in vivo* complementation assay; we established that knockdown of the zebrafish ortholog results in dose-dependent microcephaly and shortened body length, which could be rescued by introduction of human WT *TMEM184B* mRNA. We compared rescue efficiency of p.Val88Met, p.Leu127Arg and p.Gly288Ala to WT and found inability to restore head size phenotype suggesting that these alleles result in loss of protein function and a haploinsufficiency basis for disease. To characterize further the consequences of *tmem184b* depletion we painted wholemount larvae with acetylated tubulin antibody to visualize the central nervous system; we observed reduced numbers of intertectal axons with reduction in optic tecta size. To study the cellular impact of TMEM184B variants, we monitored localization of TFEB (transcription factor EB), a regulator involved in endosomal transport. Ectopic expression of disease-associated variants in HEK293T cells caused aberrant nuclear localization of TFEB, suggestive of disrupted nutrient signaling pathways. This defect was the most pronounced for p.Lys184Glu, and resulted in significantly increased apoptosis *in vitro*, which may be linked to microcephaly in humans. Together, our clinical, genetic, and functional studies suggest that *de novo* variants in *TMEM184B* impair protein function through diverse molecular effects but converge upon overlapping neurodevelopmental phenotypes.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2021T: Detection of deletions as second variants by whole-genome sequence analysis overlooked by whole-exome sequence analysis in autosomal recessive disorders

Authors:

A. Mitsutake[1,2], H. Naruse[2], K. Orimo[2], J. Mitsui[2], T. Matsukawa[2], H. Ishiura[3,2], Y. Shiio[4], T. Nishizawa[1], N. K. Iwata[1], S. Tsuji[1], T. Toda[5]; [1]Intl. Univ. of Hlth.and Welfare, Tokyo, Japan, [2]The

Univ. of Tokyo, Tokyo, Japan, [3]Okayama Univ., Okayama, Japan, [4]Tokyo Teishin Hosp., Tokyo, Japan, [5]The Univ. of Tokyo, Bunkyo-ku, Japan

Abstract:

In certain instances of autosomal recessive genetic disorders, despite exhibiting typical clinical presentations, only one heterozygous variant is identified in genes responsible for the condition. This study aimed to identify the second variant using whole-exome sequence (WES) and whole-genome sequence (WGS) data. Patient 1 is a 36-year-old man who developed disturbance of consciousness. Hyperammonemia and high level of citrulline indicated the diagnosis of adult-onset type II citrullinemia. WES analysis showed a heterozygous splice site variant (c.1311+1G>A) in *SLC25A13* (NM_014251.3), previously reported as pathogenic. Patient 2 is a 40-year-old woman who developed tremor and dystonia at the age of 32. Brain MRI at the age of 40 showed hypomyelinating leukodystrophy. WES analysis was performed to search for variants in genes related to hypomyelinating leukodystrophy. A heterozygous missense variant c.3430C>G, p.Leu1144Val in *EPRS1* (NM_004446.3) encoding an aminoacyl-tRNA synthetase, was detected. This variant has not been reported and is not registered in gnomAD or jMorp. We conducted WGS analysis in these patients to search for second variants. Manta analysis using WGS data revealed a deletion in patient 1 (chr7:96,126,323-96,131,989; 5,667 bp, GRCh38/hg38) involving the entire exon 6 of *SLC25A13*. In patient 2, a deletion was found (chr1:220,006,124-220,006,343; 220 bp) involving part of exon 15 in *EPRS1*. Compound heterozygosity in patient 2 was confirmed by analyzing parental DNA samples. To retrospectively verify if this variant could be detected using WES data, the mean exon coverage was calculated for each patient. The ratio of exon coverage in the patient compared to other patients on the same slide was 0.34 for exon 15 and 0.37 for exon 1 in *SLC25A13* in patient 1, and 0.53 for exon 15 in *EPRS1* in patient 2. Thus, it was hard to detect the deletion in *SLC25A13*. There are still difficulties in detecting single exon deletions using WES. Reanalyzing with WGS can improve diagnostic accuracy.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2022T: Developing a pipeline to identify rare and novel pathogenic variants for inborn errors of metabolism

Authors:

A. Westfall[1], C. Xing[2], R. Deberardinis[3]; [1]UT Southwestern, Dallas, TX, [2]Univ Texas SW Med Ctr, Dallas, TX, [3]UT Southwestern Med Ctr, Dallas, TX

Abstract:

The Genetic and Metabolic Disease Program (GMDP) seeks to diagnose and treat children with genetically determined metabolic diseases through the integration of clinical expertise and cutting-edge research. These diseases, known as inborn errors of metabolism (IEM), arise from mutations to the genes encoding metabolic enzymes and transporters that cause wide ranging symptoms, including hypoglycemia, acidosis, and developmental delays. If not identified and treated at a young age, IEMs can cause death or permanent disability. However, many IEMS are treatable once the specific genetic basis of the disease has been identified. Whole exome sequencing was performed patients with IEMs and their families from US (n = 278) and Pakistani (n = 135) cohorts to determine potentially pathogenic variants. We constructed a bioinformatics pipeline to that could identify novel disease-causing variants in genes through annotation of regions of homozygosity and large structural variants, such as exon deletions, copy number variation, and inversions. We placed emphasis on integrating improved approaches to identify autosomal dominant, compound heterozygous, and structural variants to expand from an existing focus on primarily recessive, homozygous point mutations. This expanded pipeline has successfully identified known pathogenic variants within each cohort, including an known multi-exon deletion in CYP21A2 causal for congenital adrenal hyperplasia that had been otherwise undetected and a case of maternal uniparental disomy 7 that caused homozygosity for a rare deletion with an atypical presentation. We are now exploring options to expand variant filtering and prioritization to whole genome sequencing and additional rare disease datasets. Leveraging an extensive bioinformatics pipeline to assess whole exome sequencing data - and extending this pipeline to whole genome sequencing - at a population scale provides unprecedented insights into the genetic underpinnings of IEMs, enabling novel diagnostic and treatment options to improve care and quality of life for hundreds of children.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2023T: Diagnosing the undiagnosed: Lessons from a large heterogeneous cohort of patients with Cornelia deLange and related diagnoses

Authors:

R. Rajagopalan[1,2], M. Kaur[3,4], S. Raible[3,4], J. Blair[3,4], I. Krantz[5,6,4]; [1]Div. of Genomic Diagnostics, Children s Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Pathology and Lab Med., Perelman Sch. of Med., University of Pennsylvania, PA, [3]Div. of Human Genetics, Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Roberts Individualized Med. Genetics Ctr., Children's Hospital of Philadelphia, PA, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Div. of Human Genetics, University of Pennsylvania, PA

Abstract:

Cornelia de Lange Syndrome (CdLS) is a multisystem developmental disorder caused primarily by pathogenic variants in structural and regulatory proteins associated with the cohesin complex, a critical regulator of gene expression during development. A diagnosis of CdLS can be confirmed by molecular testing of the 5 known CdLS genes (including *NIPBL, SMC1A, SMC3, RAD21* and *HDAC8*). Several phenocopy syndromes to CdLS exist, and a genome-wide, multi-omic approach is required to identify molecular etiologies of such phenotypes. In a cohort of 716 CdLS and CdLS-like probands we were able to confirm a molecular etiology in 422 (59%) using standard of care molecular diagnostic approaches such as targeted panels, exome sequencing, and exome slice on a genome platform. 358 of these (85%) had a mutation in a known CdLS gene (271 *NIPBL* (64%), 40 *SMC1A* (9%), 25 *HDAC8* (6%), 16 *SMC3* (4%), 6 *RAD21* (1%). In addition, 64 probands (15%) had mutations in 32 different genes that phenocopied CdLS. In 294 (41%) probands a definitive molecular etiology was not identified. Expanding the genome sequencing analysis workflow to include copy-number and structural variation and using transcriptome sequencing as a complementary assay (when available) in 104 of the previously negative probands revealed molecular etiologies in an additional 41 probands (39%). These were enriched for variants beyond standard exonic SNV/ indels including cryptic non-coding variants (promoter, deep intornic, large insertion), copy number variation including mosaic events, chromosomal rearrangements including inversions and balanced translocations, and variants in additional genes that phenocopy CdLS. Several of the cryptic noncoding variants did not have the computational evidence for a splicing abnormality (SpliceAI < 0.2) and yet produced aberrantly spliced mRNA. This has resulted in an overall diagnostic yield of 74% of probands with a molecular diagnosis explaining their clinical phenotype. For those who underwent genome sequencing, a cohort enriched for molecularly uncharacterized probands using standard screening methods, a diagnostic yield of 59% was obtained and was enriched for CdLS phenocopy diagnoses. Our results underscore the need for the whole genome (and transcriptome) sequencing and a comprehensive, phenotype agnostic, unbiased analytical protocol including structural and noncoding variants in order to maximally mine a phenotypically and genetically heterogeneous cohort to maximize the diagnostic yield and provide families with accurate prognostic information and counseling.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2024T: Discovery of cryptic splice variants prompts recalculation of Van der Woude syndrome and popliteal pterygium syndrome mutation rates

Authors:

K. Singleton[1], K. Robinson[2], B. Carter[1], J. C. Murray[3], E. J. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Emory Univ., Decatur, GA, [3]U of Iowa, Iowa City, IA

Abstract:

Van der Woude syndrome (VWS) is a rare single gene syndrome that often causes orofacial clefting (OFC). The distinguishing feature of VWS from non-syndromic OFCs is the presence of lower lip pits or mounds, a phenomenon virtually exclusive to VWS. Popliteal pterygium syndrome (PPS) is closely related to VWS as they both can be caused by mutations in *interferon regulatory factor 6* (*IRF6*). PPS differs from VWS however, with a primary phenotypic feature being pterygium, or webbing across different joints. Previous estimates in the literature report *IRF6* mutations to account for roughly 70% of VWS cases and nearly all (~97%) cases of PPS. Additionally, *Grainyhead-like 3* (*GRHL3*) mutations are estimated to account for 5% of VWS cases. Cumulatively, this suggests that 25% of VWS and about 3% of PPS cases are unsolved. To address this gap, we analyzed whole genome sequence data for causal variants in 21 individuals/families with VWS or PPS with no mutations in *IRF6* or *GRHL3* found by Sanger sequencing. We identified causal variants for 5 individuals, 3 of which were cryptic splice variants in *IRF6* or *GRHL3*. One individual diagnosed with PPS had the hotspot *IRF6* missense variant (R84C) commonly found in PPS that was missed by prior Sanger sequencing. Another individual harbored a heterozygous splice variant in *FBN2*, which is associated with congenital contractural arachnodactyly, a syndrome which could resemble and be mistaken for PPS. We then revisited our collection of 194 VWS and PPS families to recalculate the fraction of each condition attributable to various genes. Out of 166 VWS families, 138 (83%) had *IRF6* mutations of any kind, including cryptic splice (N=2) and enhancer mutations (N=1). Only 3.6% (6) had *GRHL3* mutations, including one cryptic splice variant (NM_198173.3:c.738C>T:p.Gly246Gly) with a donor gain SpliceAI score of 0.69. The 22 unsolved VWS cases account for 14% of the cohort. Out of 28 PPS cases, the majority had variants in *IRF6* (N=23, 82%), while the rest had mutations in *RIPK4* (N = 2, 7.1%), *FBN2* (N=1, 3.5%), and *GRHL3* (N=1, 3.5%), leaving a single unsolved case. Notably the same *GRHL3* cryptic splice variant was found in VWS and PPS; to our knowledge PPS has not previously been associated with *GRHL3* variation. In this dataset, cryptic splice variants accounted for 2% of all individuals with VWS or PPS, suggesting that reanalysis of prior genetic testing data may be warranted.

These updated statistics can assist clinicians in diagnostics for both new and unsolved VWS and PPS families.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2025T: Distinct founder haplotypes across global *CHCHD10* p.Arg15Leu amyotrophic lateral sclerosis patients

Authors:

K. L. Williams[1], L. Henden[1], A. N. Smith[1], N. Grima[1], Z. Zussa[1], S. Chan Moi Fat[1], R. Fell[1,2], D. Southwood[1], D. B. Rowe[1,2], M. C. Kiernan[3], E. Rogaeva[4], D. Kurzwelly[5], P. Weydt[5], Project MinE ALS Sequencing Consortium; [1]Macquarie Univ. Motor Neuron Disease Res. Ctr., Faculty of Med., Hlth.and Human Sci., Macquarie Univ., Sydney, Australia, [2]MQ Hlth.Neurology, Macquarie Univ., Sydney, Australia, [3]NeuroSci. Res. Australia, Sydney, Australia, [4]Tanz Ctr. for Res. in Neurodegenerative Diseases, Univ. of Toronto, Toronto, ON, Canada, [5]Neurocenter, Univ.Hosp. Bonn, Bonn, Germany

Abstract:

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease leading to muscle weakness, paralysis, and death, on average, 27 months after symptom onset. 140,000 new cases of ALS are diagnosed globally each year. ALS is genetically heterogenous, and heritability evidence suggests there is a genetic component underlying both inherited and sporadic ALS, with more than 30 genes implicated to date. Disease-causing sequence variants in the *CHCHD10* gene [HGNC:15559] were first described in familial (hereditary) ALS cases in 2014. Additional *CHCHD10* variants have been subsequently identified either segregating in ALS families, or present in individual probands. There has been conflicting evidence of pathogenicity of variants in this gene, with a ClinGen Gene-Disease Validity classification of Moderate (ALS GCEP, SOP9, 13/09/2022).

Recently, we identified a new Australian ALS family with a segregating *CHCHD10* c.44G>T (p.Arg15Leu) variant in first cousins (3rd degree relatives). This variant has been previously reported in 20 ALS patients globally (both familial and sporadic ALS cases) and is absent from >1,000,000 alleles in gnomAD v4. In ALS, this specific variant can show reduced penetrance and an atypical ALS phenotype, with slow disease progression and predominant lower motor neuron involvement. To investigate whether the two Australian *CHCHD10* p.Arg15Leu cases were a distant branch of an already reported pedigree, we obtained genome-wide genotype data (either SNP microarray or WGS) from 9 global ALS patients with this mutation (Australia n=2; Germany n=2; Canada n=1; USA n=2; Netherlands n=1; United Kingdom n=1). Population

stratification using PLINK determined all samples were of European ancestry.
Using XIBD software on unphased genotype data from all samples, we performed pairwise identity-by-descent (IBD) analysis to identify IBD segments >2cM (representative of shared ancestry up to 25 generations). IBD segments overlapping the *CHCHD10* gene on chromosome 22 were extracted to investigate haplotype sharing between cases. Five pairs of samples were identical-by-descent over the *CHCHD10* locus in two distinct clusters, indicative of two founder haplotypes. Cluster 1 contained 4 individuals (USA, USA, Germany, Germany) sharing a 5.2Mb haplotype while cluster 2 contained the two Australians with a 9.2Mb shared haplotype.
We are currently extracting the phased founder haplotypes to perform mutation dating, and to determine if the remaining 3 cases with the same genetic variant (from Canada, UK, and the Netherlands) share a small portion of one of the two founder haplotypes, suggesting a more distant genealogical relationship.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2026T: Driver gene ontology analysis highlights key pathways for male infertility

Authors:

M. Yang[1,2], K. Lovrenert[1,2], N. Thirumavalavan[2], F. Schumacher[3], C-H. W. Wu[1,2]; [1]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. Sch. of Med. and Univ. Hosp., Cleveland, OH, [2]Dept. of Urology, Case Western Reserve Univ. Sch. of Med. and Univ. Hosp., Cleveland, OH, [3]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Abstract:

Male infertility is a heterogeneous condition that affects up to 10% of the male population. An increase in the number of next-generation sequencing studies in male infertility has revealed monogenic causes, yet a majority of infertile men (70%) remain without a clear genetic diagnosis. We performed a literature review which revealed 534 genes with at least preliminary evidence as a cause of male infertility. 161 genes were identified as having strong evidence (determined by consistency of results in sequencing studies and functional validation), representing a 54% increase compared to the 104 genes identified in 2020.
Traditional categorization into pre-testicular/endocrine, testicular/spermatogenesis, and post-testicular/obstructive groups provides a broad overview of the underlying etiologies of male infertility. Our study aims to delve into biological pathways of potential mechanisms of male infertility and enhance care of these patients.

We performed gene ontology (GO) over-representation analysis of the 161 high-evidence genes to reveal driver GOs that represent key etiologies of male infertility. All three GO databases, Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) were examined. In total, 25 GO nodes were identified as drivers, of which 14 were derived from BP, 7 from CC, and 4 from MF.
Although hypogonadotropic hypogonadism has been widely studied as a pre-testicular/endocrine cause of male infertility, we indicate that steroid biosynthesis is a predominant driver in BP (p-adj 6.2E-11). This is also evidenced by the independent over-representation of steroid hydroxylase activity (p-adj 9.9E-4) as a driver in MF.
At the testicular/spermatogenesis level, we revealed the following over-represented drivers in BP: cilium movement (p-adj 7.2E-52), axoneme assembly (p-adj 6.4E-49), and left/right pattern formation (p-adj 1.7E-16). This was supported by the independent identification of these drivers in MF: minus-end-directed microtubule motor activity (p-adj 1.6E-8) and dynein intermediate chain binding (p-adj 4.9E-9). And likewise from CC: cilium (p-adj 8.1E-45), and dynein axonemal particle (p-adj 2.3E-10). This underscores the importance of the sperm tail and centrosome for fertility, but interestingly not the acrosome.
No driver GO was significantly overrepresented among the only 2 genes in the post-testicular/obstructive category for male infertility.
As revealed, ciliary function and steroid metabolism are significant drivers. They involve many organ systems and this highlights the importance of multidisciplinary management in care of patients with male infertility.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2027T: Early Diagnosis and Optimized Care of Vascular Ehlers-Danlos Syndrome Through AI-Powered Facial Analysis: Results from the Montalcino Aortic Consortium

Authors:

D. Murdock[1], A. Suresh[1], E. Calderon[1], I. Marin[1], F. Marin[1], A. Yetman[2], S. Morris[3], D. Milewicz[4]; [1]UTHlth.Houston, Houston, TX, [2]Children's Hosp. & Med. Ctr. of Omaha, Omaha, NE, [3]Texas Children's Hosp. & Baylor Coll. of Med., Houston, TX, [4]Univ Texas McGovern Med Sch, Houston, TX

Abstract:

Background: Vascular Ehlers-Danlos syndrome (vEDS), caused by *COL3A1* pathogenic variants, is a rare heritable aortic and arterial disorder associated with early mortality, mainly due to

spontaneous arterial and aortic dissections and ruptures. Despite continued improvements in management, 70% of affected individuals are diagnosed only after a vascular event or death. Additionally, vEDS is often conflated with hypermobile EDS (hEDS), a more common condition that lacks vascular complications. New and rapid screening methods for vEDS are thus essential to prevent misdiagnoses. Here, we describe an artificial intelligence (AI)-based approach to screen for vEDS utilizing facial feature analysis, aiming for high-fidelity early diagnosis.
Methods: Using the PhenoScore framework, we implemented an AI facial recognition model using a support vector machine (SVM) trained on frontal facial images of vEDS patients from the Montalcino Aortic Consortium (MAC), control images from the Chicago Face Database, and publicly available hEDS images. Cross-validation was used in training the SVM model, and statistical measures, including the Brier score, area under the receiver operating characteristic curve (AUC), and corresponding P values, were calculated. Local Interpretable Model-agnostic Explanations (LIME) was used to generate facial heatmaps highlighting the features driving the model's predictions.
Results: Overall, the AI classifier showed excellent performance in distinguishing vEDS from age-, sex-, and ethnicity-matched controls (Brier = 0.11; AUC = 0.94; P = 9.90e-03) and hEDS patients (Brier = 0.075; AUC = 0.98; P = 9.99e-04). Notably, no statistically significant difference was observed between the control and hEDS groups (Brier = 0.27; AUC = 0.46; P = 0.89), suggesting that hEDS patients have similar facial features to the general population. Additionally, there was no significant difference when comparing variant types (glycine substitutions vs. null variants) within *COL3A1* (Brier = 0.25; AUC = 0.47; P = 0.36). The model's performance improved with a larger training set but remained strong with as few as eleven vEDS images used for training (Brier = 0.16; AUC = 0.84; P = 2.63e-04). LIME confirmed the characteristic features, such as prominent eyes, that distinguish vEDS from controls and hEDS.
Conclusion: This study demonstrates that an AI-based facial analysis approach can potentially diagnose vEDS and differentiate such cases from hEDS. Thus, these studies have the potential to develop an inclusive and generalizable tool that ensures equitable and accurate screening for vEDS across all ages, ethnicities, and sexes.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2028T: Efficacy of whole genome sequencing in the diagnosis of rare neurodevelopmental disorders: experience from a tertiary health care center and the Danish National Genome Center

Authors:

B. Schantz-Conlon[1], B. Bertelsen[1], A. Hansen[1], Z. Lohse[2], A. Rasmussen[1], F. Skovby[2], L. L. Borgwardt[1]; [1]Univ. Hosp. of Copenhagen, RigsHosp.et, Copenhagen, Denmark, [2]Univ. Hosp. of Zealand, Roskilde, Roskilde, Denmark

Abstract:

Neurodevelopmental disorders are a heterogenic group of diseases, and their diagnosis and treatment represent a global healthcare challenge. While phenotypic presentation is vital for an initial diagnosis, the range of potential disorders and phenotypes make it challenging to provide a precise diagnosis, prognosis and estimate of recurrence risk. Comprehensive genetic analysis has become an essential part of diagnostics for rare neurodevelopmental disorders. Implementation of the Danish National Genome Center, aiming to sequence 60,000 genomes (~1% of the population) in a five-year period, has expanded the availability of whole genome sequencing (WGS) in Denmark for several disease groups, including neurodevelopmental disorders. Until now, the evidence for the diagnostic yield of next generation sequencing (NGS) in this patient group has primarily been limited to experiences from studies using whole exome sequencing and targeted NGS gene panels. Reports of the diagnostic yield of WGS in this patient cohort are therefore still to be fully explored, thus making our analysis of WGS-based diagnostic results extremely relevant. For 5.5 years (2018-2024) WGS analyses were conducted at a tertiary healthcare institution in Copenhagen, Denmark, as part of the diagnostic procedure. WGS-based analyses were performed using a standardized workflow with only minor optimizing adjustments in laboratory protocol, bioinformatic pipeline and variant interpretation strategy. A total of 684 individuals comprising 250 pediatric and adult patients with a neurodevelopmental disorder and 434 family members were included in our cohort. Of these patients, 10% were analyzed in single-case analyses and 90% in family analyses. Preliminary results from our cohort reveal a diagnostic yield of 51.6%, covering both causal variants and strong candidate variants. Further results from our cohort will be presented at the ASHG Annual meeting 2024, including genetic findings, diagnostic yield in subgroups of neurodevelopmental disorders and description of the clinical consequences of systematic WGS. These clinical outcomes include diagnostic clarification, changes in treatment or follow up, family, or prenatal testing, all of which are crucial for the optimal diagnostic strategy for the benefit of patients. Our results highlight the importance and value of WGS as a diagnostic tool for neurodevelopmental disorders. Furthermore, by combining phenotypic data with genotypic data, our results broaden the insight into new and known genotype-phenotype correlations.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2029T: Elevated Iron Levels In Tears of Patients Diagnosed With WDR45 X-Linked Optic Atrophy

Authors:

M. Michelson[1,2], A. Zahavi[3,4], I. Maharshak[5], L. Blumkin[6], K. Yosovich[2], D. Lev[7], N. Goldenberg-Cohen[8]; [1]Wolfson Med. center, Holon, Israel, [2]Genetic Inst., Wolfson Med. Ctr., Holon, Israel, [3]Depatment of Ophthalmology and Lab. of Eye Res., Felsenstein Med. Res. Ctr., Rabin Med. Ctr., Petach Tikva,, Israel, [4]Depatment of Ophthalmology and Lab. of Eye Res., Felsenstein Med. Res. Ctr., Rabin Med. Ctr., Petah Tikva, Israel, [5]Dept. of Ophthalmology, Wolfson Med. Ctr., Holon, Israel, [6]Pediatric Neurology Clinic, Wolfson Med. center, Holon, Israel, [7]Genetic Inst., Wolfson Med. Ctr., Holon, Israel, [8]Ophthalmology Dept., Bnai-Zion Med. Ctr., Haifa, Israel

Abstract:

Background: Pathogenic variants in the X-kinked *WDR45* gene have been related to diverse neurodevelopmental disorders and include beta-propeller protein-associated neurodegeneration (BPAN), intellectual disability, early-onset epileptic and developmental encephalopathy, and West syndrome. The disorders are most commonly reported in females. Clinically, *WDR45*-related phenotypes encompass a broad spectrum ranging from early-onset developmental delay and intellectual disability to neurodegeneration and multiple epileptic syndromes. The hallmark sign is the presence of iron deposits in the basal ganglia, a feature observed in the majority of patients within this spectrum. Here we describe two brothers presenting with optic atrophy with a novel pathogenic variant in the *WDR45:*c.238G>T, p.Val80Leu, identified by Whole exome sequencing and confirmed by Sanger sequencing. The RNA studies have shown expression differences in *WDR45* transcripts, however, they did not precisely clarify the underlying mechanisms responsible for this unique presentation. This study aimed to determine iron levels in tears of patients with *WDR45* X-linked optic atrophy compared to healthy controls. Tear samples were analyzed for trace elements using a particle-induced X-ray emission (PIXE) technique. Concentrations were calculated as the difference from unused Schirmer strips. Findings were compared with 40 healthy subjects, 33 female and 7 male, aged 17-63 (mean 31.44 years). **Results:** High iron levels were measured in the tears of the two brothers with *WDR45* X-linked optic atrophy, compared to the control group. **Conclusions:** This is the first study to show higher-than-normal iron levels in the tears of patients with X-linked optic atrophy due to a variant in *WDR45.* The findings broaden our understanding of the role of the *WDR45* gene in disorders related to abnormal iron pathophysiology. PIXE analysis is a susceptible technique for measuring trace elements in tears. Further clinical studies are needed to investigate potentially novel genotype-phenotype correlations. The application of these findings in clinical practice may benefit patient care.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2030T: Elucidating the biophysical properties of TANGO2 protein; an essential step towards future therapies for TANGO2 deficiency disorder

Authors:

L. Ghaloul Gonzalez[1], S. Lovell[2], A. Powers[1], A. Cooper[2], K. Battaile[3], J. Vockley[4], A-W. Mohsen[5]; [1]UPMC Children's Hosppital of Pittsburgh, Pittsburgh, PA, [2]Protein Structure and X-ray Crystallography Lab., The Univ. of Kansas, Lawrence, KS, [3]NYX, New York Structural Biology Ctr., Upton, NY, [4]UPMC Children s Hosp. of Pittsburgh and Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [5]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

TANGO2 deficiency disorder (TDD) is an autosomal recessive multi-system disorder. TDD patients display variable phenotype including neurodevelopmental delay, seizures, and metabolic and cardiac crises/arrhythmias. Mitochondrial dysfunction, abnormalities in lipid homeostasis and impaired Golgi-ER trafficking have been reported however, the function of TANGO2 and the exact pathophysiology of the disease are unknown. Currently, there is no FDA-approved therapy for TDD. While Vitamin B5 has been reported to reduce metabolic crises in patients, other symptoms such as neurodevelopmental delays, arrhythmias, and rhabdomyolysis persist. There is currently no confirmed crystal structure of the TANGO2 protein and no knowledge of its quarternary state with only a predicted three-dimensional structure. Understanding the biophysical properties of TANGO2 protein is vital for ascertaining its function and facilitating the development of potential therapies using structure-based drug design methods.To answer these questions, we employed a purified TANGO2 protein by cloning wild-type TANGO2 cDNA with C-terminal His-tag coding motif into pET-27b(+) expression plasmid and transformed into BL21 cells, followed by fast protein liquid chromatography (FPLC). Protein purity was assessed using SDS PAGE. By using size exclusion chromatography (SEC), with the column calibrated using a set of molecular weight standards, we compared the TANGO2 elution peak against those of the protein standards. We estimated recombinant TANGO2 size of ~29 kDa supporting the conclusion that TANGO2 exists as a monomer, further supported by the newly obtained TANGO2 crystal structure.We concentrated purified TANGO2 and incubated it in different conditions to obtain crystals; with the best sample obtained from 1.0 M sodium/potassium phosphate pH 6.9. PISA analysis did not reveal any higher order quaternary structures, consistent with the monomeric form observed in the SEC experiment. Analysis of the

crystals showed that TANGO2 consists of a β-sheet core formed by 7 sheets and another group of 6 sheets, flanked by a cluster of α-helices on either side. A large, putative ligand-binding cavity lined with positively charged residues and two smaller cavities were found near one opening of the β-sheet core.Comparing wild-type with mutant TANGO2 constructs is an essential first step to understanding the dysfunction on the molecular level. Obtaining these structures will not only help us understand the molecular structure of TANGO2, but will eventually be necessary to identify pharmacophore target sites to develop small molecule chemical stabilizers to target missense unstable mutants.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2031T: Encoding diseases and phenotypes using large language models reveals linkages between Mendelian disorders and complex diseases

Authors:

G. Siwo, A. K. Waljee, E. R. Bowen; Univ. of Michigan Med. Sch., Ann Arbor, MI

Abstract:

Mendelian disorders are often difficult to diagnose, and the discovery of genetic and environmental modifiers of these diseases is challenging, given their rare nature and clinical heterogeneity. Studies of Mendelian disorders are further complicated by the dichotomous classification of genetic diseases into Mendelian and complex diseases, which obscures phenotypic relationships across diseases. We hypothesize that encoding all diseases (Mendelian and complex) using unified computable and multi-dimensional phenotypes could help overcome these challenges, opening new opportunities for understanding the genetics of Mendelian disorders by integrating information from complex diseases. To test this hypothesis, we investigate whether representations of ICD-derived computable phenotypes (Phecodes) using low-dimensional vectors derived from language models (LLMs) capture latent relationships between Mendelian and complex diseases. Mendelian disorders and complex diseases with highly similar LLM-derived encodings are enriched for comorbidities and share genetic dependencies and pathophysiological mechanisms. For example, the encodings of Cystic fibrosis are highly like those of several cancers that have a higher incidence in Cystic fibrosis patients, including breast, pancreatic, and respiratory cancers. The LLM-derived encodings also reveal shared relationships between Mendelian disorders, suggesting the need for integrated analyses of distinct Mendelian disorders, which could enhance statistical power and inform therapeutic targeting. For example, Sickle cell disease (SCD) and Von Willebrand disease are

highly similar based on their encodings, consistent with increasing evidence of the Von Willebrand factor in the pathophysiology of vaso-occlusive events in SCD patients and clinical studies therapeutically targeting the role of Von Willebrand factor in SCD. Our findings include known and potentially novel Mendelian-Mendelian, and Mendelian-complex disease relationships. Our results align with previous studies suggesting that Mendelian disorders represent the extremes of a spectrum of disease pathologies related to complex diseases, reflected in part by shared clinical phenotypes between these diseases. This work could inform integrated genetic studies of Mendelian disorders and complex diseases and has implications for developing diagnostics and therapeutics for Mendelian disorders.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2032T: Enhanced detection and genotyping of disease-associated tandem repeats using HMMSTR and targeted long-read sequencing

Authors:

K. Van Deynze[1], C. Mumm[1], C. Maltby[1], J. Switzenberg[1], P. Todd[2], A. Boyle[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ Michigan, Ann Arbor, MI

Abstract:

Tandem repeat sequences comprise approximately 8% of the human genome and are linked to more than 50 neurodegenerative disorders including ALS, Ataxia, and Huntington's Disease. Accurate characterization of disease-associated repeat loci remains resource intensive and often lacks high resolution genotype calls. We introduce a multiplexed, targeted nanopore sequencing panel and HMMSTR, a modified, profile-HMM, sequence-based tandem repeat copy number caller. This workflow does not require pre-alignment and is robust to sequencing errors including indels and substitutions. We benchmark our method's functionality on Nanopore sequencing data from plasmids constructed with varied repeat counts and motifs, as well as whole-genome Nanopore and PacBio Circular Consensus Sequencing (CCS) data and Cas9-enriched nanopore sequencing data for known TR sequences in the NA12878 cell line. When compared to 2 sequence-based and 2 signal-based methods, HMMSTR achieves a 1.1 average copy number difference from the CHM13 reference genome, 2.5 copies closer than the closest method. When downsampled to 5x coverage, HMMSTR maintains less than a 3 copy difference on average. Thus, we show HMMSTR outperforms current signal- and sequence-based callers relative to two assemblies and performs with high accuracy in heterozygous regions and at low read coverage. The flexible panel allows us to capture 60 known and candidate disease associated regions at an average coverage of >150x. Using these tools, we successfully

characterize known or suspected repeat expansions in patient derived samples. In these samples we also identify unexpected, expanded alleles at tandem repeat loci not previously associated with the underlying diagnosis including rare alleles in patients with ALS as well as co-occurring *FGF14* expansions in CANVAS patient samples with pathogenic expansions at *RFC1*. This genotyping approach for tandem repeat expansions is scalable, simple, flexible, and accurate, allowing for addition of newly discovered and candidate expansion targets. This approach offers significant potential for diagnostic applications and investigation of expansion co-occurrence in neurodegenerative disorders.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2033T: Expansion of the NDST1 syndrome and evidence for abnormal heparan sulfate

Authors:

C. Carlston[1], B. A. Borden[1], S. Sacharow[1], G. T. Berry[1], M. Pasquali[2], L. Rodan[1]; [1]Boston Children's Hosp., Boston, MA, [2]Univ Utah/ARUP Lab, Salt Lake City, UT

Abstract:

The *NDST1* gene encodes a bifunctional enzyme involved in the initial steps of heparan sulfate synthesis. These proteoglycans can be found on cell surfaces and in the extracellular matrix. In mice, loss of NDST1 results in hypoplastic lungs as well as developmental defects of the head, forebrain, a lack of olfactory bulbs, and axon guidance defects. The *NDST1* ortholog ndst1b morpholino knockdown in zebrafish (*Danio rerio*) causes delayed development, craniofacial cartilage abnormalities, shortened body and pectoral fin length. No human with biallelic loss of function of this gene have been identified. Biallelic missense variants in the NDST1 gene have been reported in association with autosomal recessive intellectual disability with variable motor delay, hypotonia, ataxia, seizures, postnatal growth deficiency, respiratory distress and dysmorphic facial features. Here we report a four-year-old boy with hypertelorism, agenesis of the corpus callosum with hydrocephalus, tracheoesophageal fistula, mixed sleep apnea, laryngomalacia, hypotonia, short stature, brachydactyly, patent ductus arteriosus, ventricular septal defect, cryptorchidism, hydronephrosis, seizures, and profound global developmental delay. A homozygous missense variant in the in the heparan sulfate N deacetylatase domain of NDST1 was identified in our patient, and he is to our knowledge the first patient with biallelic variants in this domain (all previously reported patients have at least one missense variant in the sulfotransferase domain of NDST1). In addition, glycosaminoglycan testing in urine and blood showed elevations in heparan sulfate (HS) and a non-reducing end assay suggested an

abnormality in the composition of HS with a marked decreased in N-sulfoglucosamine. This expands the phenotype of this rare neurodevelopmental and multiple congenital anomaly condition and proposes a biologically plausible novel biochemical marker.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2034T: Frequency of *FGF14* Intronic GAA Repeat Expansion in Patients with Multiple System Atrophy and Undiagnosed Ataxia in the Japanese Population

Authors:

T. Kakumoto[1], K. Orimo[1], T. Matsukawa[1], J. Mitsui[1], Y. Suzuki[2], S. Morishita[2], Japan Multiple System Atrophy Registry Consortium, T. Toda[1], S. Tsuji[1,3]; [1]The Univ. of Tokyo, Tokyo, Japan, [2]The Univ. of Tokyo, Kashiwa, Chiba, Japan, [3]Intl. Univ. of Hlth.and Welfare, Chiba-ken, Japan

Abstract:

Objective Multiple system atrophy (MSA) is a neurodegenerative disorder characterized by autonomic nervous system dysfunction, and cerebellar ataxia or parkinsonism. Recently, expanded GAA repeats (≥ 250 repeat units) in intron 1 of the *FGF14* gene have been shown to be responsible for spinocerebellar ataxia type 27B (SCA27B), a late-onset ataxia with autosomal dominant inheritance. Patients with SCA27B may also exhibit autonomic nervous system dysfunction, potentially overlapping with the clinical presentations of MSA patients. This study investigated the frequency of expanded GAA repeats in *FGF14* in patients with MSA and undiagnosed ataxia to evaluate involvement of the expanded GAA repeats in these conditions. **Methods** We analyzed 548 Japanese patients with MSA, 476 Japanese patients with cerebellar ataxia of unknown etiologies, and 455 healthy individuals. We conducted long-range polymerase chain reaction (LR-PCR) followed by agarose gel electrophoresis. LR-PCR products larger than 750 bps were gel-purified and subjected to long-read sequence analysis using PacBio Sequel II. On the basis of waterfall plots, we classified the repeat configuration of *FGF14* into three categories: (1) expanded pure GAA repeats, (2) complex configuration with expanded pure GAA repeats, and (3) complex configuration without expanded pure GAA repeats. We then counted repeat units of the expanded pure GAA repeats, allowing < 10% mismatches. Repeat-primed PCR analysis targeting the expanded GAA repeats was performed as previously described. **Results** We identified 12 individuals carrying expanded pure GAA repeats: ten had ≥ 250 repeat units and two had 200-249 repeat units. Among those with (GAA) ≥ 250, there were one MSA patient (0.2%), seven patients with undiagnosed ataxia (1.5%), and two healthy individuals (0.4%). The clinical presentations of patients with undiagnosed ataxia carrying (GAA)

≥ 250 included late-onset ataxia, downbeat nystagmus, episodic symptoms, and atrophic cerebellar vermis on MRI, consistent with those previously described in patients with SCA27B. The MSA patient with (GAA) ≥250 exhibited typical clinical features of MSA, not SCA27B. **Conclusions** We confirmed the presence of expanded pure GAA repeats in 1.5% of Japanese patients with undiagnosed ataxia, supporting the diagnosis of SCA27B. The clinical manifestations of SCA27B in the Japanese population were similar to those documented previously. Of the 548 MSA patients, we identified only one case of expanded GAA repeats, who presented typical clinical features of MSA. Further research is warranted to explore the possibility of potential association of expanded GAA repeats in *FGF14* with MSA.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2035T: Gaucher disease type 1 patient treated succesfully with Taliglucerase alfa and Eliglustat: a case report

Authors:

J. Pascual, C. Bartolomei Aguilar, L. Konczal; Univ. Hosp., Cleveland, OH

Abstract:

Gaucher Disease (GD) is an autosomal recessive disorder leading to accumulation of glucosylsphingosine in lysosomes due to impaired function of the glucocerebrosidase enzyme[1].Depending on residual function of the enzyme, there is a range of severity from perinatal lethal to asymptomatic. GD type 1 is a subtype leading to bone disease, hepatosplenomegaly, anemia, thrombocytopenia, interstitial lung disease, but absent primary central nervous system disease[2]. We present a 31 y/o female with GD Type I, diagnosed in 1995 by biochemical testing; molecular testing revealed two GBA gene pathogenic variants in trans: c.1226A>G and c. 115+ 1G>A. Since 2021, she has been treated with combined taliglucerase alfa, enzyme replacement therapy (ERT), and eliglustat, substrate reduction therapy (SRT). Prior to the addition of the SRT, she had been treated with ERT alone (changing from imiglucerase-> velaglucerase alfa-> taliglucerase alfa due to development of neutralizing antibodies with symptom progression). GD associated complications resulted in partial splenectomy, bilateral below knee amputations, multilevel degenerative disc disease s/p surgery, compression fractures, osteopenia. Prior to adding eliglustat, she was on taliglucerase alfa alone but experienced worsening fatigue and bone pain but negative taliglucerase alfa neutralizing antibody testing. At this point, eliglustat (SRT) was added in early 2021. Eliglustat is metabolized by the enzyme CYP2D6. Activity of CYP2D6 varies significantly due to genetic polymorphisms that can affect the drug's efficacy and safety. The patient was identified as an “extremely poor

metabolizer". Therefore, Cerdelga 84 mg every other day was initiated with EKG monitoring (patient on an SSRI, a CYP2D6 inhibitor). Patient reports significant clinical improvement/ improved activities of daily living after starting dual therapy. Before dual therapy, she had: frequent bone crises, fatigue, low energy, shortness of breath, orthostatic hypotension, vertigo. All symptoms have improved greatly: bone crises decreased in frequency and intensity to once every few months. Biomarkers have improved (predual therapy > postdual therapy): CHITO 823.8 - >233 nmol/hr/mL (<78.5 nmol/hr/mL); ACE 44- >34 IU/L (32-107.9 IU/L); TRAP 13.63->2.0 IU/L (0.28-9.84 IU/L); WBC 2.8 ->5.8x103/uL (4.4-11.3x103/uL); platelets 123>264x103/uL (150-400 x103/uL). Abdominal MRI shows reduction in liver size (159.3 mm - >146.7 mm) and splenic size (99.6 mm - >90.2 mm). She continues taliglucerase alfa enzyme replacement therapy (60 U/ kg q2w) and recently eliglustat was increased to 84 mg PO daily since her EKG is normal and SSRI was discontinued.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2036T: Generating and sharing valuable rare disease data via the GREGoR Consortium ⭐

Authors:

M. Wheeler[1], M. Conomos[1], S. Gogarten[1], B. Heavner[1], S. Jhangiani[2], S. Marwaha[3], A. O'Donnell-Luria[4], P. Petrowski[3], A. Stilp[1], C. Tong[1], M. Wheeler[3], J. Yu[3], J. Zhen[3], Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium; [1]Univ. of Washington, Seattle, WA, [2]Baylor Coll. Med., Houston, TX, [3]Stanford Univ., Palo Alto, CA, [4]Boston Childrens Hosp. and the Broad Inst., Boston, MA

Abstract:

The Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium is working to advance rare genetic disease research through the application and standardization of new omic technologies and diagnostic approaches that are not routine in clinical practice. This ongoing effort includes rapid Consortium-wide sharing of data generated by participating Research Centers, Partner Members, and Research Grant Awardees, including controlled-access release of a rich dataset to the broader scientific community.
The GREGoR Consortium Dataset is broadly consented for research and is available via controlled access on NHGRI's AnVIL platform, a secure and scalable cloud environment for data sharing and analysis. Researchers may request access to controlled-access GREGoR Consortium Data by submitting a Data Access Request through dbGaP or DUOS. Additionally, the GREGoR Consortium has developed a publicly available open-access variant browser to enable searches

for variants or genes of interest across the Consortium joint callset. The GREGoR Consortium Variant Browser is available at https://variants.gregorconsortium.org.
As of June 2024, the GREGoR Consortium Dataset consists of data generated for participants from more than 1,800 families. In addition to short-read exome sequencing data, short-read whole genome sequencing data, long-read genome sequencing data, and RNA-sequencing data, this Dataset includes pedigree and phenotype information, candidate genetic findings from participating Research Centers, and detailed experimental metadata. Short read whole genome sequencing data is uniformly processed by the GREGoR Data Coordinating Center and includes a Consortium joint callset of single nucleotide variants and indels. The GREGoR Dataset is organized by an expandable data model, which is publicly available at https://github.com/UW-GAC/gregor_data_models. This data model serves as a robust framework to facilitate data interoperability across participating Research Centers and supports secondary data use by the broader scientific community.
We will present additional details about the GREGoR Dataset, including additional information about how to obtain access, and the range of tools available to support rare disease research, including family-based annotated variant filtration and searching candidate genes across the cohort.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2037T: Genetic epidemiology and clinical characteristics of SCA424 patients with mutations in *CACNA1G*based on J-CAT study

Authors:

Y. Hama[1], H. Date[2], Y. Takahashi[3,1], H. Mizusawa[4], J-CAT (Japan Consortium of Ataxias); [1]Dept. of Genomic Med., Natl. Ctr. Hosp., Natl. Ctr. of Neurology and Psychiatry, Kodaira, Japan, [2]Dept. of Genomic Med., Natl. Ctr. of Neurology and Psychiatry, Kodaira, Japan, [3]Dept. of Neurology, Natl. Ctr. Hosp., Natl. Ctr. of Neurology and Psychiatry, Kodaira, Tokyo, Japan, [4]Natl. Ctr. of Neurology and PsychiatryNatl. Ctr. of Neurology and Psychiatry, Kodaira, Japan

Abstract:

Background and objectives Autosomal dominant spinocerebellar ataxia (SCA) is a clinically and genetically heterogeneous group of neurodegenerative disorders characterized by progressive cerebellar ataxia. Whole genome sequencing analysis (WGS) reveals genetic epidemiology of rare ataxia other than repeat expansion diseases. The aim of the study was to elucidate the molecular epidemiology and clinical characteristics of patients with pathogenic variants in *CACNA1G* (SCA42) based on Japan Consortium of Ataxias (J-CAT) study. **Methods** From 2016

through November 2022, patients with cerebellar ataxia have been registered in J-CAT nation-wide. WGS was conducted for all patients with negative results in the initial screening (SCA1, 2, 3, 6, 7, 8, 12, 17, 27B, 31, 36, CANVAS, DRPLA and HD), and the frequency and characteristics of cases with pathogenic variants in *CACNA1G* was extracted from the J-CAT database. **Results** During the period, 1895 patients were enrolled and WGS was conducted for 1058 patients. Among 1058 patients, we detected six patients with a known mutation, p.R1715H in *CACNA1G,* from different families. No other known mutations in *CACNA1G* were identified. This variant was located in segment 4 (S4) of domain IV. S4 contains many positively charged amino acids and is the voltage sensor. The mean age at onset of all patients with the variant was 41 years old and the mean age at registration was 54 years. All cases had a slowly progressive disease course. Five of the six cases had a family history. The mean modified Rankin scale at enrollment was 2 points, and the mean SARA score at enrollment was 11 points. The initial symptoms were gait disturbance in 3 cases and dysarthria in 2 cases. Hearing loss, increased tendon reflexes, decreased tendon reflexes, or cognitive impairment was observed in one case, respectively. None had headache, muscle weakness, involuntary movement, sensory loss, autonomic dysfunction, or seizures. Brain MRI was performed on all patients. All patients had cerebellar atrophy, and one patient had cerebral white matter lesion. One patient underwent ^{99m}Tc-ECD brain SPECT, which revealed decreased cerebellar blood flow. **Discussion and Conclusion** SCA42 with p.R1715H variant in *CACNA1G* gene shows typically a slowly progressive pure cerebellar ataxia, but sometimes it may be complicated by extracerebellar involvement such as hearing loss and cognitive impairment. To our knowledge, this is the first case of SCA42 with p.R1715H with cerebral white matter lesions, though there is a possibility of accidental complication. J-CAT includes ultra-rare ataxias and may be a good foundation for their researches. **Funding sources related to this work**: None.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2038T: Genetic Factors in Incontinentia pigmenti patients with subclinical autoimmunity

Authors:

M. Ursini[1], E. Spinosa[2], S. Picascia[1], C. Casale[2], A. Colella[1], A. Pescatore[2], M. B. Lioi[3], F. Fusco[1]; [1]IGB ABT CNR, Naples, Italy, [2]IGB-ABT, CNR, Naples, Italy, [3]Univ. of Basilicata, Potenza, Italy

Abstract:

Background:Incontinentia Pigmenti (IP; OMIM#308300) is a rare genetic disorder caused by mutations in the IKBKG/NEMO gene. Recently, we identified neutralizing auto-antibodies (Auto-Abs) against interferon (IFN)-α2 and IFN-ω in 25% of IP patients, which were present before SARS-CoV-2 infection. In this study, we provide a molecular characterization of a cohort of IP patients, analyzed through neutralizing assays in collaboration with the IMAGINE Institute. Methods:Clinical data and biological samples (DNA and plasma/serum) from 21 Italian IP family trios were collected and stored at -80°C in the Incontinentia Pigmenti Genetic Biobank (IPGB; http://www.igb.cnr.it/ipgb). Additionally, peripheral blood mononuclear cells (PBMC) from 4 patients were isolated using Ficoll-Paque gradient centrifugation. Long-range PCR and Sanger sequencing were performed to identify alterations in the IP locus. Results:All patients (mean age 30 years; range 1-60 years) exhibited the conventional IP skin phenotype. Alterations in the IP locus were identified in 81% (17/21) of the patients. Specifically, the NEMOΔ4-10 recurrent deletion was found in 14 patients, and point mutations were identified in 3 patients (c.628_651delinsCG, c.646delG, c.976_978del). Conclusions:These findings suggest that hidden autoimmunity is a new clinical aspect of IP and indicate that further genetic aspects should be investigated. Grants:This research was supported by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT)

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2039T: Genetic testing diagnostic yield by age and test modality across 856 patients in a cohort of adult and pediatric patients with epilepsy at a tertiary care center

Authors:

E. Moura Coelho da Silva[1,2], T. Brünger[1,2,3], G. Taylor[1,2], L. Montanucci[1,2], C. Leu[1,2], A. Merket[1,2], A. Cherukara[1,2], M. Sinha[1], R. A. Tai[1], S. D. Lhatoo[1,4], G. K. Von Allmen[1,5], D. Lal[1,2,3]; [1]Dept. of Neurology, McGovern Med. Sch., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Ctr. for Neurogenetics, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Cologne Ctr. for Genomics (CCG), Cologne, Germany, [4]Texas Inst. for Restorative Neurotechnologies, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]Div. of Child and Adolescent Neurology, Dept. of Pediatrics, McGovern Med. Sch., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract:

Epilepsy can arise from various causes but genetic factors are increasingly recognized. The number of epilepsy cases with a clinically identifiable genetic cause varies based on age of

onset, comorbidities, epilepsy type, and genetic testing modality used. For example, the diagnostic yield of multigene panels (MGP) is up to 25% in epileptic encephalopathy but closer to 10% in adults. This study aims to determine the diagnostic yield of genetic testing in pediatric and adult epilepsy patients at a tertiary care center treating many individuals with intractable epilepsy. We retrospectively searched, aggregated, and curated clinical genetic testing data for epilepsy patients who had testing between 2007 and 2024. Screening strategies included computational searches for epilepsy-related genetic content in electronic health records and downloading genetic test results from diagnostic lab portals. After data curation, we identified 856 unique patients with epilepsy and 978 genetic test reports. On average, each patient underwent 1.14 genetic tests. The most ordered test was exome sequencing (ES) (44.7%, 437/978), followed by MGP (43.6%, 427/978), chromosomal microarray (CMA) (8.0%, 78/978), and single gene testing (SG) (3.7%, 36/978). Almost half (44.6%, 436/978) of test reports revealed no variant, 34.3% (335/978) listed only variants of uncertain significance (VUS), 20.4% (200/978) reported at least one likely pathogenic (LP) or pathogenic (P) variant, and 7.2% (7/978) contained only likely benign or benign variants. Of the 200 reports with at least a LP/P variant, 81.5% (163/200) were considered potentially diagnostic based on inheritance pattern. In 76.7% (125/163) of these results, a LP/P variant was found in a gene or copy-number variant known to cause seizures, while the other 23.3% (38/163) were incidental findings, such as cancer-associated genes. Thus, the overall diagnostic yield was 12.8%. When stratifying by genetic testing type, the highest diagnostic yield was ES (15.6%), followed by SG (13.9%), MGP (11.5%), and CMA (3.8%). Stratifying by age at genetic testing, the highest diagnostic yield was for those <1 year (23.8%), followed by >1-5 years (12.9%), and >5 years (9.9%). ES had the highest diagnostic yield, consistent with literature. Before 2023, 38.3% of tests ordered were ES versus 75.1% after 2023, aligning with guidelines for ES as first-tier testing for unexplained epilepsy. Roughly one in four LP/P variants was an incidental finding, stressing the importance of thorough informed consent. The large number of VUSes in epilepsy patients underscores the need for further evaluation, like bioinformatic analysis or segregation testing.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2040T: Genome Sequencing in Early-onset Psychosis

Authors:

A. Besterman[1], C. Blucher[2], J. Lenberg[2], B. Schultz[2], L. Olsen[2], H. Sweet[1], A. Omorodian[1], K. Patel[1], E. Patel[1], A. Ramalingam[2], S. Batalov[2], J. Friedman[1], S. Kingsmore[3], C. Bartley[4], M. Bainbridge[2]; [1]UCSD, San Diego, CA, [2]Rady Children's Inst. for Genomic Med., San Diego, CA, [3]Rady Children s Inst. for Genomic Med., Coronado, CA, [4]NIMH, Bethesda, MD

Abstract:

Background: Early-onset psychosis (EOP), defined as any psychiatric diagnosis with pronounced psychotic symptoms before age 18, is linked to lower premorbid psychosocial function, more hospitalizations, poorer cognitive functioning, and worse overall prognosis than adult-onset illness. Recent reports suggest an elevated rate of recurrent copy number variants in EOP compared to unaffected controls and those with autism spectrum disorder. However, the full spectrum of genetic variation underlying EOP remains unclear, and both the clinical and personal utility of genetic testing in EOP is also unclear. Therefore, we initiated a prospective, observational study utilizing genome sequencing (GS) in youth with EOP. Guardians completed the Genome Empowerment Scale (GEmS).
Methods: We performed CLIA-grade, diagnostic, reflex-trio, GS on 20 consecutive youth aged 11-17 admitted to the Rady Children's Hospital Inpatient Psychiatry Service with EOP. GS results were returned to guardians, who then completed the GEmS, consisting of four scales: meaning of a diagnosis, emotional management, seeking information and support, and implications and planning.
Results: Twenty-four youth with EOP were enrolled, with a mean age of 15.5 ± 1.7 years. 54.2% were female and 66.7% identified as Hispanic. One patient had a pathogenic variant in *SLC20A2* (Primary Familial Brain Calcification), resulting in a pathogenic/likely pathogenic (P/LP) diagnostic yield of 4.2%. Five patients had variants of uncertain significance (VUS) (4 single nucleotide variants (SNVs) and 1 copy number variant). The VUS occurred in genes associated with neurodevelopmental, including chromatin modification, synapse formation, and neural progenitor proliferation. GEmS subscale scores were: 1) Meaning of a diagnosis: 43.9 ± 13.6 2) Emotional management: 26 ± 6.7 3) Seeking information and support 28.8 ± 6.7, and 4) Implications and planning: 23.9 ± 12.3.
Conclusion: P/LP variants can be identified in patients with EOP using GS at rates similar to autism, but lower than intellectual disability or epilepsy. The VUS we identified suggest that there are likely more under-recognized EOP-associated rare genetic variants and that the true yield of GS in EOP may be higher than reported here. Larger GS studies in EOP would greatly enhance our ability to establish a wider range of high-confidence EOP genes. Most of the VUS identified were SNVs, not recurrent CNVs as previously reported, suggesting GS may broaden our diagnostic capabilities in EOP. Lastly, GEmS results suggest that receiving a genetic diagnosis for a child with EOP is significantly empowering.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2041T: Genomic analyses of 317 consanguineous kindreds from the Middle East and North Africa facilitate the discovery of novel autosomal recessive neurodevelopmental rare disease traits

Authors:

A. Saad[1,2], R. Duan[1], D. G. Calame[1,3,4], H. Du[1], J. M. Fatih[1], P. Kaur[1], Z. Dardas[1], G. M. H. Abdel-Salam[5], M. Y. Issa[5], H. Elbendary[5], K. Rafat[5], N. W. Saadi[6], H. M. Moussa[7], M. Zakarriah[8], E. El-Anany[8], L. Bastaki[9], H. Alrohaif[9], M. Djuric[10], M. Arabi[11], M. Ashkanani[12], I. Herman[1], T. Mitani[1], S. N. Jhangiani[13], Z. Coban Akdemir[1,14], R. A. Gibbs[1,13], D. Pehlivan[1,3], J. E. Posey[1], M. S. Zaki[5], D. Marafi[1,15], J. R. Lupski[1,3,13]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Med. Molecular Genetics Dept., Human Genetics and Genome Res. Inst., Natl. Res. Ctr. (NRC), Cairo, Egypt, [3]Texas Children's Hosp., Houston, TX, [4]Div. of Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [5]Clinical Genetics Dept., Human Genetics and Genome Res. Inst., Natl. Res. Ctr. (NRC), Cairo, Egypt, [6]Coll. of Med., Univ. of Baghdad, Baghdad, Iraq, [7]Specialties Hosp. Ctr., Nouakchott, Mauritania, [8]Section of child neurology, Adan Hosp., Ministry of Hlth., Kuwait, Kuwait, [9]Kuwait Med. Genetic Ctr., Al-Sabah Hosp., Ministry of Hlth., Sulaibikhat, Khaldiya, Kuwait, [10]Section of Child Neurology, Dept. of Pediatrics, Farwaniya Hosp., Ministry of Hlth., Sabah Al-Nasser, Kuwait, Kuwait, [11]Dept. of Neurology, Ibn Sina Hosp., Sabah Med. Area, Kuwait, Kuwait, [12]Dept. of Neurology, Amiri Hosp., Ministry of Hlth., Kuwait, Kuwait, [13]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [14]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [15]Dept. of Pediatrics, Faculty of Med., Kuwait Univ., Safat, Safat, Kuwait

Abstract:

Rare variants contributing to Mendelian disease can occur in different allelic configurations: monoallelic, biallelic, and multigenic/multiallelic. Variant transmission patterns vary across different population structures, swaying the contributions of *de novo* and novel variants to recessive or complex traits. To investigate allelic transmission in presumed autosomal recessive traits, we established a cohort study of 317 unrelated families from the Middle East and North Africa (MENA), with 270/317 (85.2%) reporting parental consanguinity by familial history. The MENA cohort, comprising 1265 individuals, included 466 affected subjects with neurodevelopmental disorders (NDDs) and has served as a valuable resource for 'disease gene' discovery. Family-based exome sequencing (ES) and rare variant analyses yielded molecular findings in 203/317 (64%) of the families, including (81%) with known disease genes/loci and (19%) with novel candidate disease genes. Overall, there are 206 discrete NDD genes/genomic loci in which a plausible disease-contributing pathogenic variant allele was uncovered; these

include 191 distinct genes with single nucleotide variants (SNVs) and small indels and 15 different genomic intervals associated with copy number variants. Most variants (74%) were homozygous due to parental consanguinity; as anticipated, these variant loci were embedded within the absence of heterozygosity (AOH) blocks derived through identity-by-decent. Notably, the genome-wide distributive AOH pattern partially contributed to the evolution of multiple pathogenic variants (MPVs) in 23/203 (11.3%) cases. Remarkably, one case featured a homozygous whole-gene deletion of *SLC13A5* generated by *Alu/Alu*-mediated genomic rearrangement. Other allelic patterns, including *de novo* heterozygosity (11%), X-linked hemizygosity (3.1%), and biallelic compound heterozygosity (1.6%) were less frequently observed. Large genomic rearrangements (8.7%), and chromosomal abnormalities (1.6%) were also uncovered. Still, 63.7% of the identified SNVs in known disease genes were novel variants. Proposed novel and candidate NDD-associated genes, including *ASTN1, CAPZA2, CCDC53, CDK9, DNAJA3, ESM1, GOLGA2, LIMCH1, MDGA2, NAV2, PPP1R3F, STX6, TRAK2, UGGT1, USP19, WDR6* are currently undergoing further investigations. These findings highlight the power of family-based genomic analysis, comprehensive clinical phenotyping, and international collaborative data sharing to delineate allelic series and expedite the discovery of novel NDD genes/loci, paving the way for a deeper understanding of the underlying neurobiology of disease

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2043T: GSK3β inhibition is a candidate pharmacological mechanism for treating NGLY1 deficiency

Authors:

E. Maclaughlin, C. Chow; Univ. of Utah, Salt Lake City, UT

Abstract:

NGLY1 deficiency is an exceedingly rare genetic disorder caused by mutations in the N-glycanase 1 gene, resulting in developmental delays and a range of neurological symptoms. N-glycanase 1 is a cytosolic enzyme that deglycosylates misfolded proteins before they enter the ubiquitin-proteasome system (UPS). One substrate is transcription factor NRF1, which acts as a sensor for proteasome stress and requires NGLY1 processing to upregulate UPS subunits to overcome proteasome stress. Apart from NRF1, the underlying molecular mechanisms contributing to the pathophysiology of NGLY1 deficiency remain poorly understood, though dysregulation of proteostasis and altered signaling pathways are implicated as contributors to the disease phenotype. A previous drug repurposing screen from our lab demonstrated that both genetic

and pharmacological inhibition of GSK3β offer protective benefits against lethality and proteasomal stress in a *Drosophila NGLY1* deficiency model. In this study, we expand on GSK3β inhibition as a therapeutic option in human cell culture models of NGLY1 deficiency and find differential effects of drugs targeting GSK3β on cellular phenotypes. We are performing an exhaustive analyses of all GSK3β inhibitors on NGLY1 deficiency cells. Strikingly, some GSK3β inhibitors can increase cell proliferation in NGLY1 deficiency cells, but do not rescue sensitivity to proteasomal stress. We will present pharmacological and genetic data revealing the mechanism for the ability of GSK3β inhibition to rescue NGLY1 deficiency. Understanding the commonalities and molecular mechanisms underlying NGLY1 cellular responses to different GSK3β inhibitors may pave the way for the development of novel therapeutic strategies for NGLY1 deficiency and related neurodevelopmental disorders.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2045T: HNRNP-Related Neurodevelopmental Disorders: Expanding knowledge of known conditions and prompting novel studies into candidate genes.

Authors:

M. Gillentine[1], A. Hansen[2], C. Da Silva[3], L. Phillips[3], S. Plecas Maricic[3], A. B. Beaudet[3], K. Caine[3], M. Ferguson[4], K. Murias[3], B. Hallgrimsson[3], M. Balasubramanian[5], J. Bain[6], A. Kline[4], P. Au[7]; [1]HNRNP Family Fndn., Seattle, WA, [2]Geneial, Houston, TX, [3]Univ. of Calgary, Calgary, AB, Canada, [4]Greater Baltimore Med Ctr., Baltimore, MD, [5]Univ. of Sheffield, Sheffield, United Kingdom, [6]Columbia Univ. Med. Ctr., New York, NY, [7]Univ. of Calgary, Cumming Sch. of Med., Calgary, AB, Canada

Abstract:

The HNRNP-Related Neurodevelopmental Disorders (HNRNP-RNDDs) are a group of ultra-rare NDDs caused by damaging variants in the heterogeneous nuclear ribonucleoprotein (hnRNP) gene family. To date, eight HNRNP-RNDDs have been described and multiple candidate HNRNP-RNDDs are being investigated. Here we are utilizing direct interactions with patient families through the HNRNP Family Foundation supported by a novel platform, Geneial, to integrate three natural history studies in order to compare phenotypes within and across the *HNRNP* genes. To date, over 100 individuals representing HNRNPH2-RNDD, HNRNPK-RNDD (Au-Kline syndrome), HNRNPQ/SYNCRIP-RNDD, HNRNPU-RNDD, and HNRNPUL2-RNDD are included in this joint natural history study. which increases effective sample size when investigating features shared between individual HNRNP-RNDDs. Geneial is providing a

consistent, unified data management and research interface for these studies, some of which are maintained in REDCap. For research participants, our model includes remote enrollment (informed consent) and captures electronic patient-reported outcomes., expanding our reach to those unable to be evaluated in person. High level phenotyping has shown that the HNRNP-RNDDs are more alike than when compared to other NDDs, and this study will allow for deeper phenotyping to identify similarities and differences. Furthermore, incorporation of the studies allows for inclusion of the rarer, "orphan" *HNRNP* genes that have evidence for pathogenicity but lack both clinical and research attention. This international endeavor aids in understanding of these disorders and serves as a model for investigation of other related conditions into gene families.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2046T: Identification of *ERCC2* as a Novel Waardenburg Syndrome Gene Through Analysis Human Neural Crest Developmental Regulatory Networks ⭐

Authors:

J. Cheng[1], H. Yuan[1], y. wang[2]; [1]West China Hosp., Chengdu, China, [2]Academy of Mathematics and Systems Sci., Beijing, China

Abstract:

Hearing loss (HL) is the most common sensory disorder in humans, and identifying new causal genes related to inherited HL remains a challenge. Previous studies have often overlooked the developmental context and interactions of HL genes. We investigated these genes in neural crest cells during ear development by integrating paired gene expression and chromatin accessibility data across multiple developmental stages. This led to the construction of a multi-stage regulatory network, ihReg-CNCC, which captures intricate regulatory patterns. Using PageRank in ihReg-CNCC, we predicted candidate HL genes and screened them in a large HL cohort comprising 25,512 patients across China through whole genome sequencing. We revealed a *de novo* frameshift mutation, c.812_813delTC, in the *ERCC2* gene in a Waardenburg syndrome proband. Additionally, the *Ercc2* cKO mice exhibited a progressive HL phenotype. Our research highlights ihReg-CNCC as a valuable tool for identifying deafness-causing genes and their interactions, emphasizing the significance of considering developmental contexts and gene interactions in future studies on HL.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2047T: Integrative analysis of genotype and phenotype data from the UK Biobank identified over 50-fold higher-than-expected prevalence of hereditary pancreatitis

Authors:

X. Zeng, J. L. Nierenberg, B. Dulken, A. Petukhov, J. B. Chang, S-G. Ji; BridgeBio Pharma, Inc., San Francisco, CA

Abstract:

The challenge in treating rare genetic disorders starts with its diagnosis, which is compounded by ascertainment bias, incomplete penetrance, and variable expressivity. As most rare diseases are studied in clinically ascertained cohorts, their diagnoses trend towards higher specificity but lower sensitivity.

Hereditary pancreatitis (HP) is characterized by early onset recurrent acute pancreatitis (AP) that eventually leads to chronic pancreatitis (CP). CP is a progressive condition that features persistent inflammation and tissue damage in the pancreas, resulting in abdominal pain, high risk of pancreatic endocrine and exocrine insufficiency, and increased risk of pancreatic cancer. HP is considered extremely rare, reportedly affecting 4 in a million individuals. Over the past few decades, genetic and functional studies confirmed the central role of elevated intra-pancreatic trypsin level and activity underlying HP, identifying disease-causing rare variants in major players involved in trypsin metabolism and secretion (*PRSS1*, *SPINK1*, *CTRC*, *CPA1*, *CEL*, *CFTR*, and *TRPV6*).

To better understand the prevalence and penetrance of HP, we identified AP and CP subjects (N = 4,971) from the UK Biobank (UKB) and systematically surveyed all known pathogenic variants (n = 961, curated from ClinVar, HGMD, Varsome, and Invitae) using whole exome sequencing data. We found that the penetrance of pathogenic variants is overall lower than reported from studies of clinical cohorts (e.g. 80% in HP pedigrees versus 38% in the UKB for *PRSS1* p.R122H). However, despite the low penetrance, we identified the per 100,000 prevalence for gain-of-function dominant (*PRSS1*), loss-of-function dominant (*CPA1* and *CEL*), and recessive (*SPINK1*, *CTRC*, *CFTR*, and *TRPV6*) variants being 4.1, 2.3, and 16.3, respectively. Collectively, the estimated prevalence of HP is 22.6 per 100,000, which is over 50-fold higher than reported in literature, suggesting adult-onset patients who are unlikely to undergo genetic testing may be undercounted in clinical studies.

The higher-than-expected prevalence in the UKB is consistent with recent reports in monogenic developmental disorders, *NOTCH3*-associated CADASIL, and other dominant disorders, highlighting a critical but under-recognized role of genetic variants in disease variability. Given

that the UKB has a known ascertainment bias towards healthy individuals and the incompleteness of its EHR data, our observed prevalence and penetrance are likely underestimates of the general population, shedding a new light on drug development opportunities for ultra-rare genetic diseases where conventional beliefs about their prevalence may be biased.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2048T: Investigating the molecular function of *THAP12*, a novel gene associated with a rare epileptic encephalopathy

Authors:

B. Rampal[1,2], K. Ochenkowska[1,2], M. Liao[1,2], E. Samarut[1,2]; [1]Université de Montréal, Montreal, QC, Canada, [2]Ctr. de Recherche du Ctr. Hosp.ier de l'Université de Montréal (CRCHUM), Montreal, QC, Canada

Abstract:

Lennox-Gastaut Syndrome (LGS), a severe form of epileptic encephalopathy developing in childhood, presents significant challenges in both diagnosis and treatment. Since the etiology of LGS in about 30% of patients is unknown, there is a dire need to understand the pathogenic mechanisms of this disease. This study aims to investigate the molecular mechanisms which contribute to the pathogenicity of LGS in two symptomatically diagnosed siblings. Whole genome sequencing performed on both siblings revealed a compound heterozygous mutation in *THAP12*, a novel gene never previously associated with neurodevelopmental disorders. Interestingly, *THAP12* contains a zinc-finger domain which has been predicted to enable DNA binding. This leads us to believe that *THAP12* acts as a transcription factor regulating the expression of key genes involved in cell signalling and proliferation, and that mutations in this gene may lead to epileptic encephalopathy. To study the pathogenicity of *THAP12* mutations and unravel its molecular mechanisms, we employed a comprehensive approach utilizing both in-vitro and in-vivo models. Fibroblasts obtained from our patients allow for examination of the cellular responses to the mutation. The development of *THAP12*-KO HEK293FT cells will allow us to identify potential chromatin binding sites through ChIP-seq. Mice models carrying the patients' mutation will be used to provide insight into embryonic development. We also created a stable line of *THAP12*-KO zebrafish to validate findings from our fibroblasts and to examine how neurodevelopmental processes may be impacted by a loss of *THAP12*. Analysis of patient fibroblasts revealed a loss of *THAP12* at the protein level and altered growth and morphology. Mice models induced with our patient mutations lead to embryonic lethality, emphasizing its

importance in vertebrate development. Our zebrafish models showed expression of *THAP12* in the developing brain, and severe neurodevelopmental defects were observed when *THAP12* was knocked out. Overall, our preliminary results indicate *THAP12* may play a significant role in the development of epileptic encephalopathy. By unravelling the pathogenic mechanisms of mutations in *THAP12*, our work has the potential to unveil actionable molecular targets, needed for the development of non-canonical therapeutic strategies.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2049T: Investigating the role of the GABA transporter Solute carrier 6 member 1 (*SLC6A1*) in Mendelian disease traits

Authors:

N. Gogate, K. Jay, R. German, V. Gomez, National Barin Gene Registry, J. E. Posey, M. Wangler; Baylor Coll. of Med., Houston, TX

Abstract:

Dysregulation of Gamma-Aminobutyric acid (GABA), the principal inhibitory neurotransmitter, is implicated in a spectrum of neurological disorders. Solute carrier 6 member 1 (*SLC6A1*) encodes the sodium- and chloride-dependent GABA transporter 1 (GAT-1) and plays an essential role in regulating synaptic neurotransmitter signaling and homeostasis. Rare heterozygous pathogenic variants in *SLC6A1* are associated with Myoclonic-atonic Epilepsy (MIM#: 616421). 90% of individuals with variants in *SLC6A1* exhibit epilepsy and developmental delay. In addition, autistic features, intellectual disability, hypotonia, sleep disorders, and attention-deficit/hyperactivity disorder have also been reported with variable severity. Currently, the lack of data for managing expanded SLC6A1-related neurodevelopmental disorders hinders the development of effective treatments. In this study, we recruited 14 patients with *SLC6A1*-related disorders through the National Brain Gene Registry to determine symptom variability by variant type. We utilized Human Phenotype Ontology (HPO) based phenotypic clustering analysis to uncover phenotypic subgroups that may be driven by distinct variants or variant types in *SLC6A1.* The clustering analysis reveals four unique phenotypic signatures, suggesting that traits associated with variants in *SLC6A1* have partially overlapping, but distinct phenotypes indicative of clinical subgroups within the *SLC6A1*-phenotypic spectrum. Additionally, the rapid neurodevelopmental assessment protocol (RNAP) highlighted that cognition and communication were the most affected domains, significantly impacting quality of life for the majority of the cases studied. Overall, we find known loss-of-function variants to induce similar and severe phenotypes, while missense variants are involved in variable clinical presentation.

We support this finding with *in vivo* evaluation in a Drosophila model. Thus, investigating the unique spectrum of clinical phenotypes of *SLC6A1*-related disorders, we can begin to understand the genotype-phenotype associations underlying the variability in disease expression. These correlations will inform clinical diagnosis, prognosis, treatment planning, and the design of targeted clinical trials.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2050T: Is Classic Galactosemia a Progressive Disease? The First Longitudinal Natural History Data in Children with Classic Galactosemia.

Authors:

E. Bailey[1], R. Perfetti[1], S. Shendelman[1], C. Dayley[2], A. Durrant[2], S. P. Dickson[2], S. B. Hendrix[2]; [1]Applied Therapeutics, Inc, New York, NY, [2]Pentara Corp., Millcreek, UT

Abstract:

Classic Galactosemia is an autosomal recessive metabolic disease resulting in an inability to metabolize the sugar galactose. Newborn screening and implementation of a galactose-restricted diet have substantially decreased mortality. However, despite dietary adherence, patients continue to develop significant long-term neurological and behavioral complications. Registry-based reporting has demonstrated that new symptoms, such as tremor, appear over time and are more common in later life. Caregivers have reported that the burden of illness increases with age, and that symptoms worsen over time. However, to date there has been no longitudinal study of patients with Classic Galactosemia, where the same patients perform the same test repeatedly over time, to determine whether the symptoms are worsening progressively, or if the children are only developing skills more slowly than their normative peers do. The AT-007-1002 clinical trial (NCT04902781) provided an opportunity to answer this question. In order to examine whether individual children lose skills and deteriorate functionally over time, we examined the change in longitudinal raw scores of children in the placebo group over the 18 months of the study compared to standard scores. Raw scores worsened over 18 months for several measures of behavior (withdrawal [mean=-6.25 (SE=1.86)], behavioral symptoms index [-24.12 (10.55)] social skills [-2.12 (1.52)], adaptive skills index [-18.12 (11.73)], daily function (activities of daily living [-0.25 (1.57)] and functional impairment index [-9.88 (3.94)]), demonstrating that patients lost skills that they previously had. Raw scores did not decline on cognition [+12.83 (1.77)] or fine motor skills [+4.51 (2.01)], whereas standard scores did, suggesting that for these outcomes children were not losing skills, but instead were not developing skills as expected compared to normative peers. This data is also supported by a

cross-sectional analysis at baseline of the 47 children aged 2-17 enrolled in the study using a consistent methodology.This is the first longitudinal data using standardized tests of behavior, function and cognition in children with Classic Galactosemia. Results indicate that the disease is progressive, with children losing skills over time with regard to behavior and daily function. On cognition, children developed additional skills over time, but not at the rate of their normative peers. In both cases, the result is a widening gap between children with Classic Galactosemia and their peers.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2051T: Jacobsen and Beckwith-Wiedemann Syndromes in a Child with Neonatal Encephalopathy and De Novo 11pter Duplication and 11qter Deletion

Authors:

P. Pakhathirathien[1], H. Vaseenon[1], W. Thammachote[2], P. Sinpitak[3], C. Tocharoentanaphol[3], N. Jinawath[2], D. Wattanasirichaigoon[1]; [1]Div. of Med. Genetics, Dept. of Pediatrics, Faculty of Med. Ramathibodi Hosp., Mahidol Univ., Bangkok, Thailand, [2]Program in Translational Med., Faculty of Med. Ramathibodi Hosp., Mahidol Univ., Bangkok, Thailand, [3]DNA Ctr., Bangkok, Thailand

Abstract:

Jacobsen syndrome (JS) is recognized as pre- and post-natal growth failure with distinctive dysmorphic features and thrombocytopenia. JS is caused by a deletion of the distal 11q. Beckwith-Wiedemann syndrome (BWS) is characterized by overgrowth of prenatal onset and a high risk of embryonal tumors during the early childhood. Genetic defects underlying BWS are complex, including duplication of 11p segment involving the imprinting region. Here, we report a male infant who initially presented with prematurity (32 weeks of gestational age) with borderline large for gestational age (2500 g.), nonspecific facial dysmorphic features, suspected nanophthalmos, thrombocytopenia, marked hypotonia, poor feeding, and profound neonatal lethargy. Multiple investigations for sepsis, inborn metabolic disorders, and chromosomal abnormalities were performed. The investigations revealed normal electrolytes, burst suppression pattern of the electroencephalography, normal amino acid profiles of the plasma and cerebrospinal fluid which excluded nonketotic hyperglycinemia. His brain computed tomography and magnetic resonance imaging showed no structural brain abnormalities. His karyotype was normal 46, XY. Whole exome sequencing did not show candidate variants that can explain his symptoms of overgrowth with eye abnormalities, thrombocytopenia, and encephalopathy. Cytogenomic microarray was subsequently performed which revealed an 8.4

Mb duplication of 11p15.4p15.5 and a 6.3 Mb deletion of 11q24.3q25. The duplication and deletion were confirmed to be a de novo occurrence by fluorescence in situ hybridization in the parents and the child. The patient's encephalopathy was gradually improved by the corrected age of 6 months. His overgrowth became more prominent over time, with the body weight over 2-4SD. His neurodevelopmental ability has been delayed in all aspects. There has been a handful case reports of the coexistence of 11p duplication and 11q deletion and all those cases were a result of pericentric inversion in one of the parents. The present patient represents the first case with expanding phenotypes of neonatal encephalopathy in 11p duplication with 11q deletion, as well as the first patient with a de novo event of this rare chromosomal abnormalities. The possible mechanisms explaining the apparently de novo chromosomal rearrangement in this patient may include the breakage of dicentric formation occurring during meiosis in one of the parental gametogenesis and gonadal mosaicism of pericentric inversion in one of the parents.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2052T: Laboratory experience with primary immunodeficiency panel testing

Authors:

J. Shaw, J. Zdrodowski, Z. Chen; Fulgent Genetics, El Monte, CA

Abstract:

Primary immunodeficiency diseases (PIDs) include a variety of genetic disorders that lead to an increased risk for infectious disease and non-infectious issues such as allergies, malignancies, and autoimmune diseases. The United States PID registry has estimated a prevalence of 1 in 1,200 people. The spectrum of PIDs is growing as we understand PIDs are under an umbrella of inborn errors of immunity and novel genes continue to be discovered. There are targeted treatments available for certain conditions and specific infection rates and comorbidities associated with specific genetic etiologies. Therefore, it is crucial to accurately molecularly diagnose affected individuals, which can lead to medical management changes. Few large-scale studies have reported the diagnostic yield of genetic testing for PIDs. Most of these studies focus on targeted populations of individuals meeting clinical criteria for a specific PID, like CVID, instead of a population with a broad array of PID symptoms. To better understand and treat people with PID it is useful to identify the molecular diagnostic rate for a population of affected people.

We performed a retrospective analysis of PID panel results analyzed by a commercial laboratory for samples received between 10/11/2018 and 3/14/2024. The PID panel included 471 genes

from 10/11/2018 until 11/20/2023, after which the panel included 592 genes.
A total of 7,690 cases were analyzed with positive findings reported in 1,312 individuals for a diagnostic yield of 17%. We identified 965 pathogenic variants and 466 likely pathogenic variants. Diagnostic results include genes associated with autosomal dominant, autosomal recessive, and X-linked inheritance patterns.
PIDs are a common group of disorders that have targeted treatments. It is vital for affected individuals to have an accurate molecular diagnosis to aid in medical management. With broader testing of individuals with symptoms of PID, the diagnostic yield will be lower than for a targeted population meeting clinical criteria guidelines for a specific PID. As more PID genes are discovered we expect the diagnostic yield to increase. Limitations of diagnostic yield based on laboratory experience include lack of clinical documentation of phenotype. It is impossible for commercial laboratories to filter panel results based on a patient's clinical features without complete patient notes, which are often unavailable. The diagnostic yield of PID provides information to be used in pre-test counseling and can aid us in better understanding these conditions and identifying targeted treatments for specific molecular etiologies.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2053T: Large-scale whole-genome sequencing analysis reveals a higher than expected prevalence of C9orf72 repeat expansion alleles associated with specific haplotype structure in the general population.

Authors:

R. Ophoff, B. Duymayan, L. Zhan; Univ California Los Angeles, Los Angeles, CA

Abstract:

An expansion of a hexanucleotide repeat in the C9orf72 gene is the most common genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Alleles with ≥ 30 repeat units in the C9orf72 gene are considered pathogenic. Families with carriers have also been reported to have high rates of schizophrenia, psychosis, suicide, and ASD. In this study, we analyzed two large autism spectrum disorder (ASD) parents-offspring cohorts with WGS data: the Simons Simplex Collection (SSC) cohort with over 9,000 individuals (~2,500 nuclear pedigrees), and the Simons Foundation Powering Autism Research for Knowledge (SPARK) cohort with 11,000 individuals (~3,500 nuclear pedigrees). We used the computational tool ExpansionHunter to call repeat alleles in the C9orf72 gene and perform accurate genotyping. A total of 24 and 26 pathogenic mutation carriers were identified in the SSC and the SPARK cohort, respectively, suggesting a prevalence rate of ~0.25% in the population. We observed no

evidence of de novo expansion events for the hexanucleotide repeat in >10,000 meioses. The pathogenic range allele carriers were validated by repeat-primed PCR. To assess the association between pathogenic expansion and ASD, we performed Fisher's exact test and observed no significant association. Furthermore, we performed linkage/association testing leveraging the pedigree structure within these two cohorts to identify the extent of variants associated with this repeat expansion. Treating the repeat length as a quantitative trait, we observed a single significant peak spanning 133 kb encompassing the locus and a haplotype with 211 variants shared across all 24 carriers in the SSC cohort. In brief, our study is one of the largest studies to explore the distribution of the repeat expansion in the C9orf27 locus using WGS data in a non-ALS/FTD cohort. Our work provides an up-to-date estimate of the population-level prevalence of the pathogenic repeat length alleles, and the lack of de novo expansions observed, and demonstrates a short haplotype strongly associated with the extended alleles in the general population. Future studies will focus on exploring the evolutionary history of this haplotype among the different ancestries and the connection of the expansion to other diseases.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2054T: *Lat* ^{Y136F} mutant mice as a preclinical model of IgG4-RD enable to assess the efficacy and safety of therapies *in vivo*.

Authors:

A. Golovko[1], Y. Li[1], S. Li[1], K. Zhou[1], D. Feng[1,2], R. Sun[1]; [1]GenoBioTX, LLC, Sugar Land, TX, [2]Foster High Sch., Rosenberg, TX

Abstract:

IgG4-related disease (IgG4-RD) is a systemic immune-mediated disease characterized by multi-organ enlargement, storiform fibrosis, and lymphoplasmacytic inflammatory infiltrate, accompanied by symptoms such as obliterative phlebitis and pancreatic failure. Although the exact pathophysiological role remains unclear, many studies have shown that IgG4-RD is associated with over-activation of B cells and CD4+T cells, as well as elevated serum IgG4 concentrations. The incidence of IgG4-RD in Japan is reported to be 0.28-1.08/100,000 people. While insidious progression can lead to irreversible multi-organ damage, treatment strategies exist.

Glucocorticoids, which are steroids, are the first-line treatment for IgG4-RD. Other options include conventional steroid-sparing drugs and B cell depletion therapy. These therapies can effectively improve symptoms such as tissue fibrosis, but limitations include recurrence, side effects, and poor efficacy in some cases. Therefore, further research on IgG4-RD is crucial to

develop more targeted treatment strategies.
To this end, we developed a linker for activation of T cell (Lat^{Y136F}) mutant mouse model. This loss-of-function mutation of LAT mimics the clinical phenotype of IgG4-RD. Compared with wild-type mice, Lat^{Y136F} mice showed differences in the gross appearance of spleen and lymph nodes. Flow cytometry revealed significantly increased proportions of CD4+ T cells and MHC class II+ $B220^{low}$ antibody-producing B cells in multiple tissues of Lat^{Y136F} mice. Additionally, pathological analysis indicated inflammation and fibrosis in multiple organs of the mutant mice. Furthermore, high levels of IgE and IgG1 in their circulation mirrored representative phenotypes observed in IgG4-RD patients.
Finally, Lat^{Y136F} mice were administered with 20 mg/kg Prednisolone, a first-line glucocorticoid, three times a week to evaluate the in vivo response. Prednisolone treatment effectively improved the pathological scores of multiple organs and significantly reduced circulating IgG1 levels in these mice. Overall, the Lat^{Y136F} mouse model provides a powerful preclinical tool to investigate the mechanisms of IgG4-RD and predict the efficacy and safety of potential treatments.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2055T: Leveraging population datasets to identify highly penetrant variants in monogenic obesity genes ⭐

Authors:

N. Chami[1], M. Preuss[1], R. Smit[1], G. Carrasquilla[2], V. Diez-Obrero[2], R. Loos[2]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Copenhagen, Denmark

Abstract:

Background:Obesity is a global health burden and a major risk factor for chronic diseases. Monogenic obesity is caused by rare variations in a single gene - mainly identified through small-scale family-based studies. The penetrance of these mutations at a population level is unknown and requires very large datasets. Leveraging data of the UK Biobank (N=500k), we aimed to characterize rare variants in known monogenic obesity genes. Methods:We extracted rare protein truncating (PTV) or missense variants (MAF<0.01%) from 6 known monogenic obesity genes: *MC4R*, *LEP*, *LEPR*, *PCSK1*, *POMC*, and *SH2B1*. For each variant, we calculated the penetrance (inc. all weight categories = [# carriers with obesity/total # of carriers] × 100) and effect (OR) on obesity (body mass index [BMI] ≥ 30 kg/m^2) defined by the odds of having obesity to the odds of having normal weight.As such, we identified 252 “high impact” variants, defined

by having high penetrance (≥ 30%) and effect on obesity (OR ≥ 2). We replicated the variants in 30,000 individuals from BioMe, Mount Sinai's EHR-linked biobank. We calculated a weighted polygenic risk score for BMI (PRS_{BMI}) in BioMe based on the largest BMI GWAS study.Results: More than half (55%) of the rare PTVs and missense variants in the 6 monogenic obesity genes had 0% penetrance in the UK Biobank and only 19% had high penetrance and effect on obesity and therefore belonged to the "high impact" subset. In the replication dataset, BioMe, carriers of high-impact variants were more likely to have obesity (51%) compared to non-carriers (35%) (P= 0.002). Consistently, carrier status was associated with obesity (OR=2, P=0.01) & extreme obesity (OR=2.3, P=0.003) in a sex and age-adjusted logistic regression model, further validating the pathogenicity of these variants. In contrast, rare PTVs and missense variants that we rendered as low impact in the UK Biobank were not associated with obesity (OR=0.87, P=0.32) nor extreme obesity (OR=0.89, P=0.24) in BioMe. Despite carrying high impact variants in monogenic obesity genes, those with a low PRS_{BMI} had lower prevalence of obesity (53% vs 69% P=0.006) and lower BMI (28.8 vs 33.4 P=0.007) compared to carriers of high impact variants and high $PRS_{BMI.}$Conclusions: Previous estimates of penetrance of rare variants in monogenic genes are inflated. Using large scale population studies, we distinguished highly penetrant variants from non-penetrant and provide a catalog of penetrance of rare variants in 6 monogenic genes. Further, despite being "high impact" variants by design, the penetrance of those variants was modified by a polygenic contribution. These results have implications for clinical testing and genomic medicine.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2056T: Lung sensitizing monogenic variant carrier frequency in idiopathic bronchiectasis

Authors:

J. Berghout[1,2], A. Oler[1,2], J. B. Lack[1,3], W. Cao[1,3], E. Karlins[1,2], R. Ghosh[1,4], B. Seifert[1,4], M. N. Similuk[1,4], K. Kumar[1], A. J. Lipton[1], K. P. Fennelly[1], K. N. Olivier[5], S. M. Holland[1], M. A. Walkiewicz[1,4], J. Yan[1,4]; [1]Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]Bioinformatics and Computational BioSci.s Branch, Office of Cyber Infrastructure and Computational Biology, Bethesda, MD, [3]Res. Technology Branch, Bethesda, MD, [4]NIAID Centralized Sequencing Program, Bethesda, MD, [5]Univ. of North Carolina, Chapel Hill, NC

Abstract:

Background

Bronchiectasis is a persistent inflammation and dilation of lung airways that increases

susceptibility to respiratory infection and can cause chronic cough and sputum production. Patients with cystic fibrosis (CF) disease, primary ciliary dyskinesia, and several other rare genetic conditions commonly develop bronchiectasis as their condition progresses, however bronchiectasis can also present idiopathically. Non-CF bronchiectasis occurs most commonly in adults of European ancestry where prevalence has been reported as 6-36 cases per 10,000 people. Using a cohort of 299 idiopathic bronchiectasis participants, we conducted computational genetic analyses determining the heterozygous carrier frequency of pathogenic variants across a panel of candidate genes, focusing on those where recessive inheritance causes a disorder with bronchiectasis as one of the features
(*CFTR*, *DNAI1*, *DNAH5*, *SERPINA1* and others).

Methods

All affected participants were enrolled through the National Institute of Allergy and Infectious Disease (NIAID) Centralized Sequencing Program (CSP). Each received a clinical evaluation and exome or genome sequencing with harmonized variant calling. Two control cohorts were assembled for comparison. The first was 765 unrelated, ancestry-matched non-proband CSP participants, excluding any families where structured phenotypes or diagnoses included bronchiectasis-related conditions. The second control cohort was a Parkinson's disease cohort with exome sequencing on the same platform. Observed carrier frequencies in these cohorts were compared using a beta binomial posterior model with 50 Markov Chain Monte Carlo draws.

Results:

Preliminary results show that the idiopathic, non-cystic fibrosis bronchiectasis cohort is enriched for the presence of heterozygous *CFTR*:p.Phe508del, with this variant observed in 5% of case participants (90% credible interval: 3-8%) compared to 3% (90% CI: 2-4%) in CSP controls, yielding an odds ratio of 2.17 (+/- 1.0). Presence of any pathogenic/likely pathogenic variant in *CFTR* aligning to CFTR2 curation had a cumulative minor allele frequency of 5% (CI: 3-7%) in cases compared to 3% (CI:2-4%) in controls.

Conclusion

These data provide support for the model of idiopathic bronchiectasis being associated with elevated rates for *CFTR* pathogenic variant carrier status, and ongoing work will determine if this generalizes to include elevated carrier rates for our other candidate genes. Together, these may help to explain a fraction of the etiology and risk for idiopathic cases.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2057T: Lysinuric protein intolerance: Clinical manifestations in a compound heterozygote with a novelpathogenic variant

Authors:

R. VanCoillie, J. Pascual, L. Konczal, C-H. Wu; Univ. Hosp., Cleveland, OH

Abstract:

A 10-year-old male that had a history of malnourishment and growth delay with a weight of 20.2kg (per centile <0.01%, Z= -3.84) and height of 121.9 cm (percentile 0.16%, Z= -2.95) presented to the hospital due to somnolence and was found to have mild transaminitis and hyperammonemia that peaked at 350 umol/L . Other labs such as metabolic panel and complete blood count were not remarkable for any acute process. He was born to non-consanguineous parents and has Russian, Eastern European, and Western European descent. There is no family history of any genetic or metabolic disease. Physical exam was most remarkable for a patient with diminished muscle mass and small stature for his age. Bone age was found to be two standard deviations below the mean. His previous workups revealed low GH demonstrated by low IGFBP and IGF. Initial emergent management consisted of IV ammonia scavenger, 10% dextrose at 1.5 times maintenance, and intralipids at 1g/kg. For his dietary management, protein intake was titrated to 50% of daily required intake considering his weight at the time. His neurological status improved with both interventions. Because the patient had a very limited diet, suspected to be secondary to an aversion to protein, he eventually needed a gastrostomy tube to meet the suggested daily protein requirements. Initial metabolic workup consisted of plasma and urine amino acids, urine orotic acid, urine organic acids, plasma carnitine, and plasma acylcarnitine profile. Plasma carnitine was remarkable for low total and free carnitine levels. Plasma amino acids were most remarkable for normal arginine and citrulline values. Orotic acid was elevated (520 mmol/mol, ref <1.5mmol/mol). Urine amino acids revealed elevated arginine (2278umol/g, ref <16umol/g Cr), ornithine (180 umol/g Cr, ref <30 umol/g Cr), and lysine (15213 umol/g, ref 25-440 umol/g Cr). Molecular testing for urea cycle disorders was sent which revealed two heterozygous pathogenic variants, SLC7A7 c.1383_1384del (p.Ile461Metfs*6) and SLC7A7 c.726G>A (p.Trp242*). The c.1383_1384del variant has not been published previously. The clinical and laboratory findings together suggested the diagnosis of lysinuric protein intolerance (LPI). The patient has since been managed with an ammonia scavenger, diet modification, and supplementation of citrulline, lysine, arginine and carnitine. Six months after diagnosis and treatment, his weight has greatly improved with a value of 30kg (13th percentile, Z= -1.12). His linear growth has not improved at the same velocity with a current value of 27 cm (0.71 percentile, Z= -2.45). Further assessment will be made for other possible interventions such as GH treatment.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2058T: Mendelian variation in a Mexican early onset psychosis cohort analyzed using an AI interpretation algorithm

Authors:

H. Nicolini[1], C. Brownstein[2], N. Lanzagorta[1], C. Bustamante[3], L. Almasy[4], E. Deaso[2], N. Ioannidis[5], S. Moreno[6], L. Cadavid[2], S. Levy[7], M. Cormier[8], M. Yandell[9], K. Green[7], C. Quinto[6], B. Krajacich[7], B. Lajoie[7], C. Walsh[2], M. Reese[10], D. Glahn[11]; [1]Carracci Med. Group, Mexico City, Mexico, [2]Boston Children's Hosp., Boston, MA, [3]Galatea Bio, Hialeah, FL, [4]Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Univ. of California, Berkeley, Berkeley, CA, [6]Galatea, Miami, FL, [7]Element BioSci.s, San Diego, CA, [8]Univ. of Utah - Sch. of Med., Salt Lake City, UT, [9]Univ. of Utah, Salt Lake City, UT, [10]Fabric Genomics, Oakland, CA, [11]Boston Children's Hosp. & Harvard Med. Sch., Boston, MA

Abstract:

Background: Early onset psychosis (EOP, onset prior to 18 years) is an extreme phenotype, with higher rates of rare deleterious mutations in EOP than adult-onset psychosis. Consequently, studying EOP cohorts provides a unique opportunity to discover rare genetic loci influencing illness risk. To date, most psychiatric genetic studies focus on European-ancestry (EA) cohorts, while excluding other ancestry groups. Latinos make up ~8% of the world population (~18% of the US population) but appear in less than 1% of published genome-wide studies. Additionally, Latinos are genetically heterogeneous, with substantial differences between Central and South American and Caribbean populations, reflecting continental-level ancestral group admixture and the substructure of local Indigenous American populations. As 62% of the Latinos in the US are of Mexican origin, findings from the Mexican population are directly relevant for the nation's largest racial/ethnic minority **Methods:** 867 Children, adolescents and their families were recruited for the research study from a single, large public pediatric psychiatric hospital in Mexico City. Whole genome sequencing was performed, followed by analysis of the first 24 genomes using an AI interpretation algorithm. Variants classified as "Likely pathogenic" or "Pathogenic" using ACMG criteria were considered as candidates, along with known structural variants with associated behavioral anomalies. **Results:** From a preliminary screen of 24 patient genomes, 8 (33%) had a SNP, CNV, or SV with an associated neuropsychiatric phenotype or an elevated risk for neuropsychiatric disease. Two individuals had private SNPs in genes linked to developmental delay (*CUL3* and *TAOK1*). One individual had a private variant in a gene known to convey an increased risk of schizophrenia (*TRIO*). One individual had a CNV loss associated with neuropsychiatric syndromes (*NRXN1*). Four individuals had genomic variation associated with

syndromes which have an elevated incidence of developmental and mood disorders (*CPOX*, Xp22.31 deletion, XXY, and *LZTR1*).**Discussion:** The first genomes from this investigation of early onset psychosis cases in Mexico reveals a high yield of structural variation and alterations in known and candidate neuropsychiatric genes.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2059T: Mitochondrial variant calling from genome sequencing identifies new diagnoses in a rare disease cohort

Authors:

I. De Dios[1], C. Xiao[2], G. Pitsava[1], A. Ko[3], V. A. Fusaro[1], L. Auriga[1], J. Lotempio[4], R. Barrick[5], C. H. King[1], M. Almalvez[1], S. Marmalejos[3], UCI-GREGoR Consortium, E. Delot[6], E. Vilain[1], S. Berger[7]; [1]Univ. of California, Irvine, Irvine, CA, [2]Univ. of California, Irvine, Orange, CA, [3]Children's Natl. Res. Inst., Washington, DC, [4]Univ. of California, Irvine, Grand Island, NY, [5]Children's Hosp. of Orange County, Orange, CA, [6]Children's Natl. Res. and Innovation Campus, Washington, DC, [7]Childrens Natl. Res. Inst., Washington, DC

Abstract:

Background:
Expanding sequencing from exome to genome (WGS) has allowed identification of a wider range of variants. While mitochondrial data exist in the same datasets, analysis typically requires a separate bioinformatics workflow, representing a missed diagnostic opportunity.
Objective:
To assess the diagnostic yield of mitochondrial variant calling of extant genomic data in a phenotypically diverse rare disease cohort.
Methods:
PCR-free WGS (minimum coverage 20X) from whole blood or buccal swab was performed on a cohort of 669 participants in the UCI-GREGoR Cohort representing 251 probands and 418 relatives with a wide range of clinical presentations. We ran the GATK's best-practices mitochondrial-variant-pipeline, followed by variant annotation with gnomAD, ClinVar, MitoMap, and APOGEE 2. Variants reported as pathogenic or likely pathogenic (P/LP) in ClinVar were evaluated for contribution to the phenotypes.
Results:
Average coverage of the mitochondrial genome was 7810x. After filtering based on population frequency and variant effects (to exclude synonymous and intergenic variants), 6 distinct P/LP variants were found in 9 probands at very varied levels of heteroplasmy (mean: 9.0%, median:

1.1%, standard deviation 21.5%).
Discussion:
Four samples carried the same chrM:3243 A>G variant in MT-TL1, which is well characterized in patients with hearing loss. Only 2 samples had levels of heteroplasmy above 5% and both patients had phenotypic features consistent with mitochondrial disease.
The variant at highest heteroplasmy (70%) was found in the blood of a patient who previously only had a partial molecular diagnosis. This new finding completed the molecular diagnosis for his additional symptoms (including hearing loss) not previously explained by his holoprosencephaly diagnosis.
In the saliva of a patient with a phenotype compatible with either immunodeficiency syndrome or mitochondrial disorder, the variant was found at 10% heteroplasmy, providing a definitive diagnosis. Variant and heteroplasmy level were confirmed by a CLIA lab on a fresh specimen.
Conclusion:
Adding mitochondrial variant calling into clinical and research genome analysis pipelines requires minimal additional bioinformatics and analysis burden while increasing ability to diagnose rare diseases.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2060T: Molecular characterization of the pathogenic exon 13-15 duplication in *PHEX*: investigating the intervals in families with and without the *231A-G variant

Authors:

K. Cabral[1], C. Brownstein[1], N. Jastromb[2], K. Schreiber[3], J. Soto Barros[4], J. Phoon[1], N. Edisis[1], S. Duan[5], L. Schneider[6], J. Sage[7], G. Gottesman[8], J. Thompson[6], S. Mumm[9], T. Carpenter[4]; [1]Boston Children's Hosp., Boston, MA, [2]Nabsys Inc, Providence, RI, [3]Bionano Genomics, San Diego, CA, [4]Yale Univ., New Haven, CT, [5]Div. of Bone and Mineral Diseases, Washington Univ. Sch. of Med., St. Louis, MO, [6]Nabsys, Inc, Providence, RI, [7]Nabsys, Providence, RI, [8]Shriners Hosp. for Children, Tampa, FL, [9]Washington Univ. Sch. of Med., Saint Louis, MO

Abstract:

Introduction: X-linked hypophosphatemia is the most common cause of inherited rickets, due to inactivating variants of *PHEX*. More than 800 variants have been described to date and one which consists of a single base change in the 3' untranslated region (UTR) (c.*231A>G) is reported as prevalent in North America. Recently an exon 13-15 duplication has been found to occur in concert with the c.*231A>G variant, and thus it is unclear whether the pathogenicity is

solely a function of the UTR variant. We present a family with XLH who harbors the exon 13-15 duplication as identified on Bionano Saphyr but does not carry the 3'UTR variant (as seen on next generation sequencing (NGS)). This family provides evidence that the duplication itself is a pathogenic variant when these two variants are found in cis, and that the duplication can be missed by NGS. However, it remained unclear whether the duplication was identical to the recurrent duplication. **Methods:** Ultra-high molecular weight DNA was nicked with BspQI and BssSI and labeled at the nick sites using Klenow and modified nucleotides. The modified nucleotides were tagged and the complex coated with RecA protein. This DNA was injected into a Nabsys OhmX instrument with a 256-channel detector. Voltage versus time measurements were made and processed into distances between tags in base pairs. Molecules were assembled into contigs and the resulting contigs aligned to the hg38 reference. Structural variants in *PHEX* were examined. The result was confirmed by Sanger sequencing using primers designed to cross the unique breakpoint. **Results:** A large insertion was identified; competitive remapping revealed a 56 kb tandem duplication spanning chrX:22,136,878-22,192,883 (hg38). Sanger sequencing confirmed that the unique exon 15-13 breakpoint was identical to previously sequenced cases of the exon 13-15 duplication with the c.*231A>G variant.**Conclusions:** This family provides evidence that this precise duplication itself is a pathogenic variant when found in cis with the c.*231A>G variant. In addition, Nabsys and Bionano Saphyr provide accurate assessment of structural variation potentially missed by Next Generation Sequencing.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2061T: *MSH3* is a modifier of repeat expansion and age at onset in X-linked dystonia parkinsonism (XDP)

Authors:

A. Mejia Maza; Massachusetts Gen. Hosp., Boston, MA

Abstract:

The insertion of a SINE-VNTR-Alu (SVA) retrotransposon in the TATA-box binding protein associated factor-1 (TAF1) gene causes X-linked dystonia-parkinsonism (XDP), whose age-at-onset (AAO) is inversely correlated with the length of a polymorphic CCCTCT repeat within the SVA. Notably, the inverse correlation of CCCTCT repeat length with AAO places XDP among a larger class of neurodegenerative diseases caused by expanded microsatellite repeat tracts, e.g. Huntington's disease (HD), which are characterized by the instability of their repeat mutations. We have previously shown that the XDP CCCTCT repeat exhibits somatic expansion that is tissue-specific and repeat length-dependent, consistent with somatic expansion as a mechanism

underlying the CCCTCT length-dependent AAO. Here, we show that genetic variation in MSH3 - single nucleotide variants and a complex polymorphic coding repeat - is associated with both somatic expansions, measured in blood DNA from XDP patients, and AAO. Variants associated with less expansion in blood are associated with a later AAO, and vice versa. This provides the first direct evidence that MSH3 is a modifier of repeat expansion in XDP and strong support for somatic expansion driving the rate of disease onset in XDP as it does in HD. These results contribute to our overall understanding of a common role of somatic instability in repeat expansion diseases and highlight the role of MSH3 as a significant disease modifier and important therapeutic target across this class of disorders.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2062T: Multi-omic Sequencing Approaches Identify Molecular Causes of Genetically Unsolved Developmental and Epileptic Encephalopathies

Authors:

C. Laflamme[1], S. Sengupta[1], E. Almanza Fuerte[1], E. Bonkowski[1], T. Allan[2], N. Lieffering[2], L. Sadleir[3], I. Scheffer[2], H. Mefford[1]; [1]St. Jude Children's Res. Hosp., Memphis, TN, [2]Univ. of Melbourne, Victoria, Australia, [3]Univ. of Otago, Wellington, New Zealand

Abstract:

Background: Despite advances in sequencing technologies, about half of individuals with developmental and epileptic encephalopathies (DEEs) remain without a genetic diagnosis. Here, we sought to investigate the utility of multi-omic sequencing modalities to uncover genetic etiologies. We leveraged our growing overlapping datasets from individuals with DEEs who previously underwent extensive screening (e.g. gene panel, microarray, exome sequencing) but remained genetically unsolved. **Approach**: We analyzed data from n=222 short-read trio genomes (proband and both parents), n=17 proband long-read genomes, and n=25 proband fibroblast or human cortical spheroid RNA-seq data. To detect causal and candidate etiologies from genome sequencing data, we employed a comprehensive suite of tools, including GATK and Rufus for SNV/Indel calling; Rufus, Delly, and Lumpy for structural variant detection; CNVnator for CNV identification; ExpansionHunter for repeat expansion profiling; MELT for SINE/LINE/SVA element detection; and OUTRIDER and rMATS for RNA-seq expression outlier and splicing analyses, respectively. **Results**: This yielded three *de novo* SNVs/Indels, two *de novo* SVs (one large inversion and one translocation across two different chromosomes), six *de novo* CNVs, and four inherited and expanded STRs. These variants encompassed both

established epilepsy genes and new candidate genes, and most would not have been detected by exome sequencing. Parental sequencing data was invaluable for prioritizing *de novo* and compound heterozygous calls. In many cases, overlapping multi-omic data was necessary to validate DNA variants. For instance, repeat expansions predicted from short-read sequencing were validated using long-read sequencing, which each produced more accurate length estimates and motif sequences compared to short-read. Analysis of RNA-seq data for gene dropout identified potential disease-causing gene expression abnormalities in two cases. **Conclusion**: Our study demonstrates the current diagnostic utility of multi-omic sequencing analyses in identifying molecular etiologies in DEEs. These findings advance our understanding of the genetic underpinnings of unsolved cases and inform a wealth of other rare neurodevelopmental disorders where a significant portion of individuals remain etiologically unresolved.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2063T: *NKX6-2*- related leukodystrophy: natural history and novel genetic therapeutic approaches

Authors:

K. Zhelcheska[1], S. Basit[2], V. M. Srinivasan[3], V. K. Gowda[3], V. Chelban[1], H. Houlden[1]; [1]UCL Queen Square Inst. of Neurology, London, United Kingdom, [2]Taibah Univ., Madinah, Saudi Arabia, [3]Indira Gandhi Inst. of Child Hlth., Bangalore, Karnataka, India

Abstract:

Background/ Objectives: Hypomyelinating leukodystrophies are a heterogeneous group of genetic disorders characterized by impaired formation of the myelin sheath in the CNS, leading to a wide spectrum of phenotypes. Bi-allelic variants in the *NKX6-2* gene were identified to cause spastic ataxia 8 (SPAX8), autosomal recessive, with hypomyelinating leukodystrophy. There are currently no disease-modifying treatments nor biomarkers available for *NKX6-2* patients, where only supportive and palliative care is available.**Methods:** We conducted a natural history study involving affected individuals with *NKX6-2*-related leukodystrophy from unrelated families to assess the progression of the disease. Immunocytochemical staining of NKX6-2 protein in *NKX6-2* patient-derived fibroblasts was conducted to evaluate the consequence of pathogenic variants on NKX6-2 expression compared to unaffected healthy controls.**Results:** We have identified affected individuals from families worldwide harbouring pathogenic variants in *NKX6-2*. Our findings reveal there is variation in phenotypic severity, both according to differences in the onset of disease and the specific disease-causing

variant. **Conclusion:** *NKX6-2*- related leukodystrophy is an untreatable neurological disorder and there is clinical need to develop new treatment strategies for this disorder. Our natural history study will be also used to identify serial biomarkers in patients with *NKX6-2*- related leukodystrophy, helping to provide outcome measures for potential future trials. Furthermore, we aim to use *in vitro* models to develop novel genetic therapeutic approaches to replace and rescue NKX6-2 protein loss.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2065T: Novel candidate gene associated in Dilated Cardiomyopathy (DCM) in a Malian family

Authors:

H. Ba[1], M. Diakité[2], S. Bamba[1,3], S. O. Koudoussou[1], O. Traore[1], S. Diarra[1,4], M. Goita[5], A. Maiga[1], A. Yalcouyé[6], M. Traoré[1], C. Guinto[1], M. Khokha[4], W. Ji[4], S. Lakhani[4], G. Landoure[1], O. Samassekou[7]; [1]Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, [2]Service de Cardiologie, Ctr. Hosp.ier Univ.ire, Bamako, Mali, [3]Pediatric Genomics Discovery Program (PGDP), Dept. of Pediatrics, Yale Univ. Sch. of Med., New Haven, CT, [4]Yale Univ., New Haven, CT, [5]Univ. of Sci. Techniques and Technologies of Bamako, Bamako, Mali, [6]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [7]Univ. of Sci., Technics and Technologies, Bamako, Mali

Abstract:

Introduction: Dilated cardiomyopathy (DCM) is a severe cardiac disorder characterized by ventricular dilation and impaired systolic function, often leading to heart failure. Despite advances in genomics, many DCM cases lack a precise genetic diagnosis. To date, 19 genes are known to cause DCM with high levels of evidence. Here, we investigate a putative candidate gene, NRXN3 found in a Malian family with DCM.
Methods: We obtained ethical approval and consent from participants. Cardiac assessments including blood chemistries, chest X-ray, and ECG were performed. DNA was extracted for genetic analysis, including exome sequencing (ES) and Sanger sequencing. Deleteriousness of putative variants were assessed using *in silico* tools.
Results: A 15-year-old boy from a non-consanguineous family was seen for palpitations, NYHA stage III dyspnea, productive cough with blood-tinged sputum, and retro-sternal pain. Cardiac examination showed gallop heart sounds, mitral insufficiency-type regurgitation, hepato-jugular reflux, facial puffiness, decreased breath sounds in the lower right lung, lower limb edema, and arterial hypotension at 12 years. Chest X-ray revealed cardiomegaly with vascular redistribution, while resting ECG showed regular sinus rhythm at 125 bpm with left atrial hypertrophy and

inferolateral repolarization disorder. ES identified a compound heterozygous variant in NRXN3 (p.Arg84His and p.Pro170Leu). These variants are predicted to be deleterious by several prediction tools with CADD 20.3 and 24.1 respectively.
Discussion: NRXN3 is not a known DCM gene. It is expressed in various tissues, including the heart. Recent findings suggest its involvement in cardiac function. Transcriptomic analysis implicates NRXN3 in cardiac fibrosis, a DCM-related condition. Spatial transcriptomic data confirm NRXN3 activity in patient tissue, highlighting its potential role in cardiac fibrosis progression. Further investigation into its potential contribution to DCM pathogenesis is needed.
Next step: Additional family members recruitments for segregation analysis and functional studies using Xenopus tropicalis models are ongoing to confirm the pathogenicity of the NRXN3 variant.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2066T: Partial methylation of a pathogenic *XYLT1* repeat expansion associated with intrafamilial variation in severity of Desbuquois dysplasia 2

Authors:

W. Gordon[1], J. Devaney[2], D. L. Earl[3], K. J. Buckingham[1], C. T. Marvin[1], H. C. Mefford[4], C. Laflamme[4], C-L. Wei[5], **M. J. Bamshad**[1,3,5,6], J. X. Chong[1,6], University of Washington Center for Rare Disease Research; [1]Dept. of Pediatrics, Univ. of Washington, Seattle, WA, [2]GeneDx, Gaithersburg, MD, [3]Seattle Children's Hosp., Seattle, WA, [4]St. Jude Children's Res. Hosp., Memphis, TN, [5]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [6]Brotman Baty Inst. for Precision Med., Seattle, WA

Abstract:

Baratela-Scott syndrome, *XYLT1*-related (BSS), formerly known as Desbuquois dysplasia type 2, is an autosomal recessive skeletal dysplasia due to biallelic variants in *XYLT1*. It is characterized by severe pre- and postnatal growth retardation, joint laxity, a flattened midface with broad nasal bridge, and patella dislocations. Some affected individuals exhibit developmental delay. We studied a family with two affected siblings with a suspected diagnosis of BSS. The brother (i.e. proband) is severely affected, with short stature (-6.5-7 standard deviations [SD]) and limbs, pectus carinatum, knee dislocation, carpal and tarsal ossification, developmental delays, and a flat midface with wide-set eyes and a depressed nasal root. His sister exhibits a milder phenotype including short stature (-6 SD), skeletal asymmetry, persistent growth deficiency, and

hypothyroidism. A third sibling and both parents are unaffected. Interestingly, the father and one of his sisters (aunt of the affected children) reportedly received growth hormone therapy during childhood (adult height 5'8" and 4'11", respectively). Clinical sequencing of *XYLT1* and a microarray did not identify variants in the gene. The family had research exome sequencing in 2015 that also did not identify any compelling candidate genes or variants. In 2019, we reported the discovery that homozygosity or compound heterozygosity for a $(GGC)_k$ repeat expansion in the 5'UTR of *XYLT1* can trigger hypermethylation of its promoter and silencing of the allele during transcription when $k > 72$ units. Subsequently, we performed short read genome sequencing and used ExpansionHunter to genotype the repeat. The results were suggestive of an expansion, but could not definitively resolve the repeat lengths. We then performed long read genome sequencing (lrGS) on the Revio platform on the proband and sister, which enabled simultaneous determination of both repeat length and methylation status. TRGT called compound heterozygous repeat expansion alleles in both siblings (proband $k = 222, 477$ and sister $k = 207, 586$). While the proband exhibited ~100% methylation at the locus, the sister was only ~56% methylated. We hypothesize that partial methylation in the sister may result in incomplete transcriptional silencing of *XYLT1* and explain her milder presentation. lrGS of additional family members is underway to investigate segregation of the repeat alleles and its possible relationship with the multi-generation paternal family history of short stature.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2067T: Pathogenic variants in ADTKD-*MUC1* can be detected using short-read sequencing data

Authors:

C. Nagano[1], N. Morisada[2], Y. Inoki[1], Y. Tanaka[1], C. Ueda[1], H. Kitakado[1], N. Sakakibara[1], T. Horinouchi[1], T. Yamamura[1], S. Ishimori[1], K. Nozu[1]; [1]Kobe Univ. Graduate Sch. of Med., Kobe, Japan, [2]Hyogo Prefectural Kobe Children's Hosp., Kobe, Japan

Abstract:

Background Autosomal dominant tubulointerstitial kidney disease (ADTKD)-*MUC1* is an autosomal dominant tubulointerstitial nephritis, which leads to end-stage kidney disease at around 50 years of age, with poor urinary findings, and is caused by frameshift variants within variable number of tandem repeats (VNTRs), a region responsible for glycosylation. Because VNTRs are 20-125 repeat sequences consisting of a GC-rich 60 base pair, variants cannot be identified by conventional Sanger method or short read sequencing. Other special methods or long read sequencing are required for diagnosis. However, due to technical and cost issues,

there are no clinically effective diagnostic methods available at this time, and many cases remain undiagnosed. **Methods** Fourteen patients with suspected ADTKD-*MUC1* and identified *MUC1* pathogenic variants by long-read sequencing or other methods at our hospital between 2015 and 2023 were included. We attempted to detect *MUC1* pathogenic variants using the VNtyper-Kestrel pipeline, a method to detect pathogenic variants in the VNTR region by software programming using short-read sequencing data as recently published by Saei et al. First, a hereditary kidney disease panel containing more than 100 genes was used, followed by an ADTKD panel containing six genes (*MUC1*, *UMOD*, *HNF1B*, *REN*, *SEC61A1* and *DNAJB11*) specifically for ADTKD who were clinically and/or pathologically suspected of having ADTKD. **Results** The FASTQ file was used as the input file for the analysis. VNtyper-Kestrel pipeline analysis using the hereditary kidney disease panel, pathogenic variants were detected in 11 of 14 confirmed ADTKD-*MUC1* cases. On the other hand, using the ADTKD panel, pathogenic variants were detected in all cases. The mean coverage of the *MUC1* VNTRs region was 188 in the hereditary kidney disease panel and 1,185 in the ADTKD panel ($p=7.41\times10^{-11}$). The average analysis time per case was 2.2 minutes for both. **Conclusion** By using the ADTKD panel, we could detect *MUC1* pathogenic variants in all cases. Improved coverage is key to effectively using this pipeline to identify *MUC1* pathogenic variants. This analysis method was significantly quicker and less expensive than conventional methods. We hope that using the VNtyper-Kestrel pipeline will facilitate the diagnosis of *MUC1* variants in the future.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2068T: Pathological Short Tandem Repeats Analysis and Imprinting Disorders by Long-Read Sequencing in Affected Individuals

Authors:

H. Mor-Shaked[1,2], B. Rivka[1], F. Ilana[1], L. Shagam[1], S. Shira[1], R. Chaggai[1], O. Elpeleg[1], T. Harel[1,2]; [1]Dept. of Genetics, Hadassah Med. Organization, Israel, Jerusalem, Israel, [2]Faculty of Med., Hebrew Univ. of Jerusalem, Israel, Jerusalem, Israel

Abstract:

Background/Objectives: Pathological short tandem repeats (STRs) are repetitive DNA sequences found throughout the genome prone to expansion and contraction, resulting in various genetic diseases. Traditional short-read sequencing technologies have limitations in detecting STRs due to their inherent instability and repetitive nature. Long-read sequencing (LRS) can provide a potential solution for accurate and comprehensive STR analysis, as well as detection of methylation. Methylation disorders are a group of conditions characterized by

abnormal methylation patterns in DNA, which can disrupt normal gene expression and lead to various disorders. **Methods:** Twenty-four individuals with confirmed pathogenic expansions across ten different loci, along with twenty-nine controls, underwent long-read whole-genome sequencing (LR-WGS). Sixteen affected individuals and twenty-eight controls were sequenced using PromethION (Oxford Nanopore Technologies). STR genotyping was performed using the Straglr software. Straglr provided repeat unit counts from the phased reads that span the collection of pre-annotated disease-associated human STRs. Eight affected individuals and one control individual underwent sequencing on Revio long-read system (Pacific Bioscience). STR genotyping was performed using the TRGT software. In addition, three individuals affected with known methylation disorders were sequenced on the PromethION. **Results:** Both Strglr and TRGT software analyses on LR-WGS data successfully identified all previously known repeat expansions, with approximately the same allele sizes as those reported in repeat-primed PCR, and provided sizing information for large loci where the size was previously unknown. False positive results were detected in twelve out of sixteen individuals for whom PABPN1 gene results were available. In addition, methylation analysis effectively detected all previously identified methylation disorders, and, in some relevant cases, provided supporting evidence for the expansion results. **Conclusion:** As the cost of genome sequencing continues to drop and new technologies emerge that can potentially detect all types of genetic variations in one test, we hypothesize a probable shift towards LR-WGS for clinical STR detection. This is further supported by the ability to analyze DNA methylation from LR-WGS data, which is relevant for several pathological STRs. Additional validation is necessary, which includes testing multiple healthy and affected individuals for each genetic locus.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2069T: PD GENEration: a genetic analysis beyond the classical 7 Parkinson's disease genes

Authors:

I. Keller Sarmiento[1], V. Pitz[2], C. Blauwendraat[3], Z-H. Fang[4], T. Foroud[5], A. Hall[6], K. Marder[7], I. Mata[8], M. Nance[9], M. Schwarzchild[10], T. Simuni[1], A-M. Wills[11], S. Lawrence[12], K. Ghosh Galvelis[6], J. Beck[6], A. Naito[6], R. Alcalay[13], N. Mencacci[1]; [1]Northwestern Univ., Chicago, IL, [2]NIH, Bethesda, MD, [3]NIH, Bethesda, MD, [4]The German center for Neurodegenerative diseases, Tuebingen, Germany, [5]Indiana Univ Sch Med, Indianapolis, IN, [6]Parkinson's Fndn., Miami, FL, [7]8. Columbia Univ., New York City, NY, [8]Cleveland Clinic, Cleveland, OH, [9]Hlth.Partners, Golden Valley, MN, [10]11. Massachusetts Gen. Hosp., Boston, MA, [11]Massachusetts Gen. Hosp., Boston, MA, [12]Navitas Clinical Res., Rockville, MD, [13]Columbia Univ., New York, NY

Abstract:

Objective To expand the analysis of targeted whole-exome sequencing (WES) data from 6034 patients with Parkinson's Disease (PD) from the PD GENEration cohort. We aimed to look for variants in genes associated with rare forms of parkinsonism, beyond the panel containing the 7 classical PD genes that are routinely reported (*LRRK2, GBA1, PRKN, PINK1, SNCA, PINK1, PARK7* and *VPS35*). **Background** PD GENEration is a multi-center study that offers CLIA-certified genetic testing and counseling to PD patients in North America for 7 genes determined to have an established primary role in causing PD. Overall, 13% of the participants carried pathogenic variants in one of these genes, with the majority being associated with *GBA1* (8%, Figure 1). However, pathogenic variants in several other genes are associated with different forms of monogenic parkinsonism, which could cause a presentation mimicking PD. **Methods** Paired-end clinical exome sequencing was performed at Fulgent Genetics, using the Illumina platform 2,500 HiSeq or NovaSeq 6,000. GATK version 4 was employed for alignment and variant calling. Variants were annotated with VEP v.105. We considered rare variants (minor allele frequency less than 0.01). Pathogenicity was determined based on the American College of Medical Genetics (ACMG) criteria. **Results** We curated a panel of 66 additional genes that can cause parkinsonism (Table 1). 29 cases (0.48%) carried pathogenic variants in one of those genes (Table 2). In particular, frontotemporal dementia (FTD)-related dominant genes were the most represented (12 patients - 1x MAPT, 5x GRN, 2x TBK1, 4x VCP1). The other mutated dominant genes
were *GCH1* (n=4), *ATP1A3* (n=3), *OPA1* (n=1), *AFG3L2* (n=1), *WDR45* (n=1), *PDGFB* (n=2), *PDE8B* (n=1). Moreover, we found four patients with potentially bi-allelic *PLA2G6* variants (phase unconfirmed). Three participants (10.3%) reported a first-degree family member affected by PD (2 cases with *GCH1* variants and 1 case with *VCP* variant). Of the 33 variants identified, 13 (38.2%) were loss-of-function. The median age at onset was 54.5 (range 21-84). The majority of the participants were of Caucasian ethnicity (72.4%). **Conclusions** Overall, our study shows that pathogenic variants in a wide spectrum of genes associated with non-classical parkinsonism are found in a small minority of patients clinically diagnosed with PD. The effort to collect detailed clinical information on the variant carriers is ongoing. Broader genetic studies, such as WES or WGS, may be necessary for cases that do not carry pathogenic variants in the 7 classical PD genes.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2070T: Phenotypic expansion and genotype-phenotype correlations in Kleefstra syndrome using data from a large clinical cohort

Authors:

Z. Frazier[1], H. Osika[1], A. O'Donnell-Luria[1], M. Horlbeck[1], L. Pais[1], T. Katz[1], T. Kleefstra[2], S. Srivastava[1]; [1]Boston Childrens Hosp., Boston, MA, [2]ErasmusMC, Rotterdam, Netherlands

Abstract:

Background: Kleefstra syndrome (KS) is a rare genetic condition caused by a loss of function of *EHMT1*. KS is characterized by intellectual disability (ID)/global developmental delay (GDD), hypotonia, distinctive facial features, and congenital cardiac and/or renal defects. Neurological features include seizures, neurobehavioral conditions, and neuropsychiatric conditions. A subset of the post-pubertal KS population may experience a triad of significant insomnia, psychiatric manifestations including psychosis and catatonia, and regression of self-help skills. Objective: Due to the rarity of this condition, a limited number of individuals have been characterized in the literature. Additionally, there exists a gap in our understanding of the features that present in post-pubertal individuals. Our study aimed to refine and expand the full spectrum of KS clinical features and reveal novel phenotypic-genotypic correlations. Design and Methods: We performed a retrospective analysis of 66 individuals with KS (pathogenic/likely pathogenic variant in *EHMT1* or 9q34.3 chromosomal deletion including *EHMT1*) who were clinically evaluated at the Boston Children's Hospital (BCH) KS Clinic. We analyzed demographic and clinical data obtained by clinicians in the BCH KS Clinic. Results: The most prevalent phenotypes included hypotonia (51/66, 77%), ID/GDD (50/66, 76%), autism (26/66, 39%), and constipation (20/66, 30%). Of the post-pubertal individuals, 37.5% (6/16) had neuropsychiatric decompensation/regression. Novel phenotypes included significant motor impairment (10/66, 15%), low-average or borderline IQ (9/66, 13.6%) rather than ID/GDD, cardiac conduction abnormalities (3/66, 4.5%), infantile spasms (6/66, 9.1%), and refractory epilepsy (6/66, 9.1%). Of the individuals with 9q34.3 deletions (n=27), there was a significantly higher rate of ID/GDD (p=0.008), cortical visual impairment (p=0.002), disrupted sleep (p=0.038), and atrial septal defects (p=0.026) compared to those with intragenic *EHMT1* variants (n=39). All but one participant with significant motor impairment had 9q34.3 deletions. All individual with low-average or borderline IQ had intragenic *EHMT1* variants. Conclusion: Our analysis provides significant insights into the phenotypic and genotypic spectrum of the KS patient population seen at BCH. Our results support existing KS literature and highlight several novel phenotypes seen within our patient cohort. Genotype-phenotype correlations may allow better prognostication for individuals with a new diagnosis.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2071T: *POLD3* haploinsufficiency is linked to non syndromic sensorineural late onset progressive hearing and balance impairment.

Authors:

E. Chouery[1], C. Mehawej[1], R. SAADE[1], R. BARAKE[1], P. Zarecki[2], C. Gennery[2], S. Corbani[1], R. Korban[1], A. Hamam[1], J. Nassereldin[1], M. Yamout[3], M. Banna[3], A. Yamout[3], F. Adhami[4], A. Megarbane[1], **M. Mustapha**[2]; [1]Lebanese American Univ., Beirut, Lebanon, [2]Sheffield Univ., Sheffield, United Kingdom, [3]Yamout hearing center, Beirut, Lebanon, [4]Adhami Advanced Audiology Ctr., Tripoli, Lebanon

Abstract:

Age-related hearing impairment (ARHI) is a significant health concern globally, influenced by genetic and environmental factors. We had identified a homozygous pathogenic variant in *POLD3* in a Lebanese patient with an autosomal congenital recessive syndromic hearing loss. This variant was found at heterozygous state in the parents, who developed progressive hearing impairment around age 40. We conducted a thorough clinical and genetic assessment of sixteen family members, including physical exams, audiometry and vestibular function evaluations. Additionally, gene expression analysis of the *Pold3* gene was performed in mice using RNAscope. Twelve individuals were heterozygous for the variant in *POLD3,* of whom nine showed bilateral ARHI, typically starting around ages 40-50, and two older patients displaying unilateral vestibular weakness. Additionally, two carriers of the variant developed cancer at an early age. RNAscope confirmed *Pold3* expression in auditory and vestibular neurons. Exome sequencing analysis excluded the presence of pathogenic variants in any known hearing impairment or cancer predisposition genes. We present herein the first evidence of a heterozygous pathogenic *POLD3* variant causing a novel form of autosomal dominant progressive ARHI and age-related vestibular impairment. We also highlight the necessity for further exploration of the role of *POLD3* in cancer predisposition.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2072T: Rare homozygous variants in *INSR* and *NFXL1* are associated with severe treatment-resistant Psychosis ⭐

Authors:

A. Kanwal[1], R. Zulfiqar[1], H. Cheema[2], N. Jabbar[2], A. Iftikhar[3], A. Butt[1], S. Sheikh[4], N. Pankratz[5], J. Pardo[6], S. Naz[1]; [1]Univ. of the Punjab, lahore, Pakistan, [2]Punjab Inst. of Mental Hlth., lahore, Pakistan, [3]Rainbow Obesity and Eating Disorders Ctr., lahore, Pakistan, [4]Psychiatry Dept., Hawkes Bay DHB, Hastings, New Zealand, [5]5Div. of Computational Pathology, Dept. of Lab. Med. and Pathology, Univ. of Minnesota, Minneapolis, MN, [6]Minneapolis VA Hlth.Care System, Minneapolis, MN

Abstract:

Psychosis, which encompasses hallucinations and delusions, is a fundamental aspect of schizophrenia and also affects approximately half of individuals diagnosed with bipolar disorder. This study was designed to molecularly characterize psychosis in multiple consanguineous families in which affected individuals were suffering from schizophrenia or psychotic bipolar disorder. Eight consanguineous families each with two patients affected by schizophrenia or psychotic bipolar disorder participated. Structured clinical evaluations were performed using different evaluation tools which supported the diagnosis of schizophrenia in twelve patients and psychotic bipolar disorder in four individuals. All other family members were unaffected. DNA was subjected to exome sequencing which revealed none, or ambiguous findings for members of six families. Copy number variant analyses (CNV) of exome data were also performed using GATK pipeline but no pathogenic CNVs were detected which segregated with the disorders. For the remaining two families, homozygous rare, variants in *NFXL1* (c.1322G>A; p.Cys441Tyr) and *INSR* (c.2232-7T>G) were identified for the participants which were validated by Sanger sequencing to segregate with the phenotype of severe treatment-resistant schizophrenia. The predicted deleterious missense variant of *NFXL1* affected an evolutionary conserved amino acid while several *in-silico* tools indicated that the intronic variant in *INSR* may affect splicing. However, cDNA analysis performed on patient's blood sample did not reveal an *INSR* aberrant transcript. Additional experiments are needed in order to determine the effect of this variant. Previously, some *INSR* variants have been reported to be significantly associated with schizophrenia in humans, while knockout mouse models manifest depressive and anxiety-like symptoms. Although a heterozygous *NFXL1* variant has been associated with autism spectrum disorder, other evidence supporting a role in psychiatric disorders is lacking. Our findings indicate a need for further evaluation of *INSR* and *NFXL1* in psychosis. Furthermore, the lack of genetic findings in six families suggests that either variants of large effects do not exist for a great number of treatment resistant psychosis, supporting a polygenic inheritance model, or different strategies such as genome sequencing need to be pursued to identify the causative loci. Continued research will delineate the genetics of these devastating disorders which would ultimately lead to the development of targeted therapies in future.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2073T: Rare variants modulating phenotypic outcome in Neurofibromatosis type 1.

Authors:

E. Pasquinelli[1,2], G. Casamassima[3,2], S. Minetto[1,2], C. Fallerini[4,2], M. Baldassarri[5,6], **A. Renieri**[1,2]; [1]Univ. of Siena, Siena, Italy, [2]Med. Genetics, Univ. of Siena, Siena, Italy, Med Biotech Hub and Competence Centre, Department, Italy, [3]Università di Siena, Copertino, Italy, [4]Univ. OF SIENA, Siena, Italy, [5]Med. Genetic, Siena, Italy, [6]Med. Genetics, Univ. of Siena, Siena, Italy, Siena, Italy

Abstract:

Neurofibromatosis type 1, is a rare genetic disorder predisposing to a pleiotropic phenotype ranging from psychosocial issues, congenital malformations, benign tumors of the nervous systems or aggressive solid cancer. We hypothesize that this phenotypic variability is due to additional rare variants in other genes, acting as an oligogenic threshold model. Clinical and Exome sequencing data were retrieved in a retrospective study from 32 patients with pathogenic/like pathogenic NF1 variants, counseled at the Medical Genetics Unit of University Hospital in Siena, Italy over a period of three years. Genetic variants were first prioritized and selected on the bases of rarity (<0.001 MAF) and classified according to ACMG classification. Gene combination analysis shows that frequently (precocious puberty or short stature) less frequently (macrocephaly or cognitive disorder) or rarely (connective disorder or polydactyly) ancillary characteristics could be linked to rare variants in additional genes. The same is true for facial characteristics e.g. Smith Lemli Opitz-like (DHCR7), Silver Russel-like (H19). In inherited cases, NF1 variant inherited from the father combines with other rare variants inherited from the mothers, contributing to the heterogeneity of the NF1 spectrum, e.g. GLI3 for polydactyly, TSC2 for pigmentary disorder and cognitive impairment. Particularly, short triplet repeats in the AR gene reinforce the NF1 susceptibility to precocious puberty. NF1 carriers with cognitive impairment have a mean of 2.7 additional variants, with at least one of them being a P/LP variant in a well known dominant neurodevelopmental disorder gene. This work illustrates that rare variants in specific functionally related sub-group of genes modulate phenotypic outcome in neurofibromatosis. Rare variants in intellectual disability related genes impact on neurodevelopment and other ones are related to pigmentation of the skin, head size, ligament laxity, puberty and even facial characteristics. These results suggest how a multigenic approach could be relevant for an amelioration of screening programs in NF1 carriers, compared to single-gene testing.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2074T: The Australian Genomic Health Alliance (AGHA) Mitochondrial Flagship: a national approach to delivering mitochondrial disease diagnoses. ⭐

Authors:

J. Christodoulou[1,2,3,4], N. Baker[2,3,4], S. Balasubramaniam[5], K. Bhattacharya[5], D. Bratkovic[6], D. Coman[7,8], A. Compton[1,2,3], M. Delatycki[1,2,3], C. Ellaway[5], M. Fahey[9], J. Fletcher[10], A. Frazier[1,2], R. Ghaoui[10], I. Goranitis[1,2,4], D. Hock[1,2,3,11], M. Kava[12], N. Lake[13,1], P. J. Lamont[12,14], J. Lee[2,15], J. J. McGill[7], J. Panetta[9], L. Phillips[10], R. Rius[1,2], M. Ryan[16], N. Smiith[6], D. Stroud[2,11], M. Tchan[17], D. R. Thorburn[1,2,3,4], M. J. Wallis[18], A. Welch[1], C. Wools[9], Y. Wu[1,2,4]; [1]Murdoch Children's Res. Inst., Melbourne, Australia, [2]Univ. of Melbourne, Melbourne, Australia, [3]Victorian Clinical Genetics Service, Melbourne, Australia, [4]Australian Genomics Hlth.Alliance, Melbourne, Australia, [5]Sydney Children's Hosp. Network, Westmead, Australia, [6]Women's and Children's Hosp., Adelaide, Australia, [7]Queensland Children's Hosp., Brisbane, Australia, [8]Wesley Hosp., Brisbane, Australia, [9]Royal Melbourne Hosp., Melbourne, Australia, [10]Royal Adelaide Hosp., Adelaide, Australia, [11]Bio21 Molecular Sci. and Biotechnology Inst., Melbourne, Australia, [12]Perth Children's Hosp., Perth, Australia, [13]Yale Univ., New Haven, CT, [14]Royal Perth Hosp., Perth, Australia, [15]Royal Children's Hosp., Melbourne, Australia, [16]Monash Biomedicine Discovery Inst., Monash Univ., Melbourne, Australia, [17]Westmead Hosp., Westmead, Australia, [18]Tasmanian Clinical Genetics Service, Hobart, Australia

Abstract:

Introduction: Genomic sequencing coupled with other "omic" technologies have transformed the diagnosis of mitochondrial respiratory chain disorders (RCD). The AGHA Mitochondrial Disease Flagship, a nationally funded research project, brought together clinicians, diagnostic and research scientists to focus on examining the clinical utility of genomic sequencing for the effective diagnosis of RCD. **Methods:** Using modified Nijmegen criteria, 140 paediatric- or adult-onset disease individuals were recruited and randomised to have either genome sequencing (GS) or exome sequencing (ES) and subsequent mtDNA genomic sequencing (mtDNAS) as needed. Where appropriate, transcriptomic, proteomic and/or cell-based functional studies were undertaken to functionally validate variants. **Results:** Overall, a genetic diagnosis was obtained in 77 (55%) of individuals in 77 different genes. The aetiology was nuclear in origin in 77% of cases, regardless of age of onset. The diagnostic rate was 71% in paediatric onset cases and 31% in adult-onset cases. The diagnostic yield from GS was not significantly different to ES/mtDNAS. Twenty-two individuals (29%) had diagnoses in genes not regarded as primary mitochondrial disease genes (so-called phenocopies), which would have been missed if only a

narrow RCD gene list had been used. We undertook a health economic analysis in paediatric-onset cases and showed that genomic sequencing was less costly than the "traditional" diagnostic approach (saving $2,000-$9,000AUD per diagnosis), and had an 11 - 14% higher diagnostic yield. Based on the estimated incidence of childhood-onset RCD, implementation of genomic sequencing early in the diagnostic odyssey could translate to an annual cost saving of up to $700,000AUD. **Discussion:** We have demonstrated that a national networked approach applying a consistent approach to selection of individuals for testing, with subsequent implementation of "multi-omic" technologies is a powerful method for efficiently delivering genetic diagnoses for this complex group of disorders, in many cases preventing the need for invasive testing. Importantly, the results of this study informed a recently successful application to obtain federally subsidised genome sequencing for individuals suspected of having a RCD.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2075T: The utility of exome sequencing in the seemingly obvious neurological diagnoses

Authors:

A. Meszarosova[1], J. Paulasova Schwabova[1], J. Haberlova[1], L. Osickova[2], J. Lastuvkova[3], V. Hurkova[4], L. Snajdrova[5], A. Musilova[1], P. Lassuthova[1], D. Safka Brozkova[1]; [1]Charles Univ. and Univ. Hosp. Motol, Prague, Czech Republic, [2]Faculty Hosp. Olomouc, Olomouc, Czech Republic, [3]Masaryk Hosp., Regional Hlth.Corp., Usti nad Labem, Usti nad Labem, Czech Republic, [4]PREDIKO, Zlin, Czech Republic, [5]Faculty Hosp. Ostrava, Ostrava, Czech Republic

Abstract:

Introduction: Spasticity and gait disturbance could be part of several different neurological conditions with similar initial clinical manifestation. Patients can be led for a long time (even many years) under a different diagnosis that mimics the correct one. The knowledge of proper genetic diagnosis is extremely necessary, because different disorders need different management and some of them are treatable. Evaluation of exome sequencing data is therefore a crucial tool for diagnosing these clinically overlapping diseases. Patients and methods: We examined more than 300 patients with gait disorders and clinical suspicion of hereditary spastic paraparesis using exome sequencing (SureSelect All Exome kit, Agilent Technologies). Approximately 100 patients were with early onset of disease (under 20 years of age). Exome data were evaluated using HPO terms and OMIM database in those patients where the hereditary spastic paraplegia was not genetically confirmed. Results: Among patients with childhood-onset we diagnosed nine with ultrarare disease (Galloway-Mowat syndrome,

Wieacker-Wolff syndrome, Branched-chain keto acid dehydrogenase kinase deficiency, SESAME syndrome and others). It represents almost 10% of all examined patients with childhood-onset. Among the group of adult-onset patients we diagnosed one with Wilson disease, Niemann-Pick disease and eight patients (subsequently their whole families) were diagnosed with X-linked adrenoleukodystrophy. All these disorders are severe with poor prognosis, but some of them are treatable to some extent. Spasticity and gait disturbances were only the primary clinical manifestations of the disease. Conclusion: In seemingly clear neurological diagnoses, clinical symptoms may be only the first manifestations of a completely different disease. In the case of such diseases, exome sequencing is an irreplaceable diagnostic method. Moreover, it may concern not only the rare diseases, but also the more common ones - e.g. Wilson disease, X-linked adrenoleukodystrophy. Grant support: The Ministry of Health of the Czech Republic NU22-04-00097.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2076T: TMC-CZI: Integrative Genomic and Functional Modeling Approaches for Novel Gene Discovery in Children

Authors:

C. Gu[1], A. Sabo[1], C. Nwaiwu[2,3], V. Korchina[1], D. Muzny[1], E. Venner[1], B. Yuan[1], C. Gijavanekar[1,4], Z. Liu[2,5], M. Wangler[4,5], S. Yamamoto[2,5], L. Satter[2,3], S. Sisley[2,3], C. Miyake[2,3], B. Vuocolo[4,5], R. Gibbs[1], H. Bellen[4,5], K. King[2,3]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [5]Jan and Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX

Abstract:

Establishing a molecular genetic diagnosis provides many benefits for patients and their families. While genome and exome sequencing (GS, ES) are increasingly becoming standard care with rising diagnostic yields, many individuals still lack a molecular diagnosis due to undiscovered novel Mendelian genes.The TMC-CZI study aims to discover novel genes responsible for severe Mendelian genetic disorders through integrative clinical genome sequence analysis and functional modeling. Participants, parents and available family members are recruited from a large local Children's Hospital, representing a diverse patient population with various phenotypes. Detailed phenotypic information, along with DNA and RNA samples, are collected for each individual. Clinical GS is performed, and a clinical GS report is issued for each participant. Data from individuals with negative clinical reports undergo further analysis

using research trio variant prioritization pipelines based on minor allele frequency, inheritance patterns, functional predictions, and information from various genomic databases. Additionally, variant prioritization includes several machine learning algorithms including AI-MARRVEL and Exomiser. When available existing GS or ES data are re-analyzed, and RNASeq analysis is conducted for a subset of participants. Top candidate variants and genes undergo cell and functional modeling in *Drosophila* and mouse animal models employing various methods.To date, 59 participants and their families have been referred to participate in the study. A majority (59%) of the participants come from groups underrepresented in biomedical research, including 43% Hispanic individuals, reflecting Houston's diverse population. The predominant phenotypes include cardiovascular (53%), gastroenterological (14%) and immunological (9%) disorders. Of the 59 participants referred to our study, 18 have been recruited with sequencing and analysis in progress. Approximately two-thirds of the recruited participants had prior clinical exome or genome testing.Results from ongoing analysis will be presented. The TMC-CZI study leverages a comprehensive approach combining clinical GS, advanced data analysis, and functional modeling in flies to identify novel genes associated with severe Mendelian disorders.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2077T: Understanding rare variant contribution to cleft palate using multiplex families

Authors:

K. Robinson[1], W. Adeyemo[2], T. Beaty[3], A. Butali[4], C. Buxo[5], L. Gowans[6], J. Hecht[7], L. Moreno[8], J. Murray[9], G. Shaw[10], S. Weinberg[11], H. Brand[12], M. Marazita[13], D. Cutler[14], M. Epstein[15], E. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Univ. of Lagos, Lagos, Nigeria, [3]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [4]Univ. of Iowa, Iowa City, IA, [5]Univ. of Puerto Rico, San Juan, PR, [6]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [7]Univ Texas McGovern Med Sch, Houston, TX, [8]Univ Iowa, Iowa City, IA, [9]U of Iowa, Iowa City, IA, [10]Stanford Univ., Stanford, CA, [11]Univ of Pittsburgh, Pittsburgh, PA, [12]MGH, Boston, MA, [13]Ctr. for Craniofacial and Dental Genetics, Unive, Pittsburgh, PA, [14]Emory Univ. Sch. of Med., Atlanta, GA, [15]EMORY Univ., Atlanta, GA

Abstract:

Cleft palate (CP) is a common craniofacial congenital anomaly, occurring in ~1 in 1700 live births. Family and twin studies show CP is highly heritable, therefore we hypothesized rare variants are a major driver of CP risk. To test this hypothesis, we analyzed 38 multiplex families from diverse populations with CP (i.e., those with two or more family members affected), as

they presumably share identifiable genetic risks. Families consisted primarily of nuclear families, including 29 trios with one affected parent and proband, 1 quad with two affected siblings, 3 affected parent-child dyads, and 5 multigenerational families of varying sizes. Variants with protein-altering consequences and minor allele frequencies of <1% in any population were included. We next filtered based on variant presence in affected or unaffected individuals. CP often exhibits an incompletely penetrant autosomal dominant inheritance pattern; thus, we selected variants which were heterozygous in all affected individuals and obligate carriers and absent in married-in unaffected individuals. We identified 5,601 total variants meeting all criteria (an average 183 variants per family). Prioritized variants fell into two categories: 1) monoallelic variants in known orofacial cleft (OFC) associated genes and 2) novel variants (absent from gnomAD v4.0). There were 57 variants in OFC genes found in 26 families, of which 6 were novel. These included missense variants in *IRF6* (pathogenic in ClinVar), *GRHL3*, *COL11A1*, and *PRICKLE1*, and frameshift variants in *FOXC2* (likely pathogenic in ClinVar) and *SMC3*, all of which are plausibly causal. There were 265 novel variants, including 24 predicted loss-of-function. Several of these have potential roles in CP. For example, one family harbors a frameshift deletion in *TMEM88B*—in zebrafish, tmem88 knockout disrupts pharyngeal pouch patterning. This may lead to CP in a similar mechanism to 22q11.2 deletion syndrome (often featuring CP), in which loss of *TBX1* disrupts pharyngeal pouch patterning. We also found a start-loss variant in *AMBRA1*, which is a regulator of autophagy and the cell cycle, and results in facial clefts when knocked out in mice. Although functional validation is needed, we identified several novel variants of interest for CP, highlighting the benefit of utilizing multiplex families. Ongoing work will further characterize all identified rare, protein-altering variants, including aggregation of families to find identical variants and/or genes harboring multiple predicted deleterious variants. Ultimately this work builds upon current knowledge of CP genetics and adds further understanding of rare variant contribution to disease.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2078T: Universal Exome Sequencing in Critically Ill Adults:A Diagnostic Yield of 25% and Race-Based Disparities in Access to Genetic Testing ⭐

Authors:

T. Drivas[1], J. Gold[2], C. Morse[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Northwell Hlth., Great Neck, NY

Abstract:

Numerous studies have underscored the diagnostic and therapeutic potential of exome or genome sequencing in critically ill pediatric populations. However, an equivalent investigation in critically ill adults remains conspicuously absent. Prevailing wisdom suggests that genetic testing yields in adults are inferior to those in pediatric patients and that genetic diagnosis would not impact emergent clinical care. Yet, this assumption lacks empirical support. Here, we retrospectively analyzed whole exome sequencing (WES) data available through the PennMedicine Biobank from all 365 young adult patients, aged 18-40 years, ever admitted to an ICU at the University of Pennsylvania Health System. For each participant, two Medical Genetics and Internal Medicine-trained clinicians reviewed WES reports and patient charts for variant classification, result interpretation, and identification of genetic diagnoses related to each patient's health and ICU admission.
Of the 365 individuals in our study, 90 (24.7%) were found to have clearly diagnostic results on WES; an additional 40 (11.0%) had a suspicious VUS identified; and an additional 16 (4.4%) had a medically actionable incidental finding. The diagnostic rate of exome sequencing did not decrease with increasing patient age. Affected genes were primarily involved in cardiac function (18.8%), vascular disease (16.7%), cancer (16.7%), and pulmonary disease (11.5%). Only half of all diagnostic findings were known and documented in the patient chart at the time of ICU admission. There were significant disparities by EHR-reported race, with genetic diagnoses known/documented for 63.5% of White patients but only for 28.6% of Black or Hispanic patients. There was a trend towards patients with undocumented genetic diagnoses having a 66% increased mortality rate, making these race-based disparities in genetic diagnosis even more concerning. A new definitive diagnosis was discovered in 11.2% of patients, with 76.6% of these conferring specific care-altering medical management recommendations.
Our study suggests that the diagnostic utility of exome sequencing in critically ill young adults is similar to that observed in neonatal and pediatric populations and does not decrease with increasing patient age. The high diagnostic rate and striking race-based disparities we find argue for broad, universal approaches to genetic testing in critically ill adults to enhance medical care for all individuals and rectify systemic racial disparities.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2079T: Variants in *FGF20* underlie a novel breast malformation

Authors:

K. Buckingham[1], C. T. Marvin[1], K. M. Shively[1], A. B. Stergachis[2], D. E. Miller[1,3], C. M. Carvalho[4], M. Lun[4], J. Mathieu[5], J. C. Carey[6], F. K. Mastrorosa[7], C-L. Wei[7], J. X. Chong[1,3], M. J. Bamshad[1,3,5,7], University of Washington Center for Rare Disease Research; [1]Univ. of Washington, Dept of

Pediatrics, Seattle, WA, [2]Univ. of Washington, Dept of Med. Genetics, Seattle, WA, [3]Brotman Baty Inst. for Precision Med., Seattle, WA, [4]Pacific Northwest Res. Inst., Seattle, WA, [5]Univ. of Washington, Inst. for Stem Cell and Regenerative Med., Seattle, WA, [6]Univ Utah Hlth Sci Ctr, Salt Lake City, UT, [7]Univ. of Washington, Dept of Genome Sci., Seattle, WA

Abstract:

Of the ~6,000 known mammals, all have mammary glands and most, but not all, have nipples. Mammary glands vary in number from 2-27 among species and show very limited intraspecies variation. Polythelia, or accessory nipples, is estimated to occur in 1-5% of humans. Typically, only one extra nipple is present, therefore the predominant presentation is unilateral, with the incidence reported to be the same in males and females. We report the discovery of the genetic basis of a very rare, isolated malformation of the breast, areola and nipple commonly referred to as intra-areolar polythelia (IAP). The condition was transmitted in an autosomal dominant pattern in an extended multiplex pedigree in which at least eighteen persons in six generations were affected. The canonical clinical finding in IAP appears to be disorganization of the nipple with accessory nipples arising within the areola, resulting in an apparent “split nipple.” Additional phenotypic features include polythelia and breast hypoplasia. Initial analysis of short-read whole genome sequencing (srWGS) of three affected individuals identified no compelling candidate single nucleotide or structural variants. Simultaneously, we genotyped 16 affected individuals using the Omni 2.5 chip and identified a linkage peak with a maximum LOD score of 4.8 on chromosome 8 using an autosomal dominant model. We searched the srWGS data for variants within the 7.6 Mb linkage interval and detected two rare overlapping, and nested duplication events on a single allele, both of which included fibroblast growth factor 20 (*FGF20*). This results in a total of four copies of *FGF20* (i.e. one WT allele and one allele with a triplication) and surrounding non-coding sequence. In mice, *Fgf20* is expressed in embryonic mammary glands and regulated by ectodysplasin A (Eda). Loss-of-function of *Fgf20* in a mouse overexpressing Eda impairs mammary bud growth. We hypothesize that the presence of additional copies of *FGF20* leads to over-expression and subsequent dysregulation of mammary bud development. FiberSeq was performed on iPSCs derived from blood on two individuals, which demonstrated that chromatin accessibility at the *FGF20* promoter and its promoter-proximal regulatory elements was preserved on all three copies of the *FGF20* triplication on the mutated haplotype, with no appreciable changes in their relative epigenetic activity. This functional data indicates that the epigenetic architecture of each copy is suitable to result in *FGF20* transcription. Sequencing of two additional IAP cases is underway and expression studies in mammary-like organoids have been initiated.

Session Title: Mendelian Phenotypes Poster Thursday Session

Board 2080T: Whole exome sequencing trio identifies a novel copy number variation linked to Canavan disease

Authors:

S. Siddique[1,2,3], A. M. Butt[4,5]; [1]Sch. of Med. and Dentistry, Univ. of Central Lancashire, Preston, United Kingdom, [2]American Univ. of the Caribbean (AUC), Sch. of Med. - UK Track, Univ. of Central Lancashire, Preston, United Kingdom, [3]Molecular Pathology Div., MEDFER Diagnostics, Islamabad, Pakistan, [4]Dept. of BioSci.s, COMSATS Univ. Islamabad (CUI), Islamabad, Pakistan, [5]Comprehensive Cancer Ctr., Faculty of Life Sci. & Med., King's Coll. London, London, United Kingdom

Abstract:

Background: Canavan disease (CD) is an autosomal recessive leukodystrophy characterized by spongy degeneration of the brain's white matter. Mutations in the aspartoacylase (*ASPA*) gene leads to impaired enzymatic activity and the accumulation of N-acetylaspartic acid (NAA) in the brain. While leukodystrophies are previously noted in the Pakistani population, only a few reach a definitive diagnosis through genetic testing. **Aim:** This study aims to genetically characterize a pediatric case suspected of CD based on clinical features and magnetic resonance imaging (MRI) findings. **Case Presentation:** A 2-year-old male, born to a first-degree consanguineous couple, was referred to our genetics clinic for a biochemical and genetic workup for leukodystrophies and inherited errors of metabolism. At presentation, notable clinical features included macrocephaly, inability to sit independently for extended periods, poor head control, generalized muscle weakness, and spasticity. Birth history was uneventful with no gross anomaly and a birth weight of 3.46 kg, length of 54 cm, and OFC of 37 cm. Concerns arose at three months of age due to poor head control, frequent crying, poor visual tracking, delayed milestones, and hypotonia. The proband experienced the first seizure at one year of age. Major MRI findings were T2/FLAIR abnormalities in the cerebral white matter involving periventricular regions and subcortical U-fibers, with sparing of the bilateral caudate, putamina, and internal capsules. **Methods:** NAA levels in urine were measured using gas chromatography-mass spectrometry. Whole blood samples from the proband and parents were collected for genomic DNA extraction and WES-Trio. Variant interpretation was performed according to ACMG/AMP guidelines. Segregation and variant(s) validation was performed via sanger sequencing and deletion/duplication analysis. **Results:** NAA levels were elevated (879 mmol/mol creatinine; control: 6 mmol/mol), consistent with CD. WES-Trio identified an 18 kb deletion on chromosome 7 (coordinates 3379454-3397758), encompassing exons 1-5 of the *ASPA*. This CNV is absent in the gnomAD and ClinVar databases. Segregation analysis confirmed both parents as heterozygous carriers. **Conclusion:** Partial or complete deletions of the *ASPA* are rare. The identification of a novel homozygous deletion on chromosome 7 expands the mutational

spectrum associated with CD. With ongoing clinical trials showing promising improvements, this finding has significant implications for understanding the disease's genetic diversity, genetic counseling for at-risk pregnancies, and development of targeted gene therapies.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3041T: A 13q14.11q21.33 deletion involving *RB1* and *ATP7B* implicating hereditary retinoblastoma and Wilson disease: A 26-year follow-up

Authors:

J. Chenbhanich[1,2], G. Weber[1,3]; [1]Ctr. for Human Genetics, Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [2]Dept. of Genetics and Genomic Sci., Case Western Reserve Univ., Cleveland, OH, [3]Dept. of Pathology, Case Western Reserve Univ., Cleveland, OH

Abstract:

Chromosome 13q14 deletion syndrome (OMIM #613884) is characterized by neurodevelopmental disorders, craniofacial dysmorphic features, and an increased risk of retinoblastoma due to loss of *RB1* at 13q14.2. Wilson disease (WD) is an autosomal recessive disorder caused by pathogenic variants in *ATP7B* at 13q14.3. The co-occurrence of these conditions was first reported in our patient by our institution, and we provide a 26-year follow-up. The patient presented with left eye pain at age 3 and was diagnosed with unilateral, unifocal retinoblastoma, treated by enucleation without chemoradiation. He had global developmental delay and learning disabilities. At age 13 years, he presented with a 2-year history of progressive dystonia and parkinsonism. Investigations revealed cytopenias, low serum copper and ceruloplasmin, high urinary copper, a Kayser-Fleischer ring, and basal ganglia hyperintensity on T2-weighted MRI. WD was diagnosed on clinical presentation; *ATP7B* sequencing was not performed at the time. He has been receiving the copper chelating medication trientine since his WD diagnosis. Chromosome analysis showed a deletion in chromosome 13: 46,XY,del(13)(q14.1q13q22.2). FISH confirmed *RB1* gene deletion. Parental chromosome studies were normal. Microsatellite analysis revealed the *de novo* deletion was on his paternal chromosome. It was inferred the WD variant was present on his maternal chromosome. At age 39, the patient returned to the genetics clinic with liver cirrhosis but no movement disorders or secondary cancer. A physical examination showed craniofacial and musculoskeletal dysmorphic features. Chromosomal microarray performed to assign precise breakpoints identified a 28.4

Mb deletion: arr[GRCh37]13q14.11q21.33(44927384_73283184)x1. Sequencing confirmed *ATP7B* gene deletion and an *in-trans* pathogenic variant c.3796G>A (p.Gly1266Arg), which allows for familial testing. All deleted genes were reviewed; he exhibited no signs of other monogenic disorders caused by haploinsufficiency of those genes. One deleted gene, *MED4*, has been shown to lower risk of retinoblastoma and secondary cancer compared to those with variants within the *RB1* gene or deletions without *MED4*. Similar to our patient, they tend to have unilateral and unifocal disease, and some did not have retinoblastoma.
Although 13q14 deletion syndrome has been reported previously, our patient is the only published report with co-occurrence of WD. We illustrate how updated genetic knowledge can inform care recommendations and prognostication, especially in individuals with limited ability to self-report cancer-related symptoms to caregivers.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3042T: A retrospective study of Sanger sequencing reduction in clinical genetic testing based on whole genome sequencing

Authors:

P. Tong[1], K. Weymouth[1], K. Dumas[1], L. Liao[1], C. Ho[1], N. Hammond[1], Y. Yang[1,2]; [1]Clinical Genomics Lab., Stanford Hlth.Care, Palo Alto, CA, [2]Dept. of Pathology, Stanford Univ., Stanford, CA

Abstract:

Complying MOL.35850 in CAP Molecular Pathology Checklist, laboratories must have a written policy determining if and when confirmatory testing on reported variants identified by Next Generation Sequencing (NGS) based clinical genetic testing should be performed. Sanger sequencing has been considered the gold standard for identifying DNA variation by sequencing PCR amplicon and been used widely in clinical testing lab as an orthogonal confirmation for NGS testing. Defining necessity of Sanger confirmatory sequencing is critical for the accuracy, cost-efficacy and curation time of NGS testing. To this end, we developed and implemented a Sanger sequencing reduction procedure identifying variants with high sequencing quality that Sanger confirmatory sequencing can be waived. Whole genome sequencing of six genome in a bottle (GIAB) samples (HG002-HG007) were performed following validated clinical workflow. Thresholds of seven quality metrics of variation calling were combined and assessed for the best performance of identifying the variants with high quality, specifically maximize Sanger reduction efficiency subject to the condition of maintaining a precision TP/(TP+FP) no less than 99.99% in all six GIAB samples. The combination of thresholds with best performance was subsequently validated with one GIAB sample (HG001) and achieved 76.18% Sanger reduction rate while the

precision of variants passed thresholds was increased to 99.98% from 99.85%. The procedure was also validated with 97 previously reportable variants, 7 out of 7 non-confirmed variants confirmed were identified as Sanger needed by the procedure, leading to a 100% sensitivity of identifying false positive calls of NGS. 70 out of the 97 variants were identified as Sanger not needed, which leads to a 72.16% Sanger reduction rate. After implementation of Sanger reduction, 77% reported variants of cardio panel identified with NGS were waived Sanger confirmatory sequencing which significantly reduced labor and turnaround time in wet lab and interpretation process. In summary, Sanger reduction effectively improved workflow, cost and turnaround time of NGS based genetic testing while ensuring the precision.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3043T: A *ZFHX3* nonsense variant in a girl with developmental delay and arrhythmia

Authors:

K. Nagai, Y. Kawai, R. Tachioka, T. Nishikawa, Y. Enomoto, T. Naruto, Y. Kuroda, K. Kurosawa; Kanagawa Children's Med. Ctr., Yokohama, Japan

Abstract:

Zinc finger homeobox 3 (ZFHX3) is a transcription factor encoded by *ZFHX3*. In 2024, Pérez Baca, *et al*. reported haploinsufficiency of *ZFHX3* causes intellectual disability. ZFHX3 is important for neuronal differentiation. Genome-wide analysis revealed that *ZFHX3* polymorphism also involved in atrial fibrillation. We report a *ZFHX3* nonsense variant (p.Gln882*) in a girl with developmental delay and arrythmia. The proposita was a 3-year-old girl. She was born to non-consanguineous parents at 40 weeks' gestation. Her birth weight was 2905 g (-0.32 SD), length was 48 cm (-0.82 SD), and head circumference was 32.6 cm (-0.63 SD). Her development was delayed, with head control at 4 months, sitting at 10 months, crawling at 11 months, and walking at 2 years and 10 months of age. She had feeding difficulty. She was diagnosed with autism spectrum disorder. Her developmental quotient was 34 at 18 months of age. She could speak some words at 3 years. She was diagnosed with patent ductus arteriosus at 5 months old and atrial tachycardia at 11 months old. Proband's exome sequencing identified a heterozygous variant in *ZFHX3* (NM_001164766.2:c.2464C>T (p.Gln882*)). The variant was confirmed to be de novo by trio Sanger sequencing. This patient demonstrated intellectual disability, autism spectrum disorder, and feeding difficulty. She had haploinsufficiency of *ZFHX3*. ZFHX3 (p.Gln882*) had not been reported as a causal variant of developmental delay. Nonsense variants reported by Pérez Baca, *et al*. are scattered throughout

the gene excluding its last exon. As with reported nonsense variants, messenger RNA transcribed from *ZFHX3* (c.2464C>T (p.Gln882*)) is likely to be targeted by nonsense-mediated decay. The reported patients showed mild to severe intellectual disability (49%), global developmental delay (63%), behavioral problems (40%), and feeding difficulty (63%); almost all probands had at least one of the symptoms. Her symptoms are consistent with those of previous patients although these reported patients did not suffer from arrhythmia. ZFHX3 knockdown in a cardiac muscle cell line and its heterozygous loss in mice revealed their relationship with arrhythmia. Her arrhythmia might be associated with the impaired ZFHX3 function. In conclusion, we reported a patient with truncating variant of *ZFHX3* demonstrating developmental delay and arrhythmia as an atypical phenotype of ZFHX3-related neurodevelopmental disorder.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3044T: Autosomal recessive variant reporting in clinical CMA, a three-year retrospective analysis

Authors:

D. Reich[1], S. ZoBell[1], Z. Lewis[1], **E. Andersen**[1,2], K. Rudd[1]; [1]ARUP Lab., Salt Lake City, UT, [2]Univ. of Utah, Salt Lake City, UT

Abstract:

Chromosomal microarray analysis (CMA) is the first-tier test recommended for children with developmental delay (DD), intellectual disability (ID), congenital anomalies, and/or autism. In the prenatal setting, CMA and chromosome analysis are the gold standard for diagnostic testing. Though constitutional CMA testing has historically focused on dominant disorders, focal copy number variants (CNVs) involving autosomal recessive disease (AR) loci may also be detected. CNV interpretation guidelines give labs discretion to develop individual reporting practices for recessive carrier status and for variants that may be related to the indication for testing. Clinical laboratories that perform CMA have adopted different policies for reporting carrier status of recessive variants.

At ARUP Laboratories, variants conferring AR disease risk are generally reported for constitutional CMA testing of all specimen types, including prenatal, products of conception (POC), and postnatal. While CNVs involving AR risk genes may relate to the patient's presentation, their true contribution to clinical diagnosis is unclear. Additionally, reporting carrier status may add costs through cascade testing for other variants and cause unnecessary anxiety for patients and families. To evaluate the clinical utility of reporting variants conferring

AR risk, we analyzed data from constitutional CMA performed between 2021-2024. We focused on deletions only reported for AR disease risk, not larger multigenic deletions potentially related to dominant disorders. Focal AR risk deletions were reported in 5.2%, 4.3%, and 5.0% of prenatal, POC, and postnatal CMA tests, respectively. The recurrent NPHP1 deletion was the most reported deletion, comprising 9.6% of all AR risk reports. Next most common were deletions of OTOA (5.5%), PRKN (3.7%), GALC (3.3%), AHI1 (2.4%), and CTNS (2.4%). Using the clinical indications provided at the time of testing, we searched for specific features related to these genes that might suggest the recessive disorder. One proband had hearing loss which may be related to his OTOA deletion, but testing for a second variant was not pursued. No other deletion carriers had clinical features specific to these most common genes, though non-specific findings including hypotonia, DD, and ID could be related to some of these AR disorders. Our review suggests that reporting variants conferring only AR risk provides limited diagnostic benefit for patients. Laboratories may consider discontinuing this practice in clinical CMA testing in the absence of specific phenotypic overlap.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3045T: Building an interpretable Bayesian quantitative variant classification system

Authors:

Y. Kobayashi, B. Eaves, L. Fresard, J. Kim, S. Klemm, T. Manders, S. Padigepati, A. Wahl, K. Nykamp; Invitae, San Francisco, CA

Abstract:

Background: The primary purpose of clinical genetic testing is to identify genetic variants and determine if they are pathogenic (P) or benign (B). In many cases, however, limited data results in lingering uncertainty, and pathogenicity cannot be asserted definitively. Variant classification (VC) is the process by which all available functional and clinical data are evaluated and summarized to communicate the magnitude of remaining uncertainty. The clinical genetics community has long sought a quantitative VC system that accurately and precisely calculates probabilities of pathogenicity (PoP), but this goal has remained aspirational.
Recent advances in AI technology, paired with rapidly expanding genomic data, has led to a growing number of algorithms for quantification of clinical or functional data. However, an overarching model that accurately combines them to calculate an overall PoP has remained elusive. Past attempts at such combined models have had limited adoption due to lack of interpretability (“black box”) and incompleteness of the evidence set incorporated (e.g.

omission of clinical data).
Methods & Results: As a proof of concept, we developed a Bayesian probabilistic VC system (BPVC) using a probabilistic graphical modeling approach for several genes expected to be rich in clinical and functional data. This BPVC incorporates all major categories of data used in ACMG guidelines, including: patient clinical, population frequency, *in vitro* and *in silico* functional, splicing, and nonsense mediated decay prediction data. To leverage domain expertise, it was designed with clinical geneticist input to map known causal relationships between categories. The BPVC was trained and validated using 9,927 classified variants from ClinVar and internal data.
Importantly, the BPVC achieved high binary classification accuracy for discriminating known P and B variants (AUROC > 0.98). We also found the BPVC learned dependent and independent relationships between evidence categories in a way that matched the expectations of clinical geneticists. Finally, the BPVC exhibited high accuracy of PoPs throughout the entire range of scores from 0 to 1 (Brier score = 0.030), suggesting that PoPs are precise and can be used for setting classification thresholds. With thresholds at ≤5% and ≥99% PoP, more than half of VUS classified with the same data using ACMG guidelines could be reclassified to B and P, respectively.
Conclusion: Our approach demonstrates how a quantitative VC system with a high degree of interpretability is attainable through a combination of clinical genetics expertise, a Bayesian probabilistic approach, and a large quantity of amassed variant data.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3046T: Characterization of a complex genomic rearrangement by optical genome mapping reveals BIRC6 as a candidate gene for intellectual disability

Authors:

J. Mazzeu[1,2], V. S. Souza[1,2], G. C. R. da Cunha[1], M. S. Cordoba[3], B. R. Versiani[3]; [1]Univ.e de Brasilia, Brasilia, Brazil, [2]Inst. Natl. de Doenças Raras, InRaras – Brazil, Brasília, Brazil, [3]Hosp. Universitário de Brasília, Brasilia, Brazil

Abstract:

Complex genomic rearrangements (CGR) are characterized by the presence of a large-scale genomic alteration involving more than one apparently simple structural variation with more than two breakpoints. Evidence from genomic disorders supports the hypothesis that pathogenic CGRs play a significant role in genetic diseases caused by structural variations,

including malformation syndromes, intellectual disability, and neurodevelopmental disorders. Here we report the case of a 14-year-old boy with intellectual disability, behavioral problems, mild facial dysmorphisms and febrile seizures. Standard G-banded karyotype revealed an inversion in chromosome 2p [seq[GRCh37]inv(2)(p13p25.1)mat], inherited from his phenotypically normal mother. To characterize the chromosomal rearrangement in more detail Optical Genome Mapping (OGM) was performed. The joint OGM analysis of the proband and his mother revealed a complex genomic rearrangement. OGM confirmed and refined the inversion in the proband ogm[GRCh37] inv(2;2)(p24.1;p15) and his mother. Moreover they both carry a 850 Kb pericentromeric deletion [ogm[GRCh37]2p11.2(87253768_88059339)x1] but only the proband has a 655 Kb duplication [ogm[GRCh37]2p22.3(32624557_33280372)x3] within the inverted segment. This duplication includes TTC27 gene and leads to the interruption LTBP1 and BIRC6. Heterozygous deletion LTBP1 was reported in a familial case of thoracic aortic aneurysm, and homozygous deletions cause cutis laxa. Consequently, interruption of the LTBP1 gene does not seem to be relevant to the patient's clinical picture. TTC27 has not been previously associated with any phenotype. However, BIRC6 gene is a member of the family of apoptosis inhibitors and acts as a mediator in cytokinesis and DNA restoration. The deregulation of this gene has already been associated with the process of tumorigenesis and is implicated in several types of cancer. CNVs involving the BIRC6 gene have been previously reported in two patients with ASD and schizophrenia. Our results suggest that disruption of the BIRC6 gene is the cause of the patient’s phenotype and therefore a strong candidate gene for intellectual disability. Furthermore, this case illustrates that inherited CGR should be investigated in detail even when inherited from a nonaffected parent.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3047T: Complex chromosomal rearrangements involving 66 breakpoint junctions on 7 chromosomes in a patient with atypical symptoms

Authors:

T. Naruto, Y. Enomoto, Y. Kawai, K. Nagai, Y. Kuroda, K. Kurosawa; Kanagawa Children's Med. Ctr., Yokohama, Japan

Abstract:

Complex chromosome rearrangements (CCRs) are structural variants (SVs) consisting of adjacent breakpoint junctions (BPJs) in one or more chromosomes. Although initially detected in cancer cells, CCRs have also been reported in the germline, and such CCRs are known to cause a variety

of disorders, including intellectual disability and multiple congenital anomalies. We identified complex rearrangements in a patient with developmental delay. The patient was 2-year-old girl demonstrating developmental delay and dysmorphic features. She controlled her head at 8 months, started sitting at 16 months, and walked with support at 2 years old. She never spoke any meaningful words at 2 years old. G-band karyotype was three-way translocations involving chromosome 2, 8, and 13. Spectral karyotyping (SKY) revealed 46, XX, der(2)t(2;13)(p11.2;q32), ins(8;13)(q24.1;q22q32), ins(13;2)(q12.3;p24p11.2)(q22;p24p25). Chromosomal microarray detected a 400 kb deletion at 15q11.2 (BP1-2), not constituting breakpoints of the complex rearrangement. We performed short-read whole genome sequencing to delineate the rearrangements. Sample was sequenced by NovaSeq, and reads were aligned to both GRCh38. The data was analyzed by Manta for SV calling. After filtering out by in-house SV data, 66 unique BPJs were determined by checking manually in IGV. Chromosome 1, 4, 11, and 21 were involved in the rearrangements, appearing to be normal in karyotyping and SKY. Three deletions were found larger than 1kb (5, 5 and 89 kb respectively). The sum of total genomic imbalance from all junction fragments was less than 100 kb. OMIM genes were not included in these deleted regions. Five OMIM genes were disrupted by breakpoints. MSH6 gene was one of the disrupted genes and might have caused cancer predisposition of Lynch syndrome. Short-read whole genome sequencing delineated the CCRs involving seven chromosomes. CCRs involves many chromosomes appearing to be normal in conventional karyotyping. Therefore, whole genome sequencing is an efficient method to overview the rearrangement in genome comprehensively.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3048T: Complex rearrangement including 7q21.3 region detected in the patient with split hand/foot malformation.

Authors:

Y. Enomoto[1], K. Yanagi[1], M. Yamamori[1], K. Naritomi[2], T. Kaname[3]; [1]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [2]Univ Ryukyus Sch Med, Okinawa, Japan, [3]Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan

Abstract:

It has been reported that 7q21 loci was involved in Split hand/foot malformation (SHFM). We previously reported a patient with SHFM who had de novo translocations of chromosomes 3 and 7, including 7q22. He was born at 40 weeks gestation with a birth weight of 3,105 g (0.03 SD) and length of 46.8 cm (-1.43 SD), and head circumference 32.9 cm (-0.38 SD). He showed ectrodactyly (rt 3rd finger, bilat 2/3/4 toes absence), uvula bifida, diastasis recti, high-arched

palate, and normal intelligence. The G-banding resulted his karyotype as 46, XY, inv ins(3;7)(q2l;q34q22). As the detailed genomic structure, including chromosomal breakpoints, responsible for the patient's SHFM was unknown, we performed detailed analyses such as FISH, spectral karyotyping (SKY) and Nanopore long-read sequencing. The SKY analysis identified complex structural changes such as translocations to chromosome 8 in addition to translocations between chromosomes 3 and 7. Results showed (a) insertion of a chromosome 7 fragment into chromosome 3, (b) addition of a chromosome 8 fragment to chromosome 7 and (c) addition of a chromosome 7 fragment to chromosome 8. For further detailed structural analysis, Nanopore long-read sequencing was performed. The results revealed breakpoints at chromosomes 7 and 8, located at 7q21.3 and 8q23.3 respectively.The chromosome 7q21.3 region where the breakpoint was found contains three genes associated with SHFM, *DLX5*, *DLX6* and *DYNC1I1*. *DLX5* is known to be the causative gene for Split-hand/foot malformation 1 (OMIM: 183600). However, the breakpoint on chromosome 7 in this patient was not located at the *DLX5* gene. On the one hand, it has been reported that a region of *DYNC1I1* may act as an enhancer and regulate the expression of *DLX5* and *DLX6*. Topologically Associating Domain (TAD) boundaries including *DLX5*, *DLX6* and *DYNC1I1* have also been reported. Our results support the importance of this region in limb development and demonstrate the importance of using new techniques to reveal detailed structures in previous cases where G-band chromosome analysis was available.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3049T: Copy number variation detection without using external baseline

Authors:

Q. Zeng[1], W. Xin[1], N. Russell[1], C. Xu[2], K. Phillips[3], J. Chappell[4], M. Yan[5], S. Letovsky[1], A. Kenyon[1]; [1]Labcorp, Westborough, MA, [2]Labcorp, San Diego, CA, [3]Labcorp, Durham, NC, [4]Labcorp, Atlanta, GA, [5]Labcorp, Lakeway, TX

Abstract:

Copy number variation (CNV) plays important roles in human health and diseases. For example, the presence of an abnormal number of chromosomes in a cell (“aneuploidy”) is the most frequent cause of human mental retardation and pregnancy loss. Additionally, CNVs in sub-chromosomal regions (e.g. genes, exons, etc.) have been linked to cancer, cardiac defects, limb abnormalities, immunodeficiency, and many other genetic disorders. The ability to detect aneuploidies, gene level CNVs, and exon level CNVs is critical for clinical testing. CNV detection

by next-generation sequencing (NGS) technologies is widely used in clinical testing. Currently available NGS CNV analysis methods generally use preselected baseline samples with no known CNVs in the target regions as the reference, and then compare the coverage profiles between the testing samples and baseline samples to detect the CNVs. This approach faces several challenges. For example, baseline samples with no known CNVs in the target regions are very hard to find, especially for whole exome sequencing and whole genome sequencing projects. Efforts have been made to get around this by mixing dozens to hundreds of samples to create synthetic baseline samples, which by itself is costly and could be challenging for smaller labs. In addition, NGS sequencing baseline could shift with time or vary between batches ("flowcells"), leading to higher false positives and false negatives. If detection of CNVs on ChrX and ChrY is required, labs must maintain static baselines for each gender as well. We have addressed these issues by developing a new CNV detection method called MASCOT (for "Median Coverage of All Samples for Copy Number Outlier Testing"). MASCOT is based on the assumption that most samples in a flowcell have similar coverage profiles and no CNVs in most regions. This allows us to build a normal coverage/copy number profile for each target region, and then find CNVs via outlier detection by comparison of each segment in each sample with the corresponding normal coverage/copy number profile for that segment. MASCOT accounts for the samples' genders and significantly reduces the noise encountered with copy number analysis when pre-selected baseline samples (not from the same batch) are used as the control ("batch effect"). MASCOT is also able to detect CNVs at exon, gene and chromosome levels simultaneously. Based on analysis of clinical samples with confirmed CNV calls and a large number of simulated samples with known CNV truth, MASCOT has demonstrated better sensitivity and specificity compared with previously reported CNV detection methods and can detect CNVs down to a single exon level if desired.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3050T: Detection of a balanced translocation between chromosomes 11p and 14q by fluorescent in situ hybridization- A Case Report

Authors:

C. Kumar, M. McDonald; Westchester Med. Ctr., Valhalla, NY

Abstract:

Balanced reciprocal translocations occur in approximately 1 in 600 individuals. Most carriers of balanced reciprocal translocations have normal phenotypes. Molecular analyses have shown

that the apparently balanced reciprocal chromosome translocations in patients with abnormal phenotypes are accompanied by a chromosome imbalance at the translocation breakpoints or elsewhere in the genome. Carrier of balanced reciprocal translocations can result in partial aneuploidies in the offspring due to unbalanced meiotic segregation during gametogenesis. Here we present a new case of 13-year old female with microcephaly, developmental delay and aggression. The cytogenetic, chromosomal microarray and molecular test for Fragile X Syndrome were indicated. Chromosome analysis was performed on GTG banded metaphases prepared from cultured lymphocytes that revealed an additional unidentified material at the short arm of chromosome 11 and the karyotype as: 46,XX,add(11)(p15.5). Fluorescence in situ hybridization (FISH) test was performed on metaphase cells using four different probes (Abbott Molecular Inc.), which include Telvysion probes: 4p (Spectrum Green) and 4q (Spectrum Orange) for chromosome 4, 11p (Spectrum Green) for chromosome 11; and LSI IGH Dual Color Break Apart probe for chromosome 14 which hybridizes to band 14q32.3 and produces a yellow fusion signal. The microarray analysis performed on peripheral blood indicated a normal result and Fragile X syndrome result was negative in this patient. FISH analyses for 4p and 4q showed normal signals, and abnormal signal patterns were observed for 11p and IGH probes. FISH analysis for 11p showed two signals: one normal signal on one 11p and the other abnormal signal on one 14q; and for IGH, it showed two signals: one fusion signal on the normal chromosome 14 and the other fusion signal on 11p of the abnormal chromosome 11, confirming the breakpoint in chromosome 14 is proximal to the 14q32.3 band. FISH studies identified a cryptic reciprocal translocation between the short arm of chromosome 11 and the long arm of chromosome 14, with breakpoints at 11p15.5 and 14q32.2, respectively. To the best of our knowledge, this translocation t(11;14)(p15.5;q32.2) is the first report of this combination of chromosomal abnormalities. In conclusion, this case demonstrates the effective use of cytogenetics, microarray and FISH studies to characterize the unidentified balanced translocation for patient's diagnosis and clinical management. Parental chromosome analyses were recommended to determine whether this abnormality is inherited or de novo in origin. Clinical correlation and genetic counseling were recommended.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3051T: Evaluating consensus policy in rare disease diagnosis using automated HPO match

Authors:

Y. Jiang, J. Westbrook, K. Ouyang; Invitae Corp., San Francisco, CA

Abstract:

Introduction: Commercial laboratories currently use insurance or consensus policies to determine medical necessity for patients who ordered genetic testing for rare diseases. We developed a method using human phenotype ontology (HPO) match to automatically assess whether each patient meets consensus policy. Our goal is to evaluate whether individuals meeting consensus policies have a higher diagnosis rate compared to those not meeting consensus policies.
Methods: We selected seven frequently ordered rare disease panels representing different disease areas or sub areas with consensus policy available for evaluation. Panel-specific HPO terms were derived from each consensus policy. These terms are then compared to each patient, whose HPO terms were extracted from the indication for testing texts using an in-house application. Related HPO terms were systematically included allowing fuzzy matching (e.g., 'hearing impairment' and 'sensorineural hearing impact' would both be matches for 'hearing loss'). Percentage of HPO matches between consensus policy and indication for testing were calculated and compared between patients with or without diagnosis.
Results: HPO match rate is significantly higher ($p < 0.05$) among patients with diagnosis in four panels with consensus policies, including Invitae® Epilepsy, Comprehensive Neuromuscular, Inherited Retinal Disorders and Aortopathy panels. The largest percentage difference was observed in the Inherited Retinal Disorders panel, where 47.3% of patients with an HPO match had a molecular diagnosis and 38.4% patients without an HPO match had a molecular diagnosis. No significant differences were observed in the other three panels, however, percentages of diagnosis were slightly higher in patients with HPO matches than the ones without.
Conclusion: Our study shows that patients with HPO matches are more likely to get molecular diagnosis in some panels, further supporting the medical necessity of these genetic tests. Our study also demonstrated how HPO terms derived from consensus policies and patient indication texts can contribute to automated analysis, paving the way for artificial intelligence (AI) usage in clinical, data and billing aspects of genetic testing and optimizing the predictive power of AI in genomic medicine.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3052T: Expanding the diagnostic toolkit using long-read RNA-sequencing

Authors:

K. Montgomery[1,2], E. Gustavsson[1,3,4], K. Tuschl[1], E. Clement[5], P. Mills[1], H. Houlden[2], M. Ryten[1,3,4,6,7]; [1]UCL GOS Inst. of Child Hlth., London, United Kingdom, [2]UCL Queen Square Inst. of Neurology, London, United Kingdom, [3]NIHR Great Ormond Street Hosp. BioMed. Res. Ctr.,

London, United Kingdom, [4]UK Dementia Res. Inst. at The Univ. of Cambridge, Cambridge, United Kingdom, [5]North East Thames Regional Genetic Service, London, United Kingdom, [6]Dept. of Clinical NeuroSci.s, Cambridge, United Kingdom, [7]Aligning Sci. Across Parkinson's (ASAP) Collaborative Res. Network, Chevy Chase, MD

Abstract:

Molecular diagnosis of inherited rare disease plateaus at around 50% and with progress in the development of disease-modifying treatments for rare diseases diagnostic speed has become increasingly important. However, significant diagnostic challenges remain and include technical limitations of short-read sequencing, novel gene-disease relationships yet to be discovered and the high proportion of variants of uncertain significance (VUS). In this study we evaluated the use of targeted long-read sequencing as an additional tool for the rapid assessment of aberrant splicing in patient samples to raise diagnostic yield. We assessed this technology in the context of undiagnosed paediatric patients with suspected neurometabolic disorders, previously tested by WES or WGS, and following an additional multidisciplinary review that suggested a monogenic disease with <5 genes of particular interest (based on a combination of clinical phenotype, the presence of VUS or a single pathogenic variant in the context of a probable autosomal recessive disorder). Focusing on a set of 10 patients for whom there was high confidence in a small number of candidate genes, we used patient blood samples or fibroblasts to generate cDNA, PCR for genes of interest and generate full length reads using Oxford Nanopore Technologies single molecule sequencing. Using this approach we identified probable pathogenic splicing events in two patient samples. We used not only existing transcriptomic annotations, but also publicly available short- and long-read sequencing data from control individuals to determine the rarity of the events detected. Furthermore, by enabling the prediction of both ORFs and protein structures, the availability of full-length reads greatly increased our ability to identify events expected to disrupt gene function rapidly. Taken together, our findings suggest that targeted long-read RNA-sequencing in patients who remain undiagnosed following extensive testing could be an important tool particularly when a treatable genetic diagnosis is being considered.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3053T: Experience of germline genetic testing for inborn errors of immunity: using multigene panel testing compared to exome sequencing at a diagnostic laboratory

Authors:

D. Pineda-Alvarez[1], T. Williams[2], Y-L. Ting[2], S. Aradhya[2]; [1]Invitae, Wellington, FL, [2]Invitae, San Francisco, CA

Abstract:

Inborn errors of immunity (IEI) are challenging to diagnose because of their heterogeneous presentation. Next-generation sequencing (NGS) has proven a valuable tool to diagnose these conditions because it can interrogate many genes concurrently and has enabled a quick expansion of IEI-related genes. Currently, both multigene panel testing (MGPT) and exome sequencing (ES) are available. While ES can analyze novel and established IEI genes, fixed MGPTs are still broadly used. The aim of this study was to examine the molecular diagnosis (MolDx) rate from both MGPT and ES, and the phenotypic pattern of patients referred for ES.
Patients were referred for MGPT or ES between March 2017-May 2024 at a diagnostic laboratory. MGPT contained up to 574 genes and were curated based on the International Union of Immunological Societies (IUIS) phenotypic classification list of genes related to IEIs and expert opinion. Patients in the ES cohort were selected based on clinician-provided ICD-10 and Human Phenotype Ontology (HPO) terms, grouped under their top-level HPO terms. We required patients in the ES cohort to have at least one HPO term under “Abnormality of the immune system” to be included. Odds ratios (OR) and p-values were calculated using G-tests; p-values<0.05 were considered statistically significant.
The overall MolDx was higher in the ES cohort (378/2167; 17.4%) versus MGPT (3754/40994; 9.2%) (OR 2.1, $p<2.2x10^{-16}$). The most frequent top-level HPO term was “Abnormal nervous system,” so we further stratified the ES cohort into individuals with higher proportions of terms under abnormality of the immune system (ES-IS cohort) versus those with a higher proportion of terms under abnormal nervous system (ES-NS cohort). Patients in the ES-IS cohort had a lower MolDx rate (123/896; 13.7%) than the ES-NS cohort (255/1271; 20.1%; OR 1.58, p=0.0001) but a higher MolDx rate than the MGPT cohort (OR 1.58, $p = 1.1x10^{-5}$). Out of the 121 patients in the cohort tested by ES and MGPT: nine patients had a positive result in ES, but not in MGPT; eight individuals had new findings unrelated to IEIs; while one had a finding in an IEI-related gene not present in the version of the MGPT at testing. One individual had concordant results in ES and MGPT.
Despite the growing number of genes associated with IEIs, the increase in MolDx rate from ES cannot be exclusively attributed to novel IEI-related genes. This difference may be explained by the indication for testing, which suggests patients who present with an IEI phenotype and involvement with another organ system may benefit from ES. Granular characterization of the phenotypic spectrum of patients who receive a MolDx from ES is warranted.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3054T: Genetic causes of autism spectrum disorder in twins: Molecular delineation and variability in autism parameters ⭐

Authors:

A. Mohamed, O. Eid, K. Amr, R. Mahrous, M. Hemaimy, F. Hussein, M. Farid, K. Refaat, S. Hussein, N. Meguid; Natl. Res. Ctr., Cairo, Egypt

Abstract:

Introduction: Autism spectrum disorders (ASD) are neurodevelopmental disorders with significant genetic components. This study aims to uncover the genetic causes of ASD in twins and explore the differences in clinical evaluation parameters between them. **Methods**: This study included 10 twins diagnosed with ASD, including two pairs of identical twins. Diagnoses were made using the DSM-5 criteria and the Autism Diagnostic Interview-Revised (ADI-R). We conducted multiple ligation probe amplification (MLPA), chromosomal microarray (CMA), and whole exome sequencing (WES). **Results**: One identical twin had a mutation in the AGTR2 gene, located on Xq23, a strong candidate gene for autism. This mutation was inherited from their normal carrier mother. In the dizygotic twins, one twin had a duplication in the 16p11.2 critical region for autism, inherited from their normal carrier mother. We found variations in the parameters of the Childhood Autism Rating Scale (CARS) and the Autism Diagnostic Interview-Revised (ADI-R) in both identical and dizygotic twins. **Conclusion**: Molecular diagnosis of the genetic causes of autism is crucial for genetic counseling. For X-linked gene mutations, genetic counseling and preimplantation genetic testing are essential. The female protective effect against ASD may explain the presence of normal carrier mothers passing mutations to their affected sons. The differences in evaluation parameters of autism in monozygotic twins may be due to epigenetic effects. Additionally, transcriptional and methylation analyses may help identify other causes of autism.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3055T: Identification of diagnostic and potential causative variants from critically ill neonates through rapid WGS

Authors:

M-H. Park[1], J-M. Kim[1], H-W. Cho[1], D. Shin[1], H-Y. Yu[1], S. Park[2], H. Kim[2], M. Yang[2], S. Ahn[2], M-A. Jang[2], Y. Chang[2], H-Y. Park[1]; [1]NIH, Cheongju, Korea, Republic of, [2]Samsung Med. Ctr., Seoul, Korea, Republic of

Abstract:

Whole genome sequencing (WGS) is being used to uncover the genetic causes of rare diseases, including congenital anomaly (CA). When applied to patients with CA in the neonatal intensive care unit (NICU), WGS not only reveals the genetic causes but can also save the infant's life through appropriate treatment. However, conventional WGS takes a long time to obtain results, making it unsuitable for critically ill newborns. Therefore, we constructed a workflow for rapid WGS, from patient enrollment to the return of genetic diagnosis results. We established an internal pipeline using DRAGEN-GATK-Hail for trio-based WGS data analysis. The average turnaround time for eleven critically ill neonates with suspected genetic disorders was seven working days. Among them, six patients were diagnosed genetically based on OMIM database and variant pathogenicity, resulting in a diagnostic rate of 54.5%. The six diagnostic genes identified are *AGT*, *ANK1*, *TCAP*, *NSD1*, *SERPINC1*, and *OTC*. The other five patients have potential causative genetic variants, labeled as 'Variants of Uncertain Significance,' because they are novel genes that have not been associated with the patients' phenotypes, or phenotypes may not have appeared yet because the patients are newborns or infants. The results of rapid WGS for the six diagnosed patients led to changes in medical decisions, including starting immediate treatment and making changes to long-term treatment plans. Therefore, the internal pipeline for rapid WGS is a highly valuable diagnostic tool for the early identification of diagnostic variants, enabling prompt medical intervention in critically ill NICU patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3056T: Improvement of Copy Number Variant Analysis from Exome Sequencing Using a Backbone Probe Enhancement

Authors:

E. Sward[1], S. Drury[2]; [1]Geneyx, Bozeman, MT, [2]Geneyx Genomics, Sittingbourne, Israel

Abstract:

Introduction: Clinical Exome Sequencing (ES) serves as a comprehensive diagnostic tool, capable of capturing both single nucleotide variants (SNVs) and copy number variations (CNVs) within a single workflow. However, challenges remain in accurately calling CNVs from exome data due to factors such as the need for a panel of normals, low coverage regions, and presence of pseudogenes. To address these challenges and enhance CNV detection, backbone-probe enhanced exome panels have been developed. Implementing enhanced target capture, and a compatible downstream solution, ensures accurate CNV detection and a viable alternative to traditional methods like chromosomal microarrays (CMA). **Methods:** Twist Bioscience has

developed an enhanced Twist Exome 2.0 Plus Comprehensive Exome with CNV Backbone Spike-in panel, featuring 25kb, 50kb, and 100kb interval probes evenly distributed across the genome. In parallel, Geneyx has integrated the DRAGEN CNV caller from Illumina, specifically designed for the target-capture of NGS data, to detect CNV events ranging from single-exon to whole-chromosome variations. This study evaluates the effectiveness of the Twist Exome 2.0 Plus Comprehensive Exome with CNV Backbone Spike-in panel and the DRAGEN pipeline, as implemented in Geneyx, in identifying known CNVs using the Coriell CNVPANEL01 reference set, which comprises 43 'genome in a bottle' samples. **Results:** Integration of the enhanced capture panel with the DRAGEN CNV pipeline in Geneyx demonstrated ~100% concordance in detecting known CNV events. This encompassed the accurate detection of events larger than 10kb, akin to those identified by Chromosomal Microarray (CMA), as well as single-exon events smaller than 10kb, typically captured using Multiplex Ligation-dependent Probe Amplification (MLPA). Leveraging these integrated technologies reliably detected CNV events across the 43 benchmark samples and minimized low-quality CNV calls, thus validating the robustness of this combined approach. **Conclusion:** The backbone enhancement allows for more accurate localization and characterization of CNVs, thereby providing a more detailed and reliable genetic analysis. Such improvements are crucial for clinical diagnostics and research applications, as they enhance the overall ability to identify and interpret genomic variations with greater accuracy, and potentially allow ES to replace traditional methods such as CMA in a clinical setting.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3057T: Long-molecule genomics analyses elucidating the hidden structural variations associated with *MECP2* duplication syndrome

Authors:

Q. Liang[1], Y. Uchiyama[1,2], I. Kuki[3], N. Tuchida[1,2], E. Koshimizu[1], A. Fujita[1], S. Miyatake[1,4], T. Mizuguchi[1], N. Matsumoto[1]; [1]Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [2]Dept. of Rare Disease Genomics, Yokohama City Univ. Hosp., Yokohama, Japan, [3]Dept. of Pediatric Neurology, Osaka City Gen. Hosp., Osaka, Japan, [4]Dept. of Clinical Genetics, Yokohama City Univ. Hosp., Yokohama, Japan

Abstract:

MECP2 duplication disorder (MDS, MIM300260) is caused by duplication involving the *MECP2* gene. Recent short-read next-generation sequencing can detect structural variations (SVs) of the MDS region, but it has been difficult to define the precise genomic structure because it involves highly repetitive sequences. To reveal the precise SVs, we conducted exome-

based copy number variation (CNV) analysis for two MDS families in which exome their analysis found simple duplications spanning the entire *MECP2* gene: cases 1 and 2 had a 458-kb duplication and a 6.5-Mb terminal duplication both at Xq28, respectively. Optical Genome Mapping (OGM) could detect 5 major pathogenic SVs including 1 SV in case 1, and a complexed combination of 4 SVs in case 2, respectively. Subsequent nanopore sequencing with adaptive sampling could detect the precise breakpoints of all SVs except 1 inversion in case 2 which could not be solved due to the repetitive sequences. Combination of OGM and nanopore sequencing can be powerful in indetermination of the complete genomic structures associated with rare genetic disorders.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3058T: Long-read sequencing of a neurodevelopmental disorder patient reveals complex rearrangement involving the *ARID1B* gene

Authors:

T. Sneddon, M. Xiong, K. E. Weck, B. C. Powell, K. Foss, E. L. Heinzen, A. Berglind, M. Gucsavas-Calikoglu, Y. Shiloh-Malawsky, Z. Fan, S. E. Hunter; Univ. of North Carolina, Chapel Hill, NC

Abstract:

The aim of the University of North Carolina's (UNC) Genetic Determinants of Neurological and Developmental Disorders (GDNDD) IRB-approved study is to identify novel genetic variants via genome sequencing. One GDNDD participant was a 5-year-old male with a history of febrile seizures, epileptic encephalopathy, and expressive speech delay. Prior clinical genetic testing in the participant included a karyotype with a t(6:11)(q25.3;q22.3) apparently balanced translocation, a chromosomal microarray with a 1.48 Mb interstitial deletion within 4p13 deemed not significant, a normal fragile X assay, and no significant findings on epilepsy and neurotransmitter disorders gene panels.
Short-read sequence (SRS) was generated at the UNC high throughput sequencing facility (HTSF), and the data analyzed by the UNC clinical genomic analysis (GENYSIS) core facility to identify single nucleotide, copy number, and structural variants, but the case was determined to be negative. Subsequently, long-read sequence (LRS) was generated at the HTSF, and structural variants were called using the Sniffles2 algorithm. This revealed that the t(6;11) translocation disrupted the chromosome 6 *ARID1B* gene within intron 2, with a concomitant 7.7 Mb inversion at the chromosome 11 derivative breakpoint. Parentage was confirmed by DNA fingerprinting, and all three breakpoints were confirmed to be *de novo* by Sanger sequencing in the proband and both parents in the UNC clinical molecular genetics laboratory.

Loss-of-function heterozygous *de novo* variants in *ARID1B* are associated with the *ARID1B*-Related Disorder (*ARID1B*-RD) that constitutes a clinical continuum, from classic Coffin-Siris syndrome (CSS) to intellectual disability with or without nonspecific dysmorphic features. Other findings seen in individuals with *ARID1B*-RD include feeding difficulties, ophthalmologic abnormalities, hearing impairment, seizures, and autistic features. Upon review of the GDNDD participant's medical record, in addition to seizures and cognitive and speech delays, the participant had distinct facial features (relatively large head circumference, mild midface hyperplasia with upturned nose, and prominent ears) consistent with a diagnosis of *ARID1B*-RD. The Sanger-confirmed *de novo ARID1B* translocation was thus reported in the participant's medical record as a pathogenic variant.
This case highlights the benefits of LRS to detect balanced translocations, which are often missed by SRS, and to further delineate the genomic architecture underlying such complex rearrangements that likely account for some of the missing diagnostic yield from clinical genetic testing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3059T: Molecular genetic testing for rare diseases in Peru´s National Pediatric Institute: A 9-year journey

Authors:

F. Vasquez, F. Chavesta, R. Punil, M. Trubnykova, H. Abarca; INSN, Lima, Peru

Abstract:

The present work illustrates the diagnostic journey of approximately 3500 pediatric patients in a children's hospital in Peru. Patients aged 0-18 years from all parts of the country, presenting with malformation syndromes or neurological developmental disorders, were studied for the diagnosis of rare diseases. We performed several tests such as X-fragile, CMA and WES. In total, here we show the results of over 3000 CMA tests and 800 WES tests. The diagnostic yield for CMA was 29,8%, which included findings of CNVs related to syndromes such as Angelman/Prader-Willi (5,2%), Williams (3,5%), 22q11.2 deletion (4,5%). In addition, we found less known pathogenic CNVs in chromosomes 11 and 19 (1,1% y 0,6%, respectively), associated with craniofacial dysmorphia, intellectual disability, short stature, cataract, cryptorchidism, microcephaly, delayed psychomotor development, hypotonia and cardiac malformations. Some of the patients with a normal result in CMA were directed to a WES study, where the total diagnostic yield was 27.3%. We found pathogenic and likely pathogenic variants in 131 different genes. Among the most common findings, we discovered pathogenic and likely pathogenic SNVs

in *MECP2* (3.75%), *CHD7* (2.5%), *FBN1*(2.5%), *PTPN11* (2.5%) and *SCN1A* (2.5%), associated with delayed psychomotor development, intellectual disability, epilepsy, neuroregression, obesity, tall or short stature, cardiac malformations, microcephaly, craniofacial dysmorphisms, micropenis, cryptorchidism, ocular malformations, sensorineural hearing loss, cafe-au-lait macules and dysphonia. Our genetic department holds the only molecular laboratory in the country that offers tests such as CMA and WES to pediatric patients. Moreover, these tests are entirely covered by the public insurance. Since the implementation of CMA in 2015 and WES in 2020, the time needed for diagnosis since the patient is admitted to the hospital to the delivery of genetic results has been shortened. Consequently, the referral time of patients for treatment, interdisciplinary medical management or surgery has improved over the last nine years. Moreover, considering only CMA and WES, 34.2% of the patients that have obtained a genetic diagnosis through these tests were between 0-5 years old, and this percentage increases yearly, indicating that more children are having access to early genetic diagnosis. As we continue our work, we aim to improve our diagnostic yield by acquiring other genetic platforms which can aid in the search for the genetic causes of rare diseases in our patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3060T: Molecular genetic testing of congenital adrenal hyperplasia with massive parallel long-read sequencing

Authors:

Q. Lao; NIH Clinical Ctr., Bethesda, MD

Abstract:

Congenital adrenal hyperplasia (CAH) is a group of autosomal recessive diseases of steroid hormone biosynthesis that affect 1 in 200-1,000 in its mild non-classic form and 1 in 16,000 in its severe classic forms. Approximately 95% of CAH cases are due to deficiency in 21-hydroxylase (21-OHD) encoded by *CYP21A2*. Neonatal screening for 21-OHD is mandated in the U.S. and over 40 countries worldwide. To date, the diagnosis of CAH is mainly based on a hormonal test, which is sometimes ambiguous, and a costly 60-minute stimulation test is sometimes needed. Molecular genetic testing is a vital second tier means of diagnosis and is essential for genetic counseling. However, since *CYP21A2* is mapped to a genome locus of low copy repeats where highly homologous pseudogenes and their copy number variations cause severe mapping interference, standard massive parallel sequencing (next generation sequencing) platforms are not applicable. In fact, the current gold standard for 21-OHD CAH genetic testing is still Sanger sequencing of a family trio following an allele-specific-long-range PCR, which offers by far the

most comprehensive results but is very low-throughput and laborious. Successful application of massive parallel long-read sequencing technologies, e.g., PacBio Single Molecule Real Time and Oxford Nanopore MinION platforms, on the genetic testing of 21-OHD CAH has been recently reported. Although a difficult allele-specific-long-range PCR was still required, long-read sequencing based protocols demonstrated significant improvements, especially in terms of better throughput, more complete genome locus coverage, and technological convenience of haplotyping compound heterozygous variants without the necessity of parental testing, and a "one-step" determination of *CYP21A1P/CYP21A2* and *CYP21A1P-TNXA/TNXB* chimeric gene subtypes. In our study, we developed a protocol of molecular genetic testing of 21-OHD CAH based on Oxford Nanopore MinION long-read sequencing following an allele-specific-long-range PCR targeting the genome loci of interest. Compared to previous reports with relatively smaller cohort sizes (20-50 subjects), our large CAH cohort, currently managed at the National Institutes of Health Clinical Center, is composed of 420 CAH patients with a broad spectrum of genotypes. Most patients have complete medical records including family history as well as family-wide comprehensive genetic testing results, which have been confirmed by multiple existing methodologies. Our findings offer significant insights into the feasibility and accuracy of massive parallel long-read sequencing for the molecular genetic testing of 21-OHD CAH.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3062T: Racial disparities in access to a precise genetic diagnosis are not due to differences in diagnostic yields

Authors:

J. X. Chong[1,2], V. Ustach[3], R. Torene[4,3], J-H. Yu[1,5,6], K. Retterer[4,3], P. Kruszka[3], M. J. Bamshad[1,2]; [1]Univ. of Washington, Seattle, WA, [2]Brotman Baty Inst. for Precision Med., Seattle, WA, [3]GeneDx, Gaithersburg, MD, [4]Geisinger, North Bethesda, MD, [5]Treuman Katz Ctr. for Pediatric Bioethics and Palliative Care, Seattle Children's Res. Inst., Seattle, WA, [6]Inst. for Publ. Hlth.Genetics, Univ. of Washington, Seattle, WA

Abstract:

There is widespread concern that the clinical value of ES / WGS in underrepresented minorities (URM) is limited by a lower diagnostic yield and a greater number of variants of uncertain significance than found in persons labeled as "non-Hispanic white" (nHw). We evaluated the diagnostic yield of clinical ES in a cohort of ~187,000 samples including ~73,000 cases and 52,120 trios. Cases were ascribed to *self or provider assigned racial construct* (SPARC) categories based on review of ~950 labels and aggregated hierarchically into nHw (54%) and URM (46%);

50% were < 5 and 90% were < 15 years of age. Cases were classified into 30 phenotypic categories although most (67%) were referred for either abnormalities of the nervous system (ANS; n=33,374) or multiple congenital anomalies (MCA; n=15,519). Diagnostic yield was indistinguishable between nHw and URM in the full cohort (25.2% vs. 24.7%; p=0.2) and in trios (27.4% vs. 27.1%; p=0.4), as well as between groups of cases more genetically similar to either Europeans or non-Europeans from the full cohort (24.6% vs. 24.8%; p=0.5) or in trios (26.9% vs. 27.4%; p=0.2). Comparison of annual diagnostic yield since 2013 with cases classified by SPARC or genetic similarity revealed diagnostic rates in nHw vs. URM did not differ significantly over time. These results were robust when nHw were compared only to the SPARC group "black" or to a group of cases more genetically similar to the AFR sample set in 1KG. We did observe a trend of declining diagnostic yields since peaking in 2014—likely due to offering ES to patients with a broader distribution of phenotypes. Age at testing was ~5 months earlier in URM vs. nHw overall ($p=5x10^{-13}$), but not for ANS or MCA. Diagnostic yields were similar even though URM cases were submitted with slightly fewer (p=0.002) HPO terms with lower (p=0.0002) cumulative information content. After controlling for parental relatedness, the likelihood of a *de novo*, inherited autosomal dominant, or inherited X-linked PrGD did not differ between cases grouped by genetic similarity to 1KG superpopulation sample sets. Diagnostic yields using ES are similar across human populations and we hypothesize that there will be further convergence of diagnostic yield across phenotypic categories as genotype-driven approaches become standard. More importantly, the disparities in access to a PrGD are not the result of differences in diagnostic yields but are instead due to systemic and structural barriers in access to genetic testing itself.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3063T: Rapid Genome Sequencing Augments Time to Diagnosis in Prader-Willi Syndrome

Authors:

L. Robak[1,2], C. R. DiFalco[1,2], E. Mizerik[2,1], E. M. Bland[1,2], A. C. Salgado[1,2], C. Bacino[1,2], L. Vossaert[3,1], H. Dai[3,1], C. W. Wu[3,1]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Baylor Genetics, Houston, TX

Abstract:

Background: Prader-Willi Syndrome (PWS) is a well-characterized, yet complex disorder involving defects of imprinting at the paternal 15q11.2-q13 locus. Clinical features vary, but include neonatal hypotonia, transient feeding difficulty, developmental delay, and hyperphagia.

Dysmorphic features of PWS may not be apparent in infancy, necessitating a high suspicion for the disorder. Comorbid features that overlap with other genetic disorders may also obscure this diagnosis. The average age of diagnosis varies by source but has been reported between 8 weeks to childhood. DNA methylation analysis is the gold standard for PWS assessment, which usually takes several weeks to complete. Genome sequencing (GS) is emerging as an effective comprehensive testing strategy in cases with broad differential diagnoses. Here we report two cases of PWS that were diagnosed by rapid GS that directly impacted patient management. **Case 1:** A term male presented on day of life 1 with encephalopathy, hypotonia, and apnea requiring ventilator support. His clinical instability prohibited thorough physical exams and additional diagnostic steps. Due to the broad differential for these features, rapid trio GS was sent at 6 days old. GS resulted at 14 days of life with a *de novo* heterozygous deletion in paternal chromosome bands 15q11.2q13.1 (4947.3 kb) consistent with PWS. The result allowed the patient to transition to room air and enteral feeding. **Case 2:** A term male presented at 10 weeks with hypotonia and poor weight gain. He exhibited both persistent poor oral motor skills since birth and global developmental delay. Physical exam showed nonspecific dysmorphic craniofacial features and profound hypotonia, which raised concerns for both PWS and neuromuscular disorders. Rapid trio GS was sent at 11 weeks old and resulted five days later with a *de novo* heterozygous deletion in paternal chromosome bands 15q11.2q13.1 (4988.61 kb) consistent with PWS. The medical team was able to enact a home enteral feeding plan as a result. **Conclusion:** While neonatal hypotonia is a strong clinical indicator of PWS, other features are highly variable and can obscure diagnosis. Those with significant comorbidity or delayed presentation may warrant more expansive and rapid testing. GS is emerging as the optimal test choice in patients with critical illness and complex features due to comprehensive interrogation of single gene disorders, copy number variants, and parental allele identification. We demonstrated in two unique cases that GS was able to provide a rapid diagnosis of PWS that directly impacted clinical management while simultaneously ruling out other causes of their neurological deficits.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3064T: Reducing uncertainty in variant classification by leveraging the capabilities of evidence modeling

Authors:

L. Fresard, F. Facio, E. Chen, A. Colavin, A. Wahl, S. Padigepati, Y. Kobayashi, T. Manders, H. Metz, K. Ouyang, J. Reuter, B. Johnson, K. Nykamp; Invitae, San Francisco, CA

Abstract:

Variants of uncertain significance (VUS) are a significant challenge for those undergoing genetic testing. A VUS report leads to feelings of frustration, prolonged uncertainty, and lack of timely and appropriate medical care. Strategies to reduce VUS rates are critical to fully realize the utility of genetic testing for all populations. With the growth of large-scale biological data sources and modern Machine Learning (ML) techniques, predictive modeling can impact VUS rates and increase the utility of genetic testing.
To maximize the impact of predictive modeling for clinical variant classification, we developed the Invitae Evidence Modeling™ platform (EMP). The EMP is designed to accelerate the development of ML models with data inputs from large-scale data sources and systematize the calibration of prediction scores across multiple ACMG evidence categories. We present innovative approaches that maximize the strength of evidence provided by predictive models while minimizing false positives and avoid double-counting similar evidence, ensuring that the weight of clinical and functional evidence are balanced within the classification framework. More than 800,000 variants across 1 million individuals have received EMP evidence; 42% of these variants would have been VUS without predictive modeling. Importantly, definitive classifications (P, LP, LB, B) made with EMP evidence have very high prospective concordance (>99%) with ClinVar submissions. The directionality of EMP evidence (pathogenic or benign) is highly concordant (> 90%) with definitive classifications. We find high concordance between models (>85% for benign scores; >65% for pathogenic) trained with orthogonal data sources, such as population frequency compared to functional data. We demonstrate that predictive modeling has a significant impact on VUS rates by providing new types of evidence not previously specified within the ACMG guidelines. To maximize the utility and confidence of predictive modeling for variant classification, we propose to: 1) choose input data that correspond to major evidence categories specified in the guidelines, 2) perform gene-specific training and/or validation of each model type, 3) calibrate prediction scores to the strength of evidence used for variant classification.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3065T: Revealing Hidden Splicing Variants in Inherited Retinal Degeneration

Authors:

Y-S. Huang[1,2], J. Hsu[3], C-H. Yang[2,4], T-C. Chen[2,5], P-L. Chen[3,6,1]; [1]Graduate Inst. of Clinical Med., Coll. of Med., Natl. Taiwan Univ., Taipei City, Taiwan, [2]Dept. of Ophthalmology, Natl. Taiwan Univ. Hosp., Taipei City, Taiwan, [3]Graduate Inst. of Med. Genomics and Proteomics, Coll. of Med., Natl. Taiwan Univ., Taipei City, Taiwan, [4]Dept. of Ophthalmology, Coll. of Med., Natl. Taiwan

Univ., Taipei City, Taiwan, [5]Ctr. of Frontier Med., Natl. Taiwan Univ. Hosp., Taipei City, Taiwan, [6]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei City, Taiwan

Abstract:

Inherited retinal degeneration (IRD) is a group of monogenic disorders that cause retinal cell death and ultimately result in vision loss for patients. More than 300 genes have been implicated in IRD, contributing to its phenotypic heterogeneity. In a prior study, we recruited 741 Taiwanese families and identified causal variants in 57.3 % of the patients through panel sequencing. Of these patients, approximately 6.3 % harbored variants of uncertain pathogenic significance, often located in non-canonical splice sites. Moreover, disease-causing variants remained unidentified in 42.7 % of the cases. We supposed these elusive variants might reside in genomic regions affecting splicing that are ignored in standard analyses. To identify these variants and clarify the pathogenic potential of those with uncertain significance, we employed SpliceAI and dbscSNV_ADA to predict the splicing variants and designed the minigene assay for validation. By evaluating the family history and clinical symptoms of the patients, we eventually identified 31 potential disease-causing variants: 12 in the deep-intronic regions, 10 in non-canonical splice sites, and 9 in the exonic regions. Further investigation using the minigene assay revealed that over 60 % of the candidate variants exhibited alternative splicing following point mutation. Variants located in deep-intronic regions tended to generate pseudo-exons, while most variants flanking around the exons or situated within exonic regions would result in truncated exons and rarely cause exon retention. These findings underscore the necessity of using multiple splice prediction tools for analyzing splicing variants, especially those located in non-canonical splice sites or exonic regions; SpliceAI may not always be the optimal tool. This study establishes a framework for splicing variant identification by integrating computational predictions with experimental validation, thereby enhancing our understanding of the underlying etiology of splicing variants. Improved insight into splicing variants can refine diagnostic accuracy and contribute to developing potential therapies for IRD.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3067T: Sensitivity of next generation sequencing platforms in the detection of polyalanine repeat mutations in *PHOX2B*

Authors:

G. Fischer, C. Banaszynski, A. Gerlach, C. Hashimoto, E. Mares, A. Nack, C. Noller, M. Schneider, E. Sybouts, J. Truell, A. Wick; PreventionGenetics, part of Exact Sci., Marshfield, WI

Abstract:

Congenital central hypoventilation syndrome (CCHS) is a rare autosomal dominant condition resulting in poor, shallow breathing (hypoventilation) and autonomic dysregulation due to pathogenic variants in *PHOX2B*. Clinical features typically appear in newborns and less frequently as mild symptoms in children and adults. Approximately 92% of individuals with CCHS are positive for polyalanine repeat expansion mutations (PARMs) within exon 3 of *PHOX2B*. The reference sequence consists of 20 alanine repeats, comprised of four codon combinations (GCN, where N is any nucleotide). Expansion to 24 repeats or more leads to symptom onset that correlates with repeat size (Weese-Mayer et al. 2014. PubMed ID: 20301600). Due to high GC content, next generation sequencing (NGS) of this region is difficult and concurrent Sanger sequencing of this region is performed for each patient at PreventionGenetics. However, when NGS is performed on a genome platform, the algorithm ExpansionHunter has been reported to detect repeat expansions in *PHOX2B* (Dolzhenko et al. 2019. PubMed ID: 31134279). A retrospective study was conducted to determine the sensitivity of PARM detection for NGS when carried out on a probe-based (panel or exome) or genome-based platform via ExpansionHunter. We observed 19 unique PARMs confirmed via Sanger sequencing among 112 patients. Seventy-three patients had concurrent probe-based NGS testing for *PHOX2B*, representing 18 unique PARMs based on Sanger results. NGS on a probe-based platform correctly identified the PARM in 15/73 patients, indicating a sensitivity of ~20.5%. Of these 15 cases, 4 unique PARMs located in the downstream half of the repeat (residues 11-20) were observed. In contrast, 12 unique PARMS identified by genome-based testing spanned the entire repeat region (residues 2-19) and was able to correctly identify the PARM in 13/13 patients, indicating a sensitivity of 100%. Furthermore, 8 patients had both genome-based and probe-based sequencing data. In all 8 cases, the NGS probe-based platform failed to identify the PARM. These results suggest NGS-based genome sequencing is a superior screen for the detection of PARMs in *PHOX2B*. We suspect that lack of library amplification coupled with utilization of the ExpansionHunter algorithm in genomes allows for more uniform coverage and superior variant calling, respectively, within the PARM region of *PHOX2B*. As utilization of genome sequencing continues to increase, a confirmatory Sanger sequencing assay may prove unnecessary for patients undergoing testing for CCHS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3068T: SimPheny: Matching similar phenotypes across undiagnosed disease patients for genetic diagnosis

Authors:

I. Cooperstein, A. Ward, B. Moore, G. Marth; Univ. of Utah, Salt Lake City, UT

Abstract:

Introduction: Patients with rare monogenic diseases often endure prolonged diagnostic odysseys searching for pathogenic variants among the millions present in the human genome. Current computational tools that prioritize these variants for review are hindered by incomplete knowledge of phenotype-to-gene associations and the vast complexity and diversity of clinical phenotypes that frequently diverge from known disease descriptions. We present SimPheny, a novel approach that automates the identification of patients with similar phenotypes to prioritize genetic variants for diagnostic review. Here, we demonstrate the effectiveness of SimPheny using the Undiagnosed Diseases Network (UDN) patient cohort. **Methods:** SimPheny utilizes Human Phenotype Ontology (HPO) terms to calculate pairwise phenotypic similarity scores between individuals. Using Exomiser on a benchmarking cohort of 320 genetically diagnosed UDN individuals, we generated permissive gene lists of potential disease-causing variants. These gene lists were then prioritized by identifying overlaps with diagnostic genes of phenotypically similar patients, which we refer to as “gene hits.” Statistical methods were developed to evaluate the significance of each hit and the expected false discovery rate, establishing thresholds to identify likely diagnostic gene hits in the benchmarking cohort. SimPheny was then applied to prioritize genes in 1,319 as-yet unsolved cases in the difficult-to-diagnose UDN cohort. **Results:** Our semantic similarity scoring effectively measures phenotypic similarity between patients with molecular diagnoses in the same genes within the clinically diverse UDN cohort. In the benchmarking study, Exomiser ranked the true diagnostic gene in 83% of cases, but with a median of 189 ranked genes per individual. SimPheny significantly reduced the number of potential diagnostic genes per patient, achieving a manageable list of between 1 and 8 genes at varying thresholds with high precision rates. Applying SimPheny in the undiagnosed UDN cohort yielded numerous significant hits that are currently under investigation with UDN analysts and clinicians. **Conclusion:** SimPheny leverages phenotypic and genomic data to automate patient matching, demonstrating effectiveness in gene prioritization within the UDN. This method offers a promising path to improve diagnostic rates and uncover new gene associations in rare monogenic diseases. SimPheny is publicly available as an iobio web application, ready for seamless integration into diagnostic workflows.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3069T: The development of a diagnostic test method for Prader-Willi syndrome and Angelman syndrome based on targeted long-read sequencing

Authors:

J. G. Kim[1,2], K. Lee[3], **J-S. Lee**[1,2], H-y. Lee[1], H. Lim[1], J. Jang[1,2], M. Kim[1], M-W. Seong[1,2]; [1]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of, [3]Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

Background: Prader-Willi syndrome (PWS) and Angelman syndrome (AS) arise from the loss of imprinting genes on chromosome 15. Current diagnostic methods cannot detect multiple mechanisms simultaneously. We propose a novel diagnostic approach using targeted long-read sequencing (LRS) to concurrently identify diverse mechanisms. **Methods:** 6 patients with PWS and 4 patients with AS from Seoul National University Hospital were studied. NA12878 served as the normal control. A 33 Mb region of chromosome 15 was sequenced with adaptive sampling on the Oxford Nanopore GridIon, followed by targeted sequencing of a 7.9 Mb region within it on the Illumina NovaSeq 6000. Multiple pathogenic mechanisms were analyzed using Dorado, Minimap2, Pepper-Margin-DeepVariant, Sniffles, cuteSV, and CNVkit. Loss of heterozygosity (LOH) in uniparental disomy (UPD) patients was determined by comparing B allele frequency (BAF) to NA12878. Cramer's V test and z-test were utilized to identify significant differences between patients with uniparental isodisomy (iUPD) or uniparental heterodisomy (hUPD) and controls, and to subsequently validate these findings in UPD patients with Beckwith-Wiedemann syndrome. **Results:** We identified differentially methylated regions in all 10 patients. Enriched LRS detected various pathological mechanisms, including large deletions (2 with PWS and 1 with AS), small deletion (1 with AS), hUPDs (2 with PWS and 1 with AS), iUPDs (2 with PWS), and single nucleotide variant (SNV) (1 with AS). Integrating short-read sequencing data (SRS) enhanced coverage depth, particularly refining the analysis of large deletions exceeding 5-6 Mb. BAF analysis accurately identified LOH in 5 UPD patients (4 with PWS and 1 with AS). Significant differences in iUPD patients compared to the normal control were validated by both Cramer's V test and z-test, while hUPD patients showed no significant distinctions. Applying our statistical model to the UPD patients with Beckwith-Wiedemann syndrome, those with iUPD with a LOH region on chromosome 11 exhibited significant differences from normal data, increasing the reliability of our statistical model. **Conclusions:** We illustrated the simultaneous analysis of allele-specific methylation, UPD, structural variant, and SNV via targeted enriched LRS. Integrating SRS improved low coverage, presenting various sequencing scenarios. Using an in-house statistical model, the accurate detection of LOH and differential diagnosis of iUPD from hUPD were achieved. Additional validation is essential to confirm the clinical utility of the novel diagnostic method.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3071T: The twist of *TP53*: Rare case report of 17p13.1 multigene deletion and Li-Fraumeni syndrome

Authors:

T. Porter[1], E. Macke[1], R. Hokanson[1], L. Hernandez[1], K. Zajo[1], Y. Akkari[1], K. Manickam[2], J. Hunter[2]; [1]Nationwide Children's Hosp., Columbus, OH, [2]Nationwide Children s Hosp., Columbus, OH

Abstract:

Li-Fraumeni syndrome (LFS) is a hereditary cancer predisposition syndrome resulting from alterations in *TP53*. LFS is characterized by a high incidence of breast cancer, sarcomas, brain tumors, leukemia, adrenocortical carcinoma, and melanoma, usually occurring in childhood or young adulthood. Most *TP53* alterations described are single nucleotide polymorphisms or partial gene deletions, with complete loss of *TP53* being less common. Multigene deletion of the 17p13 region, including *TP53*, has been described to result in lesser cancer risk compared to other reported *TP53* variants. To date, reports of 17p13 microdeletion syndrome primarily detail pediatric cases with limited follow-up. Here we describe a 27-year-old male patient with a history of intellectual disability, developmental delay, macrocephaly, dysmorphic features, and multiple early-onset malignant melanomas. This individual followed in the genetics clinic intermittently from the age of seven and received a stepwise approach to genetic testing. At age 12, he had a CHIP microarray and *PTEN* mutation analysis performed, both of which were negative. At age 15, testing for *PTCH1* and *SUFU* was recommended due to concern for Gorlin syndrome, but was not performed. At age 25, exome sequencing (ES) and SNP microarray were performed. ES was negative for causal variants. SNP microarray revealed a 522.54 kb pathogenic loss of 17p13.1 including *TP53*. Features of 17p13.1 microdeletion syndrome include developmental delay, dysmorphic facial features, hypotonia, and additional malformations, which was fitting to the patient phenotype. Given the loss of *TP53*, a referral to Hem-Onc ensued for appropriate follow-up. He had a baseline brain magnetic resonance imaging (MRI), whole-body MRI, and abdominal ultrasound in 2023. The brain MRI revealed a 3.8 cm tumor presumed to be a low-grade glioma, and surgical resection occurred in January 2024. Tumor-normal ES revealed the previously identified germline *TP53* loss, as well as bi-allelic loss of *TP53* in the brain tumor sample. This case highlights the limitations of ES, particularly for detecting copy number variation. Furthermore, there is a lack of data regarding adult patients with 17p13.1 microdeletion syndrome and their cancer risk, with some reports indicating these individuals are not at an increased risk for cancer predisposition. This highlights significant challenges and knowledge gaps in genetic counseling and clinical surveillance for these individuals. The tumor-normal ES described here suggests that the *TP53* germline loss, and

subsequent bi-allelic loss in the tumor are indicative of LFS in an individual with 17p13.1 microdeletion syndrome.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3072T: Two intronic variants in the SDCCAG8 gene end a patient journey

Authors:

L. Hoefsloot[1], N. A. Kuehn[1], E. Kasteleijn[1], A. I. Iglesias Gonzales[1], H. Zondervan-van der Linde[1], M. Hoogeveen-Westerveld[1], A. Kievit[2], T. J. van Ham[1]; [1]Dept Clinical Genetics, Erasmus MC, Rotterdam, Netherlands, [2]Erasmus MC, Rotterdam, Netherlands

Abstract:

We illustrate how the addition of new techniques for the analysis of genetic variants ultimately results in a diagnosis in a case that was known to us for many years. The patient is a now 39 yr. old female who has been struggling with progressive night blindness starting around 12 years of age. She subsequently developed additional symptoms, including kidney failure and hearing loss. Differential diagnostic considerations were mitochondrial or metabolic disease or ciliopathy. A broad array of metabolic tests on plasma, urine, and a muscle biopsy did not show abnormalities. Exome sequencing and analysis did not yield a candidate causal variant that could explain all the clinical symptoms, except from a variant in the WFS1 gene (NM_006005.3:c.2327A>T; p.Glu776Val) possibly explaining the hearing loss. However, upon re-evaluation this variant was classified as likely benign. Structural cilia analysis showed a minor decrease in ciliated cells once and a normal percentage upon repetition of the experiment (47 and 60 % cells with cilia ≥ 3 µm, respectively, where > 50% would be considered not suspicious). RNA-sequencing was subsequently performed to identify a genetic cause, which showed reduced expression of *SDCCAG8* (NM_006642.5) from exon 8 onward, while expression levels for earlier exons fall within the average range for controls. Visual inspection of the RNA sequencing data showed an alternative exon right before the drop in expression of exon 8. We found two deep intronic *SDCCAG8* variants *in trans* located inside this alternative exon. *In silico* predictions for the variants indicated that the first variant, c.740+300C>T, according to ESEfinder, disrupts an SF2/ASF recognition site with simultaneous creation of a new one with a lower score. For the c.740+357G>A variant, several recognition sites (two SC35 sites and one SF2/ASF (IgM-BRCA1) site) could be disrupted, and an additional potential branch point created. Strikingly, Otto et al. (Nat genetics, 2010), reported a bi-allelic variant in SDCCAG8, c.740+356C>T, in an individual with a highly similar phenotype. The compound heterozygous

SDCCAG8 variants provide a molecular diagnosis, explaining the renal and eye symptoms of the patient, compatible with Senior-Loken syndrome 7 (OMIM 613615).

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3073T: Unearthing clinical genetic test data concealed in the electronic health record to quantify progress in genetic medicine

Authors:

L. Bastarache[1], R. J. Tinker[2], B. A. Schuler[1], L. Richter[1], J. A. Phillips, III[3], W. W. Stead[1], G. Hooker[4], J. F. Peterson[1], D. M. Ruderfer[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]VUMC, Nashville, TN, [3]Vanderbilt Univ. Sch. Med, Nashville, TN, [4]Concert Genetics, Nashville, TN

Abstract:

Introduction: Genetic testing has grown significantly in the last twenty year. However, a full accounting of the growth and impact of genetic testing across clinical medicine has been challenging due to non-standard representations in electronic health records (EHRs). Moreover, population-based measures that capture a key purpose of genetic testing - to identify a clear cause for medical signs and symptoms - are lacking. **Methods:** We extracted germline genetic test results from the EHRs of over 1.8 million patients seen at an academic medical center from 2002 to 2022, using a combination of custom parsers and manual chart review. We use these data to assess trends of genetic test utilization over 21 years and developed new measure - the genetic attributable fraction (GAF) - to assess the proportion of phenotypes in a population that can be attributed to genetic diagnoses. **Results**: Our efforts revealed a wealth of genetic data, much of which was buried in the free text of clinical notes, including 103,358 genetic tests and 19,032 molecularly confirmed diagnoses. Genetic testing rates increased over the study period, from 1.0% in 2002 to 6.1% in 2022, and testing became more comprehensive with the growing use of multigene panels and exome/genome sequencing**.** These changes corresponded with a substantial increase in the number of different diseases diagnosed with genetic testing, from 51 in 2002 to 509 in 2022, and an increasing number of variants of uncertain significance (VUS). We also found evidence of the growing influence of clinical genetic testing in medicine, with testing rates >10% across a variety of inpatient clinics in 2002, including oncology, nephrology, and several pediatric subspecialty clinics. The phenome-wide GAF of 6,505,620 phenotypes observed in 2022 was 0.46%, with 74 individual phenotypes having a GAF >5%, including pancreatic insufficiency (67%), chorea (64%), atrial septal defect (24%), Microcephaly (17%), paraganglioma (17%), and ovarian cancer (6.8%). **Conclusion**: By standardizing genetic test results across a large cohort, we were able to quantify trends of genetic testing over two

decades. Our results show that genetic testing is increasingly a part of patient care. The growing ability of genetic tests to diagnose disease or explain phenotypes raises the importance of genetic testing for routine diagnostic care, while the increase in VUS reporting represents a growing challenge.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3074T: Utilizing whole exome sequencing to uncover genetic diagnoses in prolonged neonatal/infantile intrahepatic cholestasis

Authors:

T. Tanimura[1], S. Ito[1], K. Imagawa[2], H. Hayashi[3], K. Ito[1], T. Togawa[1], S. Saitoh[4]; [1]Nagoya City Univ., Nagoya, Japan, [2]Dept. of Child Hlth., Inst. of Med., Univ. of Tsukuba, Tsukuba, Japan, [3]Lab. of Molecular Pharmacokinetics, Graduate Sch. of Pharmaceutical Sci., The Univ. of Tokyo, Tokyo, Japan, [4]Nagoya City Univ, Nagoya, Japan

Abstract:

Neonatal/infantile intrahepatic cholestasis (NIIC) is a condition that requires early diagnosis and treatment and encompasses many genetic disorders, including cystic fibrosis, alpha-1 antitrypsin deficiency, and progressive familial intrahepatic cholestasis (PFIC). Owning to the similarity in clinical symptoms and the diverse range of differential diagnoses, establishing an accurate diagnosis is often challenging. In a previous publication, we documented a conclusive molecular diagnosis in 27% of the patients enrolled in our target 61-gene panel study, using the Ion Personal Genome Machine system (Thermo Fisher Scientific) in conjunction with bioinformatics pipelines. However, approximately 70% of patients remained without a definitive genetic diagnosis. In a subset of these undiagnosed cases, we performed whole exome sequencing (WES) on 18 probands and presented the findings at ASHG annual meeting 2023. To increase the diagnostic yield for cholestatic patients, we performed additional WES on undiagnosed cases from the same cohort. From May 2013 to April 2024, we included 37 probands and their parents in our study, all of whom had prolonged cholestasis with no definitive genetic diagnosis after the targeted sequencing. We performed WES using the Illumina platform. Of the 37 probands, 8 (21.6%) received a definitive molecular genetic diagnosis. The identified causative variants were in the following genes: *AKR1D1* (Bile acid synthetic defect 2), *CPS1* (Carbamoylphosphate synthetase I deficiency), *HNF1B* (Renal cysts and diabetes syndrome), *LSR* (Lipolysis stimulated lipoprotein receptor deficiency), *MTM1*(Myotubular Myopathy), *RELA* (Autoinflammatory disease, familial, Behcet-like 3), *SCYL1* (Spinocerebellar ataxia, autosomal recessive 21) and *USP53* (PFIC type 7). Seven of

these eight genes were not included in our initial 61-gene panel, except for *AKR1D1*. In conclusion, WES is useful for the genetic diagnosis of very rare diseases that are not previously known to present with NIIC. Based on this initial success, we are currently striving to uncover a novel chromosomal structural variant through additional bioinformatics pipelines.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3075T: Variant classification is impacted by the presence of biochemical data in published patient reports: A case study for Familial Hypercholesterolemia

Authors:

B. Mathias, M. Mahinic, P. Romero, Z. Young, J. Farmer Bugarin, E. Wiley, A. Grewal, A. McGill, **J. Bissonnette**, A. Kozaric, M. Kiel; Genomenon, Ann Arbor, MI

Abstract:

Familial hypercholesterolemia (FH) is an inherited disorder characterized by high levels of low-density lipoprotein cholesterol (LDL-C), which results in the early development of atherosclerotic plaques in the coronary arteries, raising the risk of early cardiovascular events. Criteria for diagnosis of FH include family history, elevated LDL-C levels >190 mg/dL, the presence of xanthomas, and genetic analysis of *LDLR*, *APOB*, *PCSK9*, and *LDLRAP1* genes. *LDLR* pathogenic/likely pathogenic (P/LP) variants are associated with >50% of all FH patients. However, the biochemical values such as LDL-C levels are often not reported in genetic studies. This study investigates the prevalence of biochemical data omission in publications relating to FH and its impact on variant classification. A semi-automated curation with manual expert review was used to collect genetic variants and biochemical values from literature to classify variants according to the ACMG/AMP criteria. A total of 42,032 FH patients were collected with a heterozygous *LDLR* variant. Of those, only 3131 (7.4%) FH patients had a reported value for LDL-C levels. Out of 3131 FH patients with a reported LDL-C value, 2572 FH patients (82%) had LDL-C levels above the threshold (>190 mg/dL). For those patients who had a reported LDL-C value, 2237 FH patients (87%) had a P/LP variant and 333 patients (13%) had VUS. Patients with a high LDL-C were more likely to have a P/LP variant identified than a VUS. In addition to patient data, *LDLR* variant data was analyzed in association with LDL-C values. Collected data consisted of a total of 2311 unique *LDLR* variants including 1359 P/LP variants (59%) and 950 VUS (41%). Only 839 (36%) variants had LDL-C values reported in literature. Furthermore, of those variants, only 745 (87%) were associated with LDL-C values >190 mg/dL. Out of 745 variants with LDL-C values >190 mg/dL, 537 variants were classified as P/LP and 208

variants as VUS. Thus, after applying a threshold of 190 mg/dl, the number of variants with data that met the threshold was reduced to 40% and 22% respectively for P/LP and VUS variants. In assessing the variant classification, we included patients with sufficient information to support a FH diagnosis regardless of the presence of biochemical data. If the cohort of patients without biochemical data was excluded from the variant assessment, our data indicates this would result in 70% of the *LDLR* variants being likely downgraded. This study underscores the need for standardized reporting for biochemical genetic conditions to increase variant curation accuracy, enhance patient identification, and positively impact the management and care of the rare disease community.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3076T: Whole exome sequencing resulted in a definitive diagnosis for individuals exhibiting common global developmental delay and intellectual disability at early stages

Authors:

L. Wang, E. Jacoby, D. Stubbs, N. Agrawal, S. Bell, M. Vilt, A. Popa, P. Mayo; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Background: Global developmental delay (GDD) and intellectual disability (ID) are common clinical features of rare disorders in pediatric patients. Historically, ACMG recommended chromosomal microarray analysis (CMA) as first tier testing for patients with GDD/ID. However, this testing approach will not identify rare disorders caused by single genes. By using whole exome sequencing (WES), we identified molecular causes for two individuals with undiagnosed developmental disorders and uninformative CMA. **Methods:** The whole exome was captured using IDT X-gen Exome (Integrated DNA Technologies) followed by sequencing on the Illumina NovaSeq platform. WES was performed on the family trio samples, and clinically significant variants were confirmed by Sanger sequencing. **Results:** Case 1: 1-year-old male with a clinical history of GDD, speech delay, febrile seizure, short stature, hypotonia, and late tooth eruption. A heterozygous, de novo variant (NM_014795.3: c.3042del, p.Leu1015Cysfs*60) in the *ZEB2* gene was identified via WES. Case 2: 3-year-old female with a clinical history of GDD, ID, hypotonia, mixed receptive-expressive language disorder, plagiocephaly, and abnormal findings on brain MRI. A heterozygous, de novo variant (NM_078480.2:c.590_591insT, p.Arg198Glnfs*23) in the *PUF60* gene was identified via WES. **Discussion:** In case 1, the 3042del (p.Leu1015Cysfs*60) variant in the *ZEB2* gene is novel and was classified as likely pathogenic.

This finding supports a molecular diagnosis of Mowat-Wilson syndrome (MOWS). In Case 2, the finding of the novel c.590_591insT (p.Arg198Glnfs*23) pathogenic variant in the *PUF60* gene supports a molecular diagnosis of Verheij syndrome (VRJS). GDD and ID are common and early manifestations in individuals with MOWS and VRJS; however, they are not specifically diagnostic. Other unique MOWS or VRJS -associated phenotypes usually present later in life. Therefore, a MOWS or VRJS diagnosis can only be established by detection of a pathogenic variant in the *ZEB2* or *PUF60* gene. **Conclusion:** Clinical assessment and CMA testing for individuals with common GDD and ID are non-diagnostic for many rare disorders. In this study, WES identified novel, de novo pathogenic variants establishing a definitive diagnosis for two GDD/ID patients at early stages. This provides further evidence that trio WES is a cost-effective strategy that can shorten the diagnostic odyssey for patients with GDD/ID; WES should be considered as first line testing for patients with non-specific GDD/ID. Given the complexity of rare disorders, WES can be performed in parallel with CMA in the affected patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3077T: Whole-genome, single-cell DNA sequencing to identify de novo copy number mutations in individual sperm cells

Authors:

T. Nicholas, A. Quinlan, X. Yang; Univ. of Utah, Salt Lake City, UT

Abstract:

Spontaneous mutations affecting DNA copy number contribute to diverse developmental disorders and human phenotypes. Previous work from our group studied multi-generational pedigrees and family-based trios and showed that 1 in 6 live births harbors a spontaneous structural variant (SV). Furthermore, aneuploidies and SV changes have been observed in sperm samples and are linked to reduced sperm motility, suggesting that these sperm are less likely to result in viable offspring. While FISH has commonly been used to identify sperm aneuploidies and larger structural changes, this method has limited resolution. With the availability of whole-genome single-cell sequencing using Primary Template-directed Amplification, which enables higher and more even genomic coverage, detecting de novo SV changes in individual sperm cells is, in principle, possible at higher resolution and with substantially reduced bias compared to prior whole-genome amplification techniques.
Using BioSkryb ResolveDNA single-cell sequencing, we have obtained whole-genome sequencing data from 64 individual sperm cells derived from 4 different donors. We also performed deep whole-genome sequencing on bulk sperm and blood samples as controls from

the same donors. With this data, we aim to demonstrate the feasibility of calling copy number mutations in individual gametes. Through comparisons to the blood controls, we anticipate being able to distinguish CNVs that are unique to the sperm samples from these donors. Specifically, we expect to identify CNVs that are private to individual sperm cells within each donor. Additionally, we plan to detect more common CNVs that are present in more than one sperm and likely arose in a specific spermatogonial stem cell lineage. With this information, we will compare rates of de novo CNV formation between individual sperm cells to rates derived from ostensibly reproductively fit sperm analyzed in healthy pedigrees. This study also motivates future work comparing these rates across differing ages and environmental exposures, such as smoking.

Session Title: Molecular and Cytogenetic Diagnostics Poster Thursday Session

Board 3078T: Williams Beuren Syndrome, 11 year clinical follow up, cytogenetic study, and a neuropsychological approach: case report

Authors:

C-D. Reyes-Rodriguez[1,2], A. Rivera-Cameras[1], A. Ramirez-Velazco[1], G. Rodriguez-Gonzalez[3], I. S. Davalos-Higareda[4], M. Dominguez-Quezada[1,2], I. P. Davalos-Rodriguez[1,2]; [1]Doctorado en Genética Humana, CUCS, Univ. de Guadalajara, Guadalajara, Mexico, [2]División de Genética, CIBO, Inst. Mexicano del Seguro Social, Guadalajara, Mexico, [3]Hosp. Regional de Alta Especialidad de Zumpango, Zumpango, Mexico, [4]Especialidad en Psiquiatria, Clinica Psiquiatrica Dr. Everardo Neumann Peña, UASLP, San Luis Potosi, Mexico

Abstract:

Introduction: Williams Beuren Syndrome (WBS) (OMIM#194050) has a prevalence of 1 in 7,000 to 25,000 births and is caused by a deletion of 1.5 to 1.8 Mb on chromosome 7q11.23 which contains approximately 25 to 28 genes. WBS clinical characteristics include dysmorphic facial features, cardiovascular abnormalities, and distinctive cognitive, social, and psychological profiles. Objective: To present a patient whose clinical and cytogenetic data led to the diagnosis of WBS, his clinical follow up over 11 years, and a neuropsychological approach. Case Report: A 13 year old male patient, second pregnancy, full term, delivered by cesarean section, normal pregnancy, mother aged 27 and father aged 32 at birth, non-consanguineous and he has two brothers apparently healthy. At 15 months, a cardiological evaluation reported mild supravalvular aortic pulmonary stenosis, physical examination revealed psychomotor developmental delay, typical WBS facial features, convergent strabismus, and bilateral cryptorchidism. The cytogenetic study reported karyotype 46,XY[16].ish

del(7)(q11.23q11.23)(ELN-)[5]. At 2 years, orchidopexy was performed, at 5 years, strabismus was corrected, at 7 years, right renal hypoplasia was diagnosed, and at 10 years, neuropsychological assessment revealed intellectual disability, comprehension language impairment, decreased social, visuomotor, and visuospatial skills. Currently, nephrology reports preserved renal function, cardiology reports no symptoms, and endocrinology reports no abnormalities; however, cognitive, and psychosocial disabilities have become more pronounced. Conclusions: Early clinical diagnosis in patients with WBS allows for a multidisciplinary management and opportune treatment of complications. In patients with WBS, it is essential to focus on social, cognitive, and psychological aspects in WBS to promote timely interventions to improve their quality of life.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6002T: A concordance study between molecular and HPLC-based methods in diagnosing sickle cell disease

Authors:

A. Kumari, S. Shakya, P. K. Chaudhuri, A. Prasad; Rajendra Inst. of Med. Sci. Ranchi, Ranchi, India

Abstract:

Background: In Sickle Cell Disease (SCD), over 20 techniques have been either developed or are in the process of being developed for its diagnosis. While traditional diagnostic methods have been effective in identifying individuals with the advancements in genetic diagnostic facilities Sanger sequencing provides a relatively more precise and reliable means of confirming the presence of the SCD subtypes. This study aimed to utilize Sanger sequencing to validate the sickle cell disease diagnoses of individuals previously identified through HPLC. **Method:** We selected a total of 85 patients, out of these 85 patients, 46 had HbSS, 23 had HbS beta, 15 had HbS trait, and 01 had HbSE from a cohort previously diagnosed with sickle cell disease based on HPLC analysis. The HPLC findings were juxtaposed with the genotyping data to evaluate concordance between these two methodologies. A threshold of A2 more than 3.5 from HPLC was used to determine the presence of HbS beta. We considered the report as “false positives” when HPLC indicated HbSS but the genotype was HbS beta or vice versa. Sanger sequencing was performed for each patient using primers flanking the *HBB* exon-1, covering prevalent mutations such as HbS, HbSC, HbSE, IVS1.1, and IVS1.5. The sequencing data was analyzed using DNASTAR-SeqMan software. **Results:** Sanger sequencing revealed the cohort of SCD patients comprised 52 HbSS (60%), 17 HbS beta (20%), 15 HbS trait (18%), and 1 HbE beta (1%).

Comparing the HPLC reports with the genotyping data, we primarily observed non-concordance between HbSS and HbS beta. The results indicated that HPLC had 100% sensitivity in diagnosing SCD but only 50% specificity in terms of detecting SCD subtypes. In 15 cases (18%), predominantly involving HbSS and HbS beta, the HPLC method misdiagnosed the SCD subtype. **Conclusion:** The Sanger sequencing demonstrated a high level of specificity in diagnosing SCD subtypes compared to HPLC. The genotyping approach offers a more comprehensive assessment of the specific hemoglobin mutations responsible for the disease, enhancing the accuracy of the diagnosis. The genetic diversity observed in sickle cell anemia, which includes the influence of concurrent beta thalassemia, can lead to variations in clinical symptoms and disease severity. This approach enables the precise identification of the molecular defect responsible for the sickle cell phenotype, facilitating more focused/ customized clinical interventions and deepening our comprehension of genotype-phenotype relationships in this intricate disorder.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6003T: A framework to characterize rare recessive diseases in diverse populations using biobank-scale data

Authors:

C. Caggiano[1], R. Shemirani[1], A. Biji[1], T. Bergquist[1], S. Asgari[1], V. Pejaver[1], G. Belbin[2], E. Kenny[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Gencove Inc., New York, NY

Abstract:

Intro: Understanding the genetics of rare recessive diseases (RRDs) is crucial for improving clinical care. RRDs are typically characterized in specialty care; however, not all populations receive this care equally and knowledge of these disorders is biased toward those that affect European ancestry groups. To prevent inequities, RRDs need to be further studied in diverse populations. Biobanks offer a unique opportunity to study RRDs by combining health records and genetic data outside of specialty contexts. We hypothesized that a framework leveraging elevated genetic sharing in founder populations would increase power and identify genomic regions associated with disease in understudied populations, giving greater insight into RRDs. **Methods:** Using a novel identity by descent (IBD) clique detection approach, iCURL, we detected more than 18 million IBD cliques in 57600 participants from the BioMe Biobank in NYC. Founder populations were identified using IBD enrichment. Then, we developed a recessive association testing framework that used phenome-wide electronic health data of 7520 disease traits and cliques of haplotypic IBD sharing within founder populations. Then, we evaluated associated

cliques by comparing to known pathogenic recessive variants within the region to highlight potential genetic mechanisms.
Results: In total, for 7 diverse founder populations (N=157 to 8320), we identified 17290 significant clique-trait pairs for 587 unique phenotypes. The results were enriched for congenital and genetic traits (p=1e-16). We recapitulated several known founder variants. For example, there was an association between cliques at *PSEN* and hereditary Alzheimer's disease in Puerto Ricans (p=4e-7) and *LRBA* and lymphocytosis in Dominicans (p=8e-4). Furthermore, in Puerto Ricans, we found an association between having 2 copies of *COL27A1*.pG697R, a variant for Steel syndrome, and congenital hip deformity, a known disease symptom (p=3e-7). Potential misdiagnoses were also identified. For instance, *COL27A1*.pG697R was associated with Klippel-Feil syndrome (p=6e-5), which has symptom overlap with Steel, but different etiology. Additionally, a clique at *MME*, causal for Charcot Marie Tooth disease, was associated with myasthenia gravis (p=7e-7). Both have muscle weakness symptoms but are distinct diseases. Lastly, we identified new RRD associations, motivating future study
Conclusions: This work represents an advance in characterizing RRDs in diverse and understudied founder populations. This may have implications for genomic medicine initiatives that aim to equitably implement genetic screening programs and improve diagnostic accuracy.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6004T: A Functional Analysis of Ramp Sequences in Data from the Clinical Proteomic Tumor Analysis Consortium

Authors:

G. Johnson, M. Cloward, M. Wilson, P. Ridge; Brigham Young Univ., Provo, UT

Abstract:

The impact of synonymous variants is often overlooked. However, a new perspective is emerging, suggesting that even common synonymous variants can influence transcription and translation. Ramp sequences are a novel and potentially significant functional element. Ramp sequences are slowly translated, or inefficient, codons clustered consecutively at the 5'-end of the coding regions of some genes. The efficiency, or speed, with which a codon is translated is highly correlated with tRNA abundances; since tRNA abundances vary across tissues, some genes in some tissues may have a ramp sequence, but not in others. Counterintuitively, ramp sequences generally increase translational efficiency, and genes with ramp sequences are often highly expressed compared to genes without a ramp sequence. Like a metered on-ramp on a highway, these sequences may slow ribosomal movement at the start of

translation, resulting in more even spacing of ribosomes on the transcript, potentially preventing ribosomal collisions and leading to enhanced translational efficiency. Finally, the presence/absence of ramp sequences is conserved, even if the actual gene sequence is not. Our objective is to characterize the role of ramp sequences. The dataset developed by the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) has extensive data for each participant, including whole genome sequence data, mRNA expression levels, and protein levels in paired tumor/normal cells. These data provide a unique opportunity to explore how ramp changes affect expression and translation within a single individual. We are analyzing data from head/neck (110 participants), lung (104 participants), kidney (109 participants), and pancreas cancer (139 participants). We previously developed ExtRamp (Miller et al., 2019, NAR; version two is presented at ASHG: Cloward et al.). Since multiple variants in cis can affect ramp status (i.e., whether or not a gene has a ramp sequence), we estimated haplotypes for each gene in each individual and next determined the ramp status of each gene. We focused on instances where the ramp status of a given gene in an individual is different in their paired samples. We found 7,632 instances where a ramp present in normal tissue was lost in cancer tissue and 8,074 instances where a ramp sequence absent in normal tissue was gained in cancer tissue. Our preliminary analyses revealed patterns consistent with what we expected. We are excited to share our detailed findings at the conference.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6005T: A genome-wide association study identifies genetic regulators of metabolism at birth

Authors:

B. Snyder[1], C. McKennan[2], N. Schoettler[3], T. Gebretsadik[1], W. Dupont[1], C. Ober[3], T. Hartert[1], for the ECHO-CREW investigators; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Univ. of Chicago, Chicago, IL

Abstract:

Background: Integration of multiple levels of molecular data can provide a more comprehensive assessment of disease mechanisms through identification of upstream/downstream targets, intermediate processes, and mediating effects. We aimed to determine the genetic contribution of variation in metabolite concentrations at birth, a period critical to understanding disease origins. *Methods:* Our study population included infants enrolled in the INSPIRE population-based birth cohort (n=1,414) with linked targeted newborn screening metabolic data and genotypes. Separate GWASs for each metabolite (n=39) and related metabolite ratios (n=14)

were conducted using multivariable linear or ordinal regression, depending on metabolite distributions. We performed analyses for each ancestral group (European, African, and Hispanic Americans) and reported the meta-analyzed results. Lead variants in each genomic region were tested for replication in two independent birth cohorts of European ancestry (MAAP: n=29, WISC: n=139; meta-analyzed results reported). *Results:* Of the 284,247,533 variant-metabolite associations tested (5,363,161 variants*53 metabolites/ratios), 1,934 associations reached genome- and metabolome-wide significance ($p<5x10^{-8}/53=9x10^{-10}$). The significantly associated variants mapped to six genomic regions (1p31.1, 5q31.1, 6q21, 9q34.11, 11q13.2, 12q24.31). The most significant variants, at 12q24.31, were associated with C4 acylcarnitine concentrations (lead variant: rs12829722, β=0.30, $p=1.7x10^{-57}$). The lead variants in the 1p31.1, 9q34.11, 11q13.2, and 12q24.31 regions replicated (p<0.05/6=0.008). *Conclusions:* We characterized the genetic contribution to inter-individual differences in metabolite concentrations at birth. Our findings suggest that metabolite concentrations are regulated by genotype and highlight the promise of metabolites as functional intermediate phenotypes in disease pathways originating at birth.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6006T: A genome-wide association study of human blood metabolite levels in the Japanese population elucidates genetic metabolic mechanisms

Authors:

Y. Takeuchi[1], M. Koido[1], Y. He[1], N. Itokawa[1], X. Liu[1], A. Nagai[2], N. Kumasaka[3], Y. Okada[4], T. Morisaki[2], K. Matsuda[2], Y. Kamatani[1]; [1]Lab. of Complex Trait Genomics, Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Tokyo, Japan, [2]Lab. of Clinical Genome Sequence, Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Tokyo, Japan, [3]Div. of Digital Genomics, Human Genome Ctr., Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan, [4]Dept. of Genome Informatics, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan

Abstract:

Human blood metabolite levels are influenced by various factors, including genetic factors. Metabolomic genome-wide association studies (GWASs) in the European population (EUR) have provided biological insights. Using nuclear magnetic resonance (NMR) measurements from serum samples, here we conducted a large-scale GWAS for 248 metabolome quantitative traits from 37,969 Japanese individuals in Biobank Japan (BBJ). We identified 2,140 locus-metabolite pairs, including 53 novel pairs ($P < 5\times10^{-8}/20$, 20 is the number of principal components

explaining 95% of the variance in metabolite levels; Bonferroni correction). Leveraging size-stratified lipoprotein measurements by NMR, we found that 488 variant-lipoprotein pairs showed heterogeneity in effect sizes (Cochran Q test $P<6.1\times10^{-5}$; Bonferroni correction). Notably, we identified novel associations between rs75326924 (missense variant in *CD36*) and the serum levels of small and medium high-density lipoprotein (HDL), but those of larger HDL were not associated. This variant is low-frequency in the East Asian population (EAS) but rarely observed in other populations, likely causing EAS-specific CD36 deficiency linked with various metabolic conditions. Using GWAS results for 115,082 individuals from UK Biobank (UKB), we identified highly correlated effect sizes across populations (Pearson correlation coefficients of 0.95 for the lead variants in BBJ and 0.81 for those in UKB). Through the meta-analysis, we identified the additional 80 novel pairs. Furthermore, a gene-based analysis leveraging bulk tissue and single-cell transcriptome profiles found the enrichments of genetic associations in 200 metabolite-tissue pairs in either cohort (false discovery rate < 5%). For example, the associations for lipoproteins were most frequently observed in liver tissue (178 pairs), particularly liver zonation near the central vein. Finally, we focused on the metabolomic scores as the weighted sum of 36 metabolite levels to predict disease onset in EUR in a previous study. We calculated the scores for five diseases (stroke, myocardial infarction, lung cancer, liver cirrhosis, and colorectal cancer) in BBJ and found that myocardial infarction and liver cirrhosis showed significant hazard ratios for the top 10% ($P < 0.05/5$; Bonferroni correction). In summary, this large-scale study using blood metabolite levels in the Japanese population demonstrates associations between CD36 deficiency and lipoproteins, high consistency of the genetic effects across populations, genetic links between metabolites and tissues, and the utility of the measurements for predicting disease onset.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6007T: A novel ancestry-specific eQTL resource of healthy lung tissue from African Americans

Authors:

J. Lifferth[1], H. Chen[2], A. Petty[1], B. Ryan[3], J. E. Below[1], M. Aldrich[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Academia Sinica, Taipei, Taiwan, [3]Natl. Cancer Inst., Bethesda, MD

Abstract:

Genetic variation between populations of diverse genetic ancestry can result in distinct patterns of gene expression. Prior studies have identified lung eQTL, but the role of ancestry-related differences in lung eQTL remains relatively unexplored. To investigate the relationship between

genetic ancestry and eQTL variation, we applied a pipeline for the identification of local ancestry-specific eQTL in tumor-adjacent lung tissue collected from self-identified African American individuals. We describe the ancestry patterns and RNA sequencing data for this unique resource.
Blood biospecimens and lung tissues were collected from an ongoing case-control study of lung cancer. Participants self-identifying as Black or African American and having both a blood biospecimen and lung tissue were prioritized for further follow-up. Samples were genotyped using the Illumina MEGA genotyping array and bulk RNA sequencing was performed on the lung tissue (n=80). Two-way global ancestry was inferred using ADMIXTURE and reference populations from the 1000 Genomes Project. Local ancestry was estimated using RFMix2 with 1,757 reference samples (gnomAD v3.1.2). Genotyping data were phased and imputed with the TOPMed Imputation Server and allele-specific gene expression was determined using phASER Gene AE software.
Among self-identified Black/African American individuals, the median global African ancestry proportion was 0.74 (sd = 0.34). Sixteen individuals had low African global ancestry (less than 3%). Using RFMix2, we inferred proportions of African and European local ancestry at 9,910 chromosomal segments genome-wide. Using local ancestry, the median African ancestry was nearly identical to estimates for global African ancestry. After performing quality control using RNA-seQC, we detected 29,242 genes with at least 5 unambiguous reads. Mean mapping rate was 94.08% across all samples. 87.45% of mapped reads were of high quality.
These data constitute a valuable resource for assessing the impact of genetic variation on gene expression in lung tissue of ancestrally diverse populations. These data will inform future work to identify ancestry-specific eQTL that may influence racial disparities in lung cancer risk and outcomes.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6008T: A zebrafish crispant model of *crim1* loss of function shows similarity to mouse models

Authors:

T. Le[1], S. Htun[2], M. Pandey[1,3], Y. Sun[4,5], A. F. Magnusen[1], S. Baranzini[4,5], **A. Slavotinek**[1,3]; [1]Div. of Human Genetics, Dept. Pediatrics, Cincinnati Children's Hosp., Cincinnati, OH, [2]Div. of Med. Genetics, Dept. Pediatrics, UCSF, San Francisco, CA, [3]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [4]Dept. of Neurology, UCSF, San Francisco, CA, [5]Inst. of Human Genetics, UCSF, San Francisco, CA

Abstract:

Heterozygous deletions predicting haploinsufficiency for *CRIM1* have been identified in two families with autosomal dominant, colobomatous macrophthalmia with microcornea and coloboma (MACOM). Crim1 encodes a type I transmembrane protein that is expressed at the cell membrane of lens epithelial and fiber cells at the stage of lens pit formation. Murine lines with decreased *Crim1* expression demonstrate reduced numbers of lens epithelial cells and defective adhesion between lens epithelial cells and between lens epithelial and fiber cells. We used CRISPR/Cas9 to generate zebrafish that were homozygous (HZ) for a 2 basepair deletion in *crim1*, c.339_340delCT:p.Leu113Leu*fs**3 (ENSDART0000050534.6). Similar to murine models of *Crim1* loss of function, HZ *crim1*-crispant larvae showed microphthalmia, smaller lenses with increased nuclei and reduced adhesion of lens cell layers compared to controls, with increased apoptosis in the crispant eyes at 24 and 72 hours post fertilization (hpf) compared to controls. 'Bulk' RNA-Seq using dissected, whole eyes from *crim1* homozygous larvae and controls at 72 hpf showed upregulation of *C1qc* in *crim1*-crispant larvae compared to controls *(*p = 0.03), amongst other genes. This finding was consistent with the dysregulation for 4 complement genes (C1qa, C1qb, C1qc, and Cd74) seen in RNA-Seq datasets from the *Crim*1^{glcr11} mouse model at P60. However, the increased *C1qc* in *crim1*-crispants was not verified on qRT-PCR or in an Elisa assay comparing C1q in lysates from crim1-cripsant and wildtype eyes. We are continuing to use 'bulk' and single cell RNA-Seq to study the dysregulated genes in our *crim1* crispant model.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6009T: Accurate polygenic association of disease to individual cells in single-cell RNA seq data

Authors:

A. Turcan[1], M. Zhang[1], K. Hou[2]; [1]Carnegie Mellon Univ., Pittsburgh, PA, [2]Univ. of California, Los Angeles, Los Angeles, CA

Abstract:

Associating individual cells to disease by integrating single-cell RNA-seq with GWAS can deepen our understanding of disease etiology, beyond insights from cell type-level analyses. However, existing methods have limited power due to excessive technical noise and do not distinguish true associations from those due to pleiotropic gene expression. Here, we introduce single-cell Disease Relevance Score+ (scDRS+), a high-powered, pleiotropy-aware approach to accurately associate individual cells in scRNA-seq to disease based on polygenic enrichment.
scDRS+ evaluates whether an individual cell exhibits excess expression levels across a set of

disease-associated genes inferred from GWAS, leveraging single-cell imputation and gene-level pleiotropy weighting to enhance accuracy. First, scDRS+ estimates the pleiotropy level for each gene based on patterns of cell-type specific expression. Second, scDRS+ constructs a disease-specific cell neighborhood graph using individual cell-level disease association statistics derived from scDRS (Zhang et al. 2022 Nat Genet). Third, scDRS+ imputes gene expression by propagating information across the neighborhood graph, then computing association statistics by applying scDRS to the imputed data. In steps 2 and 3, genes exhibiting high pleiotropy are down-weighted. In realistic simulations, scDRS+ is well-calibrated and significantly more powerful than baseline methods, (e.g., avg. 11x more powerful than scDRS across settings) . We applied scDRS+ to GWAS summary statistics for 74 diseases/traits (avg. N=346K) in conjunction with 3 scRNA-seq atlases (N=1.1M cells). First, scDRS+ achieved significantly greater power than scDRS (e.g., 490K cell-disease associations for scDRS+ vs. 70k for scDRS in Tabula Muris Senis). Second, scDRS+ results aggregated at the cell-type level recapitulated known biology, but also produced novel, biologically plausible findings, associating dendritic cells to rheumatoid arthritis (RA) (FDR=.0056 for scDRS+ vs. .27 for scDRS) and CD24 neutrophils to major depressive disorder (FDR=.049 for scDRS+ vs. .57 for scDRS). Third, scDRS+ identified subpopulations of disease-associated cells not captured by cell type annotations, such as subsets of muscle stem cells (MuSC) associated with RA. Interestingly, scDRS+ determined that the MuSC-RA associations were strongly correlated with age (PCC=-0.73, P=0.0008), as well as with protocadherin-gamma family expression (PCC=0.75, P=0.001). Fourth, scDRS+ greatly reduced pleiotropic cell-disease associations not distinguished by other methods, filtering 82% of a curated set of putative pleiotropic associations.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6011T: An Integrative Single-Cell Allele-Specific Gene Expression and Open Chromatin Analysis Method ⭐

Authors:

S. Tao[1], S. Feng[1], H. Hu[1], Z. Xu[2], R. Duerr[3], W. Chen[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of Pittsburgh, Munhall, PA, [3]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

Abstract:

Background: Genome-Wide Association Studies (GWAS) have identified enormous genetic variants that influence human diseases and traits. However, most of these variants are in non-coding regions, complicating their functional interpretation. Understanding how these variants influence chromatin accessibility and gene expression is crucial for elucidating their roles in

disease risk. As a complementary approach to Quantitative trait locus (QTL) analysis, Allele-specific expression (ASE) and open chromatin (ASoC) analyses offer a powerful approach for studying cis-regulatory effects of genetic variants. However, methods for prioritizing ASE/ASoC findings are missing. Recent advances in single-cell multiomics sequencing provide new opportunities to explore how genetic variants jointly regulate chromatin accessibility and gene expression.**Method:** we developed AGENS for Allele-specific Gene Expression and open chromatiN analysiS. As a gene/peak-based approach, AGENS partitions SNPs within each gene/peak cis-region and assigns read counts to each heterozygous site. We utilize the beta-binomial regression model from the DAESC method to identify cell-type specific allele-specific imbalance sites. To enhance this, we extended the Sum of Single Effects (SuSIE) framework from linear regression to beta-binomial regression, enabling more precise fine-mapping of ASoC and ASE results. Furthermore, we developed a novel Bayesian colocalization method to identify genetic variants that jointly regulate chromatin accessibility and gene expression.**Results:** In our simulations, AGENS demonstrated high accuracy and power in both fine-mapping and colocalization tasks. More interestingly, we have generated tri-modal DOGMA-seq data (transcripts, epitopes, and chromatin accessibility of the same cell) from ~181,000 T cells across 27 human donors and paired CITE-seq and Nanopore long-read sequencing data from ~50,000 PBMCs. Using AGENS on these in-house valuable datasets, we integrated SNP, transcript, and chromatin accessibility data to dissect T cell activation regulatory networks and immune disease loci. Our results revealed the allele-specific stimulation-responsive chromatin and transcript landscape across diverse T cell subtypes and their differences between disease-related stimulations. In conclusion, AGENS offers a robust and integrative approach for analyzing allele-specific gene expression and chromatin accessibility, enhancing our ability to uncover the regulatory mechanisms of genetic variants in a cell-type-specific context.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6013T: Application of Deep Learning Algorithms reveals Copy Number Variations (CNVs) Outside of Chr21 that associate with Congenital Heart Defects in Down Syndrome Children

Authors:

S. Torkamandi[1], Y. Liu[1], J. Glessner[2], M. March[1], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia (CHOP), Philadelphia, PA, [2]Childrens Hosp. of Philadelphia/PENN, Philadelphia, PA

Abstract:

Understanding the genotype-phenotype correlation is crucial for advancing clinical management, particularly in medical conditions with significant clinical variability such as Down syndrome (DS). Copy number variations (CNVs) play an important role in contributing to the clinical heterogeneity of complex disorders. In DS, CNVs are particularly significant in the context of congenital heart disease (CHD), contributing to the variability in clinical presentations. This study aims were designed to investigate the genetic basis of CHD, especially subtypes of CHD in DS patients, focusing on the importance of CNVs. Whole genome sequencing (WGS) was performed on 481 DS cases and 555 family healthy controls. CNVs were detected and integrated using six tools: Lumpy, Delly, CNVnator, GATK, Manta, and cn.mops. Based on SURVIVOR, an integration algorithm we developed to interpret CNVs from multiple tools, we selected CNVs that were captured by at least three tools for further analysis. Patients were classified by the presence and subtypes of CHD, including Patent ductus arteriosus (PDA), Atrial/Ventricular septal defect (ASD)/(VSD), and Ostium Secundum Atrial Septal Defect (OSTASD). Multilayer perceptron (MLP) was applied to identify significant genes, with CNVs' feature importance validated through permutation testing. The analysis identified specific genes with high importance scores for each CHD subtype, providing new insights into their roles in heart development and CHD manifestation. Pathway enrichment analysis highlighted significant biological pathways involved in heart development, such as the MAPK, Notch, and TGF-beta signaling pathways, offering a comprehensive view of the molecular mechanisms underlying these phenotypes. This study underscores the critical role of CNVs in the clinical heterogeneity of CHD and highlights the importance of genotype-phenotype correlations in clinical management. Our findings enhance the understanding of clinical variability in heart development mechanism of DS patients, revealing diverse genetic factors contributing to CHD pathogenesis. These insights help to identify the role of CNVs in the expression of clinical heterogeneity and pave the way for personalized treatment strategies and improved patient outcomes through targeted interventions based on individual genetic profiles.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6014T: Application of PEXDETEX to predict poison exons in publicly available databases

Authors:

A. Hunter[1], K. Kelchner[2], T. N. Person[3], A. E. Hare[1]; [1]Commonwealth Univ. of Pennsylvania, Bloomsburg, PA, [2]Univ. of Rochester, Rochester, NY, [3]Pennsylvania State Univ., State College, PA

Abstract:

Whole exome sequencing (WES) studies have become the leading diagnostic tool in identifying pathogenic LOF variants within the coding regions, however, the role of nonsense mediated decay (NMD) facilitated by intronic variants has been largely unexplored. Genomic variants that disrupt the conserved splice donor motif can lead to NMD due to the inclusion of an intronic sequence that introduces a premature stop codon to the resulting transcript. Variants that create these 'poison exons' (PEX) have been identified in a number of case studies, a broad survey of the prevalence PEX on a genome-wide scale has yet to be conducted. We developed an automated pathogenicity prediction algorithm (PEX-DETEX) to identify splice donor (SD) variants that facilitate PEX (SD-PEX) from VCF datafiles. PEX-DETEX utilizes SpliceAI to identify SD variants that cause introns to be retained during splicing and identifies predicted truncated protein sequences from the resulting altered mRNA transcript. Using PEX-DETEX, we created a genome-wide map of predicted SD-PEX in gnomAD v3.2. Of the 153,771 SD variants in gnomAD v3.2, 13.6% are predicted to be SD-PEX that alter the canonical mRNA transcript, suggesting a baseline tolerance of PEX. We further demonstrate the enrichment of pathogenic SD-PEX in 548 genes that are known to be associated with developmental brain disorders (DBD) in cases obtained from ClinVar and the UK Biobank compared to control individuals in gnomAD (v3.2). Using a genotype-first approach, we analyzed a total of 1495 SD variants in DBD genes from the UK Biobank, 1894 SD variants from ClinVar, and 2239 SD variants from gnomAD v3.2. Overall, 11.9% of SD in DBD genes from gnomAD were predicted to be SD-PEX, while 79.6% of the UK Biobank SD and 62% of ClinVar SD were predicted to cause PEX. This study provides increased evidence of pathogenicity for SD-PEX variants. This is the first study of its kind to develop and utilize an automated algorithm to assess the clinical implications of PEX variants on a genome-wide scale. The inclusion of this algorithm in clinical variant curation pipelines can help to elucidate the clinical manifestation of intronic variants and can be extended to the analysis of other genetic causes of PEX.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6016T: Cell type-specific RNA-sequencing analysis identifies KCNJ6 as a novel gene associated with Alzheimer's disease in East Asian populations

Authors:

M. Cho[1], S. Chaudhuri[2], S. Liu[2], Y-N. Huang[2], T. Park[2], T. Rosewood[2], A. Saykin[2], H-H. Won[3], K. Nho[2]; [1]Indiana Univ. Sch. of Med., Sungkyunkwan Univ., Indianapolis, IN, [2]Indiana Univ. Sch. of Med., Indianapolis, IN, [3]Sungkyunkwan Univ., Seoul, Korea, Republic of

Abstract:

Genome-wide association studies (GWAS) help understand the underlying mechanisms of Alzheimer's disease (AD) with an estimated heritability of 60 to 80%. Although genetic variation varies across ancestries, most studies have predominantly focused on non-Hispanic White (NHW) populations. Given that GWAS in East Asian populations have identified several AD-associated loci, investigating these loci at the single-cell level might further elucidate novel genes. In this study, we performed cell type-specific RNA-sequencing analysis from the *Religious Orders Study and Memory and Aging Project* (*ROS/MAP*) cohort, being predominantly NHW, to identify novel genes associated with East Asian GWAS loci. We analyzed 26 loci previously identified in seven GWAS in East Asian populations, including cohorts from South Korea, China and Japan. Using cell type-specific expression quantitative trait loci (eQTL), we identified two genes that were not reported in the original studies. One of the newly identified genes, *KCNJ6* (potassium inwardly rectifying channel subfamily J member 6), is involved in the synaptic transmission and serotonergic pathway. A protein-protein interaction network revealed that *KCNJ6* was connected to *CACNA1A* (calcium voltage-gated channel subunit alpha1 A) that was identified in a previous East Asian GWAS. Additionally, *KCNJ6* was highly expressed in neurons and shown to be expressed according to AD progresses. Overall, cell type-specific eQTL analysis of genetic loci in East Asian populations identified a novel AD-associated gene, *KCNJ6*, which has not been identified in European GWAS, highlighting the importance of diverse ancestries in AD GWAS.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6017T: Cell-type-specific effects of chromosome 15q11-q13 duplications in human brain

Authors:

C. Dias[1], A. Mo[2], C. Cai[2], L. Sun[2], K. Cabral[2], C. Brownstein[2], S. Rockowitz[2], C. Walsh[2]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Boston Children's Hosp., Boston, MA

Abstract:

Recurrent copy number variation represents one of the most well-established genetic drivers in neurodevelopmental disorders, including autism spectrum disorder (ASD). Duplication of 15q11-q13 (dup15q) is a neurodevelopmental syndrome that increases the risk of ASD by over 40-fold. However, the effects of this duplication on gene expression and chromatin accessibility in specific cell types in the human brain remain unknown. To identify the cell-type-specific transcriptional and epigenetic effects of dup15q in the human frontal cortex we conducted

single-nucleus RNA-sequencing and multi-omic sequencing on dup15q cases (n=6) as well as non-dup15q ASD (n=7) and neurotypical controls (n=7). Cell-type-specific differential expression analysis identified significantly regulated genes, critical biological pathways, and differentially accessible genomic regions. Although there was overall increased gene expression across the duplicated genomic region, cellular identity represented an important factor mediating gene expression changes. Neuronal subtypes, showed greater upregulation of gene expression across a critical region within the duplication as compared to other cell types. Genes within the duplicated region that had high baseline expression in control individuals showed only modest changes in dup15q, regardless of cell type. Of note, dup15q and ASD had largely distinct signatures of chromatin accessibility, but shared the majority of transcriptional regulatory motifs, suggesting convergent biological pathways. However, the transcriptional binding factor motifs implicated in each condition implicated distinct biological mechanisms; neuronal JUN/FOS networks in ASD vs. an inflammatory transcriptional network in dup15q microglia. This work provides a cell-type-specific analysis of how dup15q changes gene expression and chromatin accessibility in the human brain and finds evidence of marked cell-type-specific effects of this genetic driver. These findings have implications for guiding therapeutic development in dup15q syndrome, as well as understanding the functional effects of CNVs more broadly in neurodevelopmental disorders.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6018T: Cell-type-specific mapping of enhancer-gene pairs from multimodal single cell data ⭐

Authors:

C. Su, D. Lee, J. Zhang; Emory Univ., Atlanta, GA

Abstract:

Mapping enhancer and its target genes in disease-related cell types has been fruitful for studying the functional mechanism of genetic variants from Genome-wide associated studies (GWAS). However, most existing analysis use bulk data, which fail to capture the cell-type-specificity of enhancers, or use cultured cell lines, which may not represent primary cells or diseased cells. The recently available multimodal single-cell data measure gene expression and chromatin accessibility across the same cells from primary human tissues and from subjects with disease, facilitating the computational inference of enhancer-gene pairs in a cell-type-specific and context-specific manner. However, such analyses are challenged by high sparsity, sequencing depth confounding, and the large number of candidate pairs. These lead to limited

statistical power, excessive false positives, and extreme computational costs by existing methods. Here, we propose scMultiMap, a statistical method for identifying enhancer-gene pairs in disease-related cell types with multimodal single-cell data. Using a joint latent-variable model, scMultiMap fully leverages the quantitative information in counts, leading to higher statistical power (39.7%-62.1%) in a reproducibility analysis with 10x multiome datasets, stronger enrichment (up to 1.6-fold) for PLAC-seq validated pairs on brain cell types, and more pairs (8.0%-25.3%) overlapped with orthogonal data types (Hi-C, HiChIP and eQTL) on blood cell types, as compared to other state-of-the-art methods. By adjusting for sequencing depths in the model, scMultiMap maintains appropriate type I error control while the empirical false positive rate doubled for Signac on permuted data. Moreover, scMultiMap adopts a moment-based estimation framework and provides analytical p-values, which is statistically efficient and computationally scalable. When run on 31.1k candidate pairs with 729 cells, scMultiMap cost 12.8 seconds, which is &lt 1% of the time consumed for Signac (1.6 hours) and SCENT (32.8 hours). We applied scMultiMap to PBMC data and identified cell-type-specific gene regulatory trios in immune cell types. We further applied scMultiMap to brain samples from healthy and Alzheimer's disease (AD) subjects and found significant enrichment for AD heritability in microglia (1.3 fold higher than SCENT). scMultiMap linked a fine-mapped non-coding GWAS variant rs10792831 to *PICALM* in healthy microglia, while the pair was not significant in AD microglia, revealing the potential microglia-specific regulatory mechanism for this SNP. scMultiMap also linked rs4075111 to *INPP5D* in microglia that similarly vanished in AD.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6019T: *CNTN2* is a candidate modifier in a PIGA-CDG pedigree with reduced penetrance

Authors:

H. Thorpe, B. Pedersen, J. Bonkowsky, A. Quinlan, C. Chow; Univ. of Utah, Salt Lake City, UT

Abstract:

Loss of function mutations in the X-linked *PIGA* gene lead to PIGA-CDG, an ultra-rare congenital disorder of glycosylation (CDG), typically presenting with seizures, hypotonia, and neurodevelopmental delay. We identified two brothers (probands) with PIGA-CDG, presenting with mild developmental delay, epilepsy, and autism. Both probands carry the novel, rare $PIGA^{S132C}$ variant, a predicted damaging variant not found in the gnomAD database. Confirming this diagnosis, both probands show a 50% decrease in GPI-anchor proteins on the cell surface. Strikingly, the maternal grandfather and a great uncle both also carry $PIGA^{S132C}$, but

neither presents with symptoms associated with PIGA-CDG. We hypothesized that there might be a modifier segregating in the family that contributes to this reduced penetrance. Using whole genome sequencing and pedigree analysis, we identified all the possible susceptibility variants found in the probands and not in carriers and all the possible protective variants found in the carriers and not in the probands. This list of potential candidates included heterozygous, damaging variants in three other genes also involved directly in GPI-anchor biosynthesis, *PIGS, PGAP5,* and *DPM1*, and a small number of genes involved in other glycosylation pathways or encoding GPI-anchored proteins.
To functionally test our predicted modifiers, we used a *Drosophila* eye-based model of PIGA-CDG. We created double knockdowns (KD) of *PIGA* and the candidate modifiers and compared the eye sizes of the double KD to the PIGA eye model and single KD of the candidate modifiers. Multiple candidate modifiers showed significant interactions with *PIGA* in the fly that mimic what we predict in the family. *CNTN2*, a GPI-anchored protein harboring a nonsense variant, was the only predicted protective gene from the pedigree analyses to rescue the *PIGA* phenotype in *Drosophila*. Further testing of *CNTN2* was also performed in *Drosophila* neurological models of PIGA-CDG displaying seizures and climbing defects. Similar to the results seen in the eye model, loss of *CNTN2* rescued both seizures and climbing defects in the PIGA-CDG neurological models. This *Drosophila* functional validation makes *CNTN2* a strong candidate modifier underlying the incomplete penetrance seen in this family. However, it is unlikely it is the sole contributor. Identifying and studying rare disease modifier genes in human pedigrees may lead to pathways and targets that may be developed into therapies.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6020T: Colocalization of a bipolar disorder type 1 GWAS signal with expression and splice QTLs suggests a causal role of increased expression and cryptic splicing of *CACNA1B* in disease risk

Authors:

C. Radens[1], S. Ullrich[2], L. Addis[2], C. Mordaunt[1], D. Seaton[2]; [1]GSK, Collegeville, PA, [2]GSK, Stevenage, United Kingdom

Abstract:

Numerous GWAS have identified loci associated with complex diseases but pinpointing specific variants and the genes they act through to increase disease risk remains a significant challenge. One hypothesis is that causal variants associated with disease risk disrupt gene regulatory mechanisms, functioning as either expression (eQTL) or splicing (sQTL) quantitative trait loci.

We performed a Bayesian colocalization analyses to investigate whether a shared variant is causal for both a bipolar disorder type 1 (BD1) GWAS (Mullins et al 2021) signal and GTEx cerebellum eQTL and sQTL signals within the same locus. We observed that of the 19 genes near the BD1 locus of interest on chromosome 9, only *CACNA1B* showed eQTLs and sQTLs with a high probability (>80%) of colocalization. Our eQTL results suggest that increased overall expression levels of *CACNA1B* mRNA in cerebellum correlates with increased BD1 risk. Our sQTL results implicate increased splicing to a cryptic splice site acceptor (SSA) within the third-to-last intron (intron 44). The cryptic SSA is 545 bp upstream of the lead sQTL SNP, 9:138119688 G>C, and 1022 bp upstream of the annotated SSA. Splicing to the cryptic SSA in intron 44 encodes a putative 125 amino acid (AA) sequence before reaching a stop codon, truncating the last 329 AA of the canonical isoform. Conserved reading frames tend to be biologically important, so we ran a BLASTP sequence alignment on the 125 AA to test for conservation across species. We observed 100% identity with an annotated isoform of the *CACNA1B* homolog in chimpanzee, suggesting the "cryptic" splice isoform may be biologically relevant.
Additionally, *CACNA1B* encodes a subunit of the N type voltage-dependent calcium channel, which has been implicated in several neurological diseases. In summary, our AA alignment results motivate a closer look at "cryptic" splicing events to explore whether they are conserved with recent primate ancestors. Our colocalization findings suggest that the C allele at 9:138119688 causes increased expression and cryptic splicing of *CACNA1B* in the cerebellum, leading to an increased risk of BD1.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6021T: Comparison of BabySeq's novel gene selection approach to other newborn genomic sequencing programs' gene lists

Authors:

A. Nagy[1], T. Minten[2], S. Bick[3], S. Adelson[4], I. Holm[5], R. Green[6], M. Lebo[7], N. Gold[8], H. Zouk[9]; [1]Mass Gen. Brigham, Boston, MA, [2]KU Leuven, Leuven, Belgium, [3]Boston Childrens Hosp., Boston, MA, [4]Brigham and Women's Hosp., Boston, MA, [5]Boston Children s Hosp, Boston, MA, [6]Harvard Med. Sch., Boston, MA, [7]Mass Gen. Brigham Lab for Molecular Med., Cambridge, MA, [8]Massachusetts Gen. Hosp., Brookline, MA, [9]Lab for Molecular Med./ Harvard Med. Sch., Somerville, MA

Abstract:

Genomic sequencing of apparently healthy newborns and children (NBSeq) is being explored by many research projects and clinical initiatives and has the potential to identify genetic risks that

are not covered by traditional newborn screening. Typically, these NBseq projects begin with a well-defined gene list, often associated with severe, treatable conditions. However, between these NBSeq efforts around the world, gene list size varies widely, ranging between 134 and 613 genes (Minten 2024 PMID: 38585998). The BabySeq Project is a NBSeq multi-site randomized control trial in infants to identify monogenic disease risks (MDR) and recessive carrier status for conditions meeting reportability criteria. In comparison to other NBSeq initiatives with well-defined gene lists, BabySeq begins with a wider scope of 4299 genes that have some reported association to disease, which are then curated in depth for reportability upon detection in a case. This method allows for the inclusion of a broader set of genes, especially those with a newer association with disease.

To date, the second iteration of the BabySeq study has issued reports for 74 infants. Results returned include 140 unique variants in 125 genes. Fifteen of these individuals had monogenic disease risk (MDR) findings in 10 genes which represent a childhood-onset disease risk. The 10 genes with MDRs were compared to 25 other international NBSeq gene panels to determine which (if any) included these genes (Minten 2024 PMID: 38585998). Four of these individuals (27%) had MDRs which would not have been identified by any other NBSeq program. An additional 7 individuals (47%) had MDRs which would have only been found by 1-6 panels. Therefore, the majority of these individuals (11/15, 73%) would not have had these MDRs identified on the majority of current panels. The 4 MDR findings that were identified exclusively by the broader BabySeq gene selection method were associated with autosomal dominant, childhood-onset conditions. These include *ARFGEF1* (neurodevelopmental disorder), *KMT2E* (syndromic neurodevelopmental disorder) *TEK* (Congenital glaucoma), and *ZMYM2* (neurodevelopmental-craniofacial syndrome).

BabySeq has a unique approach to NBSeq that allows the inclusion of all genes which may be clinically relevant in childhood. The majority of MDRs identified with this approach would have been missed in more traditional settings where a defined and pre-curated gene-list approach is utilized. In the future, consideration of using a similar gene approach for NBSeq may increase the sensitivity and yield for MDRs, however the utility and feasibility of this approach on a large scale will need further investigation.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6022T: Comprehensively comparison and application of ancestry-specific eQTL from an admixed population.

Authors:

H-H. Chen[1], W. Zhu[2], M. Graff[3], M. Yaser[3], L. Petty[4], H. Polikowsky[2], J. McCormick[5], S. Fisher-Hoch[5], H. Highland[3], E. Gamazon[4], K. North[3], J. Below[4]; [1]Academia Sinica, Taipei, Taiwan, [2]Vanderbilt Univ., Nashville, TN, [3]Univ. of North Carolina, Chapel Hill, NC, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX

Abstract:

Expression quantitative trait loci (eQTL) are genetic variants that regulate the expression of nearby genes. Top findings from GWAS of complex diseases are often enriched with eQTLs and colocalizing significant GWAS results with eQTLs can inform causal gene mapping and functional interpretation. Despite the importance of eQTLs, the majority of eQTL resources are derived from samples predominantly of European descent, potentially limiting their power in other understudied populations, such as Hispanics/Latinos. In 645 Hispanic/Latino participants from the Cameron County Hispanic Cohort (CCHC) with genomic and transcriptomic data, we combined local ancestral information and allelic-specific expression to identify over 3 million local ancestry-informed eQTL for 8,523 genes (FDR-adjusted p-value < 0.05) in European- and Amerindian-specific haplotypic backgrounds. In comparison to extant European eQTL databases (GTEx whole blood N=670), we observed a positive trend between the heritability of gene expression and the genetic correlation between GTEx and CCHC European specific expression regulatory. For highly heritable genes (h^2>0.4) and the median was 0.53 for all heritable genes (h^2>0.05). In stark contrast, the correlation between European- and Amerindian-specific regulatory in CCHC were not correlated with heritability, with a median of 0.5 and 0.53 for highly heritable and heritable genes, respectively. This result suggests there is low correlation of regulatory effects across different local genetic backgrounds, regardless of how heritable the gene expression is. Next, we conducted colocalization analysis of local ancestry-specific eQTLs with Hispanic/Latino GWAS, including type 2 diabetes, blood lipids, body height and body mass index. The CCHC-derived local ancestry-informed eQTL map demonstrated a superior performance over the GTEx-derived map on interpreting Hispanic/Latino GWAS. Our ancestral specific results suggest 10, 19 and 3 colocalized regions for T2D, body height, and BMI, but GTEx dataset only indicates 1, 10, and 0 regions, respectively. For blood lipids, CCHC eQTL identified a total of 58 colocalized regions and GTEx only includes 19 regions. Moreover, we also find our local ancestry-specific eQTL map outperformed the map from standard bulk approach, e.g. 33 regions colocalized from CCHC bulk eQTL map for blood lipids GWAS compared to 58 with local ancestry-specific QTLs. Our findings show that genomic context is important model systems cannot rely on the introduction of a single variant to recapitulate the effects seen in human bodies, and emphasize the imperative call for ancestral diversity in biomedical research.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6023T: Context-aware single-cell multiome approach identified cell-type specific lung cancer susceptibility genes ⭐

Authors:

E. Long[1], J. Yin[2], S. Ju Hye[3], Y. Li[1], B. Li[2], A. Kane[2], H. Patel[2], X. Sun[1], C. Wang[1], T. Luong[2], J. Xia[4], Y. Han[5], J. Byun[5], T. Zhang[2], W. Zhao[2], M. Landi[2], N. Rothman[2], Q. Lan[2], Y. Chang[3], F. Yu[6], C. Amos[5], J. Shi[2], J. Lee[3], E. Kim[3], J. Choi[2]; [1]Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, [2]NCI/NIH, Bethesda, MD, [3]Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [4]Creighton Univ., Omaha, NE, [5]Baylor Coll. of Med., Houston, TX, [6]Guangzhou Natl. Lab., Guangzhou, China

Abstract:

Genome-wide association studies (GWAS) identified over fifty loci associated with lung cancer risk. However, the genetic mechanisms and target genes underlying these loci are largely unknown, as most risk-associated-variants might regulate gene expression in a context-specific manner. Here, we generated a barcode-shared transcriptome and chromatin accessibility map of 117,911 human lung cells from age/sex-matched ever- and never-smokers (n = 16) to profile context-specific gene regulation. To better represent lung cancer cells of origin we used fluorescence-activated cell sorting, which enriched epithelial cell populations to 40% of the final dataset. A joint-clustering of RNA and assay for transposase-accessible chromatin (ATAC) modalities identified 23 cell types in four categories supported by marker gene expression and chromatin accessibility. Candidate cis-regulatory elements (cCREs) defined by ATAC-seq peaks were largely cell type specific, where 37% were detected in a single cell type and located outside of gene promoter regions. Differential expression and chromatin accessibility analyses identified candidate genes affected by smoking status in each cell type, including previously known smoking-responsive genes. Cell-cell communication analysis using CellChatDB highlighted potentially altered communications in MHC pathways based on smoking status. Colocalization of the cCREs with 2,574 lung cancer candidate causal variants (CCVs) from 51 GWAS loci prioritized the variants for 68% of the loci. Among them, 111 CCVs from 29 loci were further supported by transcription factor abundance and footprinting. cCRE colocalization and single-cell based trait relevance score using SCAVENGE nominated epithelial and immune cells as the main cell categories contributing to lung cancer susceptibility. Notably, cCREs of rare proliferating epithelial cell types, such as AT2-proliferating (0.13%) and basal cells (1.8%), overlapped with the CCVs. A multi-level cCRE-gene linking system identified candidate susceptibility genes from 57% of lung cancer loci, including those not detected in tissue- or cell-line-based approaches. cCRE-gene linkage highlighted the context-dependent gene regulation across the loci, where half of the loci showed cell category-specific target gene connection,

while the other half involving the cCREs active in multiple cell type categories via distinct subsets of coinherited CCVs. Overall, we provide a unique single-cell resource to characterize lung cancer GWAS loci and highlight context-dependent gene regulation underlying the risk.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6024T: CRISPR screens to unravel the role of candidate genes associated with blood pressure by GWAS on endothelial cell functions

Authors:

E. Lecluze[1], B. Melissa[1], H. Belhaj[2], G. Lettre[3]; [1]Montreal Heart Inst., Montreal, QC, Canada, [2]ICM, Montréal, QC, Canada, [3]Université de Montréal, Montreal, QC, Canada

Abstract:

Hypertension is a major public health issue resulting from the interplay between genetic and environmental factors. Although genome-wide association studies (GWAS) have identified 1000s genetic variants associated with blood pressure (BP), most are non-coding and do not clarify which genes are affected or their precise biological roles. This genetic variation may influence endothelial cells, which are crucial in regulating BP, but the specific effects remain to be elucidated. This functional genomics project aims to assess the impact of candidate genes disruption on endothelial cell activity. First, we prioritized 2454 candidate genes around BP GWAS signals. Second, we designed at least four gRNAs to target these genes with the CRISPR-Cas9 (nuclease) and CRISPR-dCas9^{KRAB} (CRISPRi) systems in immortalized human endothelial cells (teloHAEC) using lentiviruses. These approaches enabled us to edit our targets or inhibit their expression, respectively, and to observe their effects on the proliferation of teloHAEC, a cellular readout used to capture the roles of these genes in vascular endothelial cells. Our CRISPR screens identified 26 genes significantly impacting cell proliferation, including the ERG gene, an important transcription factor in our cellular model, along with other new candidates (e.g. CDKN1A). These genes demonstrated notable susceptibility to disruption by BP-related variants in their regulatory regions, particularly in endothelial and various immune cells, as shown through analyses using the ENCODE-rE2G encyclopedia of enhancer-gene interactions. An enrichment analysis using the WebCSEA resource indicated that our significant genes are predominantly expressed in endothelial cells across various tissues. By integrating transcriptomic and epigenomic data in both *in vitro* and *ex vivo* experiments, we prioritized genes most relevant to endothelial cells. We applied CRISPR/Cas9-mediated base editing to install specific alleles in teloHAEC at variants near ERG and CDKN1A to pinpoint the causal BP

variants at these two loci, screening for proliferation and response to inflammatory stimuli.Together, our results highlight potential new regulators of BP that act through vascular endothelial functions, potentially paving the way for targeted therapies and precision medicine.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6025T: Cryptic donor splice-site in non-coding region of *HADHA* leads to diagnosis of Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (LCHAD)

Authors:

L. Rives, J. Phillips, A. Robertson, T. Cassini, K. Ezell, J. Cogan; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (LCHAD) is a rare, autosomal recessive condition that prevents the body from converting certain fats to energy, particularly during periods of fasting. Symptoms usually appear by early childhood and can include cardiomyopathy, cardiac arrhythmias, hypoglycemia, hypotonia, retinopathy, and rhabdomyolysis. Individuals with LCHAD deficiency are also at risk for serious heart problems, breathing difficulties, coma, and sudden death. This disease is typically caused by pathogenic variants in *HADHA* gene. Diagnosis is commonly made through genetic testing in combination with acylcarnitine analysis. A 26-year-old male was referred to the Undiagnosed Diseases Network (UDN) with a history of myopathy, muscle weakness, and pain beginning in childhood. With age, the family noted a clear association of symptoms with physical exertion, dehydration and fasting. The subject was diagnosed with rhabdomyolysis at 14 years of age. Electromyography (EMG) showed myopathic abnormalities, muscle biopsy was positive for mild fiber size variation, and biochemical testing only found intermittent mildly elevated long-chain acylcarnitine (C12-C18) species. Genetic panel testing, done prior to UDN submission, found a maternally inherited pathogenic frameshift rare variant (RV) in *HADHA*:NM_000182.4:c.274_278del-TCATC (p.S92Kfs*10), but no variant on the alternate allele was identified. We performed genome sequencing (GS) and identified a paternally inherited non-coding RV in *HADHA*:NM_000182.4:c.68-414A>G. This RV was predicted to create a cryptic splice donor site and perturb splicing. Analysis of cDNA on the proband and parents confirmed that the maternal frameshift allele was subject to nonsense mediated decay (NMD). The paternally inherited intronic allele was found to encode both normal and abnormal transcripts utilizing the cryptic donor site, with the abnormally spliced product being subject to

NMD. The presence of a measurable proportion of normal *HADHA* transcript from the paternal allele was thought to cause the proband s milder presentation of LCHAD. LCHAD is usually a life-threatening, but treatable condition. This subject presented with clinical and lab findings that were atypical, making it elusive to typical biochemical and molecular methods of detection, which prolonged the diagnostic odyssey. The findings of this case demonstrate the utility of GS, coupled with transcript analysis, to demonstrate the significance of non-coding alleles that can enable diagnosis and treatment, as seen in this subject, who responded well to the known available treatment of triheptanoin.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6026T: De novo deletion variant in the GTPase effector domain (GED) of DNM1L leads to Gross motor delay, hypotonia, axonal sensory neuropathy with peroxisomal and mitochondrial abnormalities

Authors:

S. Srivastav[1], N. Roy[2], J. Andrews[2], S. V. Jangam[2], M. Wangler[3]; [1]NRI, Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX

Abstract:

DNM1L encodes dynamin-related protein 1 (DRP1), an 80-kDa GTPase in the dynamin super-family, which is essential for mitochondrial and peroxisomal fission. Dysfunction in DNM1L has been linked to mitochondrial and neurodegenerative disorders, including Alzheimer's, Parkinson's, and Huntington's disease. DNM1L protein has 3 distinct domains the GTPase domain, middle domain and GTPase effector domain (GED). To date, different domain specific variants have been identified in the DNM1L gene among patients. These gene variants exhibit a range of neurological symptoms, from severe hypotonia causing neonatal death to developmental delay, cerebellar atrophy, encephalopathy, profound intellectual disabilities, and gross motor delay. Pathogenic variants in the middle domain and GTPase domain of DNM1L are associated with encephalopathy due to defective mitochondrial and peroxisomal fission 1 (EMPF1, MIM #614388). Additionally, familial pathogenic variants in the GTPase domain have been connected to isolated optic atrophy. Here we report a patient with a novel, de novo deletion variant (Leu700del) in the GED domain exhibiting gross motor delay, onset at infancy with hypotonia, and axonal sensory neuropathy, thereby further expanding the clinical spectrum of DNM1L associated disorders. Here we detail the functional studies involving this novel GED variant (Leu700del) in Drosophila and compare the phenotypes this variant produces

in our model system with GTPase domain (A192E), and middle domain (R403C) fly models produced in our lab. We find that our Drosophila model displays large abnormal peroxisomes and mitochondria when overexpressing the GED variant (Leu700del) and middle domain variant (R403C), an effect comparable to that of middle domain missense alleles observed in pediatric subjects with EMPF1. The results clearly indicate that our Drosophila model can produce EMPF1 phenotype, and that current study further enhances our understanding of the natural progression of an increasing number of DNM1L-related disorders.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6027T: Detection of a pathogenic *PRKN* inversion by a novel DNA-RNA combined analysis of structural variants

Authors:

E. Schultz[1], G. Sharma[1], H. Asghari[1], S. Baker[1], D. Arkadir[2], Y. Mazor[3], J. Greenblatt[3], Y. Unger[3], H. Volkov[3], T. Harel[2], R. Patro[1], G. Marcais[1], C. Kingsford[1], H. Mor-shaked[4], **W. Haagmans**[5]; [1]Ocean Genomics, Pittsburgh, PA, [2]Hadassah-Hebrew Univ. Med. Ctr., Jerusalem, Israel, [3]Geneyx Genomex LTD, Herzliya, Israel, [4]Hebrew Univ. Hadassah Med. Sch., Jerusalem, Israel, [5]Geneyx, Arcen, Netherlands

Abstract:

Background: The contribution of structural variants (SVs) to an individual's genetic makeup is not well understood. Only in recent years, with the advent of high-throughput sequencing technologies, has the extent of structural variation in both phenotypically normal individuals and those with pathological conditions been explored. Recent studies have shown that SVs account for 5-15% of genomic-based diagnoses. However, many SV calling methods in diagnostics pipelines are focused on DNA without considering the functional impact at the transcriptomic level and often yield a high number of false positive events. Here, we used a novel method - SVDuo - that intelligently integrates SV information from DNA- and RNA-seq to provide functional information about the structural variants with higher confidence. This synergistic approach provides better understanding of the underlying biology than either modality alone. Applying SVDuo to DNA- and RNA-seq data obtained from an early-onset Parkinson's patient revealed the previously identified inversion in the *PRKN* gene along with newly discovered functional effects of the SV at the transcriptomic level.
Method: SVDuo employs a multi-step pipeline. First, high-quality reads are pre-processed and aligned, with detailed quality statistics incorporated into the final output. Subsequently, dedicated DNA and RNA modules within SVDuo call SVs, applying stringent filters to minimize

false positives. Finally, SVDuo uses a novel position-based clustering approach to merge different types of SVs generating a VCF with a unified set that indicates detection in DNA, RNA, or both.
Result: Using SVDuo, a novel DNA-RNA combined SV caller, an inversion within the *PRKN* gene was detected along with novel transcripts. The inversion breakpoints are present in the intronic region surrounding exon 5. As a result, the *PRKN* gene, whose variants are known to lead to early onset Parkinson's disease, is missing expression of exon 5. Use of SVDuo highlights the importance of a joint DNA-RNA approach in investigating rare diseases as this reveals key structural variants and their functional impact with high confidence.
Conclusion: Using DNA or RNA alone presents an incomplete picture of the importance and impact of SVs. SVDuo solves this issue by performing integrated DNA-RNA analysis, providing functional annotation to structural variants with increased confidence in detection due to evidence from multiple modalities. While further studies are needed, this multi-data approach holds promise for advancing our understanding of the role SVs play in human health.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6028T: Discovering the molecular basis of *TBR1*-associated autism

Authors:

H. C. Rudolph[1], G. K. O'Brien[1], M. Co[1], J. Michaelson[2], Y. Shen[3], B. J. O'Roak[1]; [1]Oregon Hlth.& Sci. Univ., Portland, OR, [2]Univ. of Iowa, Iowa City, IA, [3]Columbia Univ., New York, NY

Abstract:

Missense variants are the most abundant type of protein-altering variants. Recent studies have shown that missense variants overall contribute to about half of attributable risk of autism among coding *de novo* variants (DNVs). However, the genetic effects of individual missense variants are largely uncertain. Additionally, these effects are more complex than binary pathogenic or benign classifications. Accurate interpretation of gain- or loss-of-function variants in these genes is of critical importance for genetic diagnosis in autism and future interventions. We have developed a new multimodal "sort-seq style" deep mutational scanning (DMS) assay for transcription factors (TFs), which can capture protein abundance, DNA binding activity, and subcellular localization in a single experiment. For piloting, we focused on *TBR1*, which is a neuronal TF and potential master regulator of many autism risk genes. *TBR1* DNVs lead to a complex clinical presentation that includes autism, intellectual disability, behavioral disturbances, and speech and motor impairments. The distribution of *TBR1* DNVs predicts at

least three major classes of pathogenic variants: 1) Early-truncating loss-of-function; 2) Missense that are highly clustered in the T-Box DNA-binding domain; and 3) Last exon late-truncating, which have more severe clinical presentations (e.g., cortical malformations, non-verbal/mobile). Using *in vitro* plasmid overexpression assays, we and others have shown that a limited set of *TBR1* DNVs are dysfunctional through different molecular mechanisms. *Importantly,* the majority of T-Box missense and late-truncating DNVs tested (e.g., p.K228E, p.T532Rfs*144) show increased protein stability. Some of these DNVs also disrupt TBR1 protein interactions. For the DMS approach, we are leveraging an improved lentivirus landing pad (LLP) HEK293T system, LLP-iCasp9-Blast (Matreyek et al., 2020). With our custom construct, we tested previously evaluated DNVs and recapitulated findings from cycloheximide assays, including relative rankings of protein abundance. All pilot T-Box missense DNVs increased in abundance, while early truncations lost nuclear specificity. Emboldened, we are scaling our DMS to include all possible (~3.6k) variants of the T-Box domain. We will use these multimodal scores (e.g., abundance x TF activity) to generate a combined functional matrix for each variant. The matrix distributions will be used to highlight likely molecular mechanisms for known patient DNVs, e.g., increased stability with abnormal TF activity. We also plan to conduct an exploratory analysis to see if these functional groupings predict autism clinical severity.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6029T: Discovery of noncoding mobile element insertions associated with aberrant splicing signatures ⭐

Authors:

A. Yenkin[1,2,3], X. Zhao[2,3,1], Y. Mostovoy[2,3], R. Yadav[1,2,3], M. Walker[2,3], E. Pierce-Hoffman[2,3], Genome Aggregation Database (gnomAD) Production Team, Genome Aggregation Database (gnomAD) Structural Variant Team, H. Brand[1,2,3], D. Gao[1,2,3], M. Talkowski[1,2,3]; [1]Harvard Med. Sch., Boston, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

Mobile element insertions (MEIs) are a diverse class of human genetic variation. They are caused by novel insertion of the three main human retrotransposons, Alu, LINE-1, and SVA elements. In many documented cases, MEIs have led to human disease through altering splicing patterns by providing novel splice acceptor/donor sites and other mechanisms. Studies examining these effects have generally been case-specific, and there is a dearth of systematic assessments of the functional impact of MEIs that result from natural genetic variation in the

population. MEIs have also largely been omitted from prior research examining variant-transcriptome associations because they are difficult to discover from short-read genome sequencing (srGS) as they include highly repetitive sequences. One challenging class of MEIs that are particularly underexplored, SVA elements, are found only in humans and great apes and occur far less frequently in the genome than LINE-1 and Alu. Here, we applied GATK-SV to profile SVs, including MEIs, from paired whole-genome sequencing (WGS) and RNA-seq in 851 individuals from the Genotype-Tissue Expression project.
We quantified exon usage and intron retention (IR) for each alternative splicing event across 18,647 genes encompassing 246,693 introns. Unique to this study, we quantified IR across the span of the intron rather than on the splice junction, which increases power to discover functional MEIs. Upon modeling these splicing events via beta regression, we then identified exon usage and IR outliers due to local MEIs (19,948 Alu, 6,631 LINE-1, 2,087 SVA). We find that introns containing Alu and SVA elements were significantly enriched for aberrant IR (Alu: OR=2.21, p=1e-60; SVA: OR=4.01, p=2e-42), generalizing the consequences of MEIs from the disease literature to the broader genome. Strikingly, introns containing SVA elements were the most enriched for aberrant IR demonstrating SVA elements as particularly potent splicing modulators. We characterized four cases of intronic SVA elements that were associated with extreme IR signatures across tissues. The genes containing these elements display novel splice junctions close to the SVA elements and discrete localization of the IR with respect to the SVA, both signatures present in previously described pathogenic SVA. We thus demonstrate that MEIs are important components of transcriptional modulation in gene regulation, and that the specific sequence characteristics and features of functional consequences associated with SVA disruption may be more widespread in disease association than previously appreciated.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6030T: Distinguishing environmental variance from measurement noise in plasma protein levels

Authors:

X. Jiang[1], S. Ritchie[2], M. Inouye[1], A. Price[3]; [1]Cambridge Univ., Cambridge, United Kingdom, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Harvard Sch Pub Hlth, Boston, MA

Abstract:

Despite intense interest in population-scale studies of protein levels (Sun et al. 2023 *Nature*), sources of variance in protein levels are poorly understood. Existing methods can quantify genetic vs. non-genetic variance, but cannot distinguish biological environmental variance from

non-biological measurement noise, which cross-platform studies show to be substantial (Eldjarn et al. 2023 *Nature*).
We propose a method, expression-wide regression (EWR), to dissect protein level variance into **G** (genetic variance), **E**(biological environmental variance), and **N** (non-biological measurement noise). EWR estimates the **G** + **E** variance of a focal protein *A* using a correlated protein *B* and a set of auxiliary traits Z_j (e.g. other proteins); EWR relies on the intuition that the components **G** + **E**, but not **N**, are partially tagged by other proteins. First, we regress covariances cov(A, Z_j) on cov(B, Z_j) across Z_j. Second, we adjust the regression slope using an error-in-variable model to obtain an estimate of **G** + **E** variance for focal protein *A* (which can be decomposed into **G** and **E** by applying standard methods to estimate **G**). Simulations show that EWR obtains unbiased estimates of the **G** + **E** variance that is tagged by auxiliary traits Z_j; for a large set of auxiliary traits, this approximates the total **G** + **E** variance.
We applied EWR to UK Biobank plasma protein levels for 238 proteins with low correlation between longitudinal measurements in the same individuals (which could be caused by large **E** or large **N**); we also estimated **G** using BOLT-REML. We reached 3 main conclusions. First, EWR obtains consistent estimates of **G + E** for a given *A* across the 10 most correlated proteins *B* (median variance ratio 30% vs. estimates of **G + E** across *A* and *B*).
Second, **E** estimates varied substantially across proteins, averaging 92% for the top 10 proteins (s.e.(avg)=0.9%) vs. 23% for the bottom 10 proteins (s.e.(avg)=0.5%) (overall s.d.=17% vs. avg s.e. of 1.9%). For example, we estimated **E**=80% (s.e. 2.4%) vs. **N**=10% (s.e. 2.2%) for METAP2, which impacts tissue repair and protein degradation; measurements of proteins with large **E** provide substantial complementary information compared to genetically predicted values. Third, many proteins had substantial values of heritability after removing noise variance, defined as **G**/(**G**+**E**): avg of 45% for the top 10 proteins, vs. avg heritability of 20% for the same 10 proteins. For example, we estimated **G**/(**G**+**E**)=37% vs. heritability of 22% for HGF, which impacts adult organ regeneration; proteins with large **G**/(**G**+**E**) are good targets for genetic prediction (Xu et al. 2023 *Nature*). In conclusion, EWR distinguishes environmental variance from measurement noise in plasma protein levels.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6031T: Effect of the rs1050828 G6PD variant on HbA1c-based risk assessment for diabetes and its complications in a model incorporating polygenic scores

Authors:

N. Arora, F. Blostein, P. Straub, M. Shuey, N. Cox, J. Hirbo; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Clinicians routinely utilize glycated hemoglobin (HbA1c) as a key measure for diabetes diagnosis and prognosis. However, HbA1c may be falsely lowered in individuals with the G202A, p.Val68Met (rs1050828) variant causing glucose-6-phosphate dehydrogenase (G6PD) deficiency, which disproportionately affects individuals of African ancestry, leading to an extended undiagnosed period of diabetes while still being hyperglycemic. We aimed to determine the risk of complications in patients with diabetes and the rs1050828 variant compared to patients with diabetes without this variant. We also sought to analyze the effect of incorporating polygenic risk coupled with the G6PD variant on a prediction model of HbA1c. For complication risk, a cohort of individuals with a coded Type 1 (T1D) or Type 2 (T2D) diabetes diagnosis with or without G6PD deficiency and specific comorbidities was compiled from BioVU (2007-2020). We utilized poisson regression modeling to determine the difference in risk of diabetic complications between individuals of African ancestry with (n=1038) and without (n=4867) rs1050828. We also generated polygenic risk scores in our BioVU cohort using large multi-ancestry HbA1c summary data (n=760,612) and checked improvement in prediction quality for HbA1C with or without the rs1050828 variant included. We estimated a 21.6% increased risk of developing diabetic complications in patients with T1D or T2D of African ancestry with rs1050828 compared to those without the variant (p=2.75E-06). In addition, there was a 60.5% increased risk of developing ketoacidosis (p=0.00218) and 76.2% increased risk of developing ophthalmic manifestations (p=0.000827) for those with the variant and T1D. Replication was performed in an All of Us cohort of individuals of African ancestry with diabetes (n=26457) +/- rs1050828. Age was found to be a risk factor for developing all diabetic complications, male gender for developing ketoacidosis (p=.003) and renal complications (p=4.7e-12), and PC1-5 and 7 for developing ophthalmic manifestations. Lastly, there was a 13.6% increase in prediction quality of HbA1c in individuals without diabetes when rs1050828 was incorporated into the model. Because the rs1050828 variant causing G6PD deficiency falsely lowers HbA1c and increases certain diabetic complication risk, we showcase the need to routinely test individuals of African ancestry for G6PD variants and utilize multiple measures of glycemic control in addition to HbA1c for diabetes diagnosis and prognosis. We additionally showcase the increased precision of incorporating a G6PD variant into a prediction model for diagnosis of diabetes using HbA1c.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6032T: Efficient Demultiplexing and Dynamic Allele-Specific Expression Quantification for Single-Cell RNA-seq and ATAC-seq Data

Authors:

R. Pique-Regi, j. wei, F. Luca; Wayne State Univ., Detroit, MI

Abstract:

Single-cell RNA sequencing (scRNA-seq) and single-cell ATAC sequencing (scATAC-seq) technologies provide high-resolution insights into cellular heterogeneity and gene regulation. We present a suite of tools designed to enhance the analysis of these data by enabling efficient demultiplexing of reads by genotype and quantification of allele-specific expression (ASE) at heterozygote loci.Our demultiplexing tool achieves accuracy comparable to the widely-used demuxlet but offers significant improvements in computational efficiency. It runs faster and requires substantially less memory, making it suitable for large-scale studies. The tool seamlessly handles cell barcodes and unique molecular identifiers (UMIs), ensuring accurate assignment of cells to the genotyped individuals, as well as detecting droplets with multiple cells from different individuals. To quantify ASE, we have extended our previous QuASAR framework. This model employs a beta-binomial distribution with a logistic link function, allowing for the control of covariates and batch effects. It facilitates the monitoring of dynamic changes in ASE across various conditions, including different cell types, treatment groups, and continuous variables such as pseudo-time or single-cell signatures. This model is particularly adept at detecting dynamic genotype-by-environment (GxE) interactions at single-cell resolution for both expression and ATAC-seq data, as well as multiome assays.Our advancements provide researchers with powerful tools to dissect the genetic and regulatory complexity of single-cell datasets, offering new insights into cellular responses and interactions. The ability to accurately demultiplex reads and quantify dynamic ASE will enhance our understanding of cellular behavior and gene regulation in diverse biological contexts.These tools represent a significant step forward in single-cell genomics, enabling more precise and efficient analysis of scRNA-seq and scATAC-seq data. By providing a robust and computationally efficient framework for ASE quantification, we aim to facilitate the discovery of novel regulatory mechanisms and the characterization of cellular heterogeneity with unprecedented detail.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6033T: Endocrine and reproductive phenotypes in the *Sox2* heterozygous knockout mouse

Authors:

J. Cassin, R. D. Nguyen, K. J. Tonsfeldt, G. A. Dunn, V. Chen, Y. E. Seung, P. L. Mellon; UCSD, San Diego, CA

Abstract:

Mutations in the *SOX2* gene have been previously linked to a syndromic form of Isolated Hypogonadotropic Hypogonadism (IHH), with additional ocular and neurodevelopmental phenotypes. IHH is a rare genetic disorder of the endocrine system with a substantial burden of unidentified genes and mutations. Recently, we published research detailing novel human variants in *SOX2*, and demonstrated a functional role for SOX2 in Kisspeptin-expressing hypothalamic neurons, establishing a mechanistic relationship between SOX2 heterozygous variants and IHH. To further test the role of *SOX2* in the hypothalamus, we utilized *Sox2*$^{tm1.1Lan}$/J mice to generate whole body heterozygous knockout (*Sox2*^{het}) mice. We found that heterozygous loss of *Sox2* significantly delayed pubertal onset in both male and female *Sox2*^{het} mice compared to wild type (WT) controls (male WT: 29.7±0.7 days, *Sox2*^{het}: 32.3±0.5 days and female WT: 28.8±0.9 days, *Sox2*^{het}: 33.8±1 days). This difference is explained in part by a growth phenotype with *Sox2*^{het} mice being significantly smaller at weaning and into adulthood despite observing no difference in weight at birth. We next examined the effect of *Sox2*^{het} on fertility in both sexes. *Sox2*^{het} male mice had a significant decrease in sperm motility compared to WT controls (WT: 40.4%±1.5, *Sox2*^{het}: 29.5%±1.3); however, this decrease was insufficient to affect time to first litter. In females, we observed a significant delay in time to first litter of *Sox2*^{het} dams compared to WT controls (WT: 20.6±0.4 days, *Sox2*^{het}: 26.5±1.7 days), which likely arose from disrupted estrous cyclicity. Although *Sox2*^{het} females did not spend an extended period of time in any one stage, they did not progress linearly through the stages as did their WT controls and therefore completed fewer cycles during our study (WT: 3.6±0.3 cycles, *Sox2*^{het}: 2.5±0.2 cycles in 21 days). Finally, we assessed the number of GnRH neurons in the adult mouse and found that *Sox2*^{het} mice have significantly fewer GnRH neurons (WT: 570.6±42.8 neurons, *Sox2*^{het}: 228.4±32.8 neurons), a potent regulator of gonadotrophin release. Together these data further substantiate a role for *SOX2* in the endocrine system by addressing the effects of *Sox2* heterozygous knockout on the hypothalamus and extend the known spectrum of *SOX2* related disease burden.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6034T: Estimating the proportion of disease heritability mediated by molecular traits

Authors:

B. Strober[1], J. Rossen[2], A. Price[3]; [1]Harvard Sch. of Publ. Hlth., Brookline, MA, [2]Harvard Sch. of Publ. Hlth., Boston, MA, [3]Harvard Sch Pub Hlth, Boston, MA

Abstract:

Recent work has estimated that only 11% of disease heritability is mediated by cis-predicted gene expression in healthy adult tissues (MESC; Yao et al. 2020 Nat Genet), raising intense interest in explaining the remaining heritability. It is widely hypothesized that incorporating gene expression in other cellular contexts, or other molecular traits such as RNA splicing or protein levels, will help close this gap. Here, we consider the possibility that estimates may be downward biased due to imperfect predictions of cis-genetic gene expression arising from finite molecular trait sample size; this bias may be exacerbated by selection causing disease genes to have lower cis-heritability and hence lower cis-genetic prediction accuracy (Mostafavi et al. 2023 Nat Genet). We introduce a statistical method that estimates the proportion of disease heritability mediated by the cis-genetic components of a set of molecular traits while accounting for finite molecular trait sample size; the method inputs GWAS summary statistics, reference LD, and individual-level molecular cis-QTL data. We first obtain a downward biased estimate of the mediated disease heritability using cis-predicted gene expression, similar to MESC. We then correct the biased estimate based on the proportion of variance in cis-predicted gene expression attributable to noise arising from finite molecular trait sample size. In addition, we jointly model correlated molecular traits to estimate the unique contribution of each molecular trait (e.g. gene expression, RNA splicing, protein levels) in each tissue. Our method provides unbiased estimates of the proportion of disease heritability mediated by cis-genetic components of molecular traits in extensive simulations involving real genotypes, correlated molecular traits, and finite molecular trait sample size (such that true cis-*genetic* components explain more heritability than cis-*predicted* molecular traits). We will report results of applying the method to 78 diseases and complex traits using bulk and single-cell RNA-seq and RNA splicing data in diverse tissues and cell types, plus protein levels in whole blood.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6035T: Evolutionary loss of chitinolytic activity in YKL-40 (HC-gp39, CHI3L1) is caused by multiple sequence modifications that did not affect its affinity to chitin

Authors:

K. Suzuki, K. Okawa, M. Ohkura, T. Kanaizumi, T. Kobayashi, K. Takahashi, E. Tabata, F. Oyama; Kogakuin Univ., Hachioji, Japan

Abstract:

YKL-40, also known as human cartilage glycoprotein-39 (HC-gp39) or CHI3L1, is a protein similar to chitotriosidase (CHIT1), an active chitinase, but, in contrast to CHIT1, it lacks chitinase activity. Since YKL-40 is increased in inflammatory disorders and cancer, it has been used as a diagnostic marker. Elucidation of the biochemical properties of YKL-40 may lead to elucidating the causes of diseases and establishing treatments. Here, we report that YKL-40 has lost its ability to degrade chitin while maintaining its structural frame with affinity to chitin. The lack of chitinase activity in YKL-40 has been generally attributed to amino acid substitutions in the catalytic motif. Therefore, we first attempted to activate YKL-40 by introducing A138D and L140E mutations at this motif generating mutant YKL-40 (MT-YKL-40). However, this intervention did not lead to the activation of the protein. Next, we aimed to locate the exons responsible for the YKL-40 inactivation by creating various chimeras between the MT-YKL-40 and CHIT1. Introducing YKL-40 exons 3, 6, or 8 into CHIT1 CatD resulted in chitinase inactivation. Conversely, incorporating CHIT1 exons 3, 6, and 8 into MT-YKL-40 led to its activation. Our recombinant proteins exhibited properly formed disulfide bonds, affirming a defined structure in active molecules. Detailed analysis revealed that the reduced chitinase activity of MT-YKL-40 correlates with specific amino acids in exon 3 (I at position 61 and W at position 69). Altering M61I and T69W in CHIT1 CatD diminished chitinase activity and increased chitin binding. Conversely, substituting I61 with M and W69 with T in MT-YKL-40 triggered chitinase activity while reducing the chitin-binding activity. These findings underscore that YKL-40, though retaining the structural framework of a mammalian chitinase, has evolved to recognize chitin while surrendering chitinase activity.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6036T: Exploration in a Latin American sample of the genotype-specific effect of eQTLs on expression of *PM20D1*, a gene linked to Alzheimer's disease

Authors:

G. Chavarria-Soley, M. Fonseca-Brenes, L. Mora-Villalobos, H. Raventos; Univ. de Costa Rica, San José, Costa Rica

Abstract:

The *PM20D1* gene has been found to be hypermethylated in Alzheimer's disease (AD). A model has been proposed in which *PM20D1* plays a neuroprotective role and is therefore overexpressed in the early stages of cognitive impairment, with an increase in methylation and decrease in expression of the gene in later stages of the disorder. Several SNPs in the gene's vicinity, which are in linkage disequilibrium (LD), act as methylation and expression quantitative trait loci (mQTLs and eQTLs). They interact with the *PM20D1* promoter through a CTCF-mediated chromatin loop. It is a well-known fact that Latin American populations (among others) are underrepresented in human genetics research. Our goal was to explore the effect of known eQTLs on *PM20D1* expression in a Costa Rican sample. The sample was recruited in the context of a previous research project. A total of 106 individuals, belonging to a family with high prevalence of psychosis, had whole-genome expression and whole-genome sequence data that we could use to focus on the *PM20D1* gene. For each person, we genotyped 6 previously reported eQTL SNPs (rs1172198, rs708727, rs823082, rs823088, rs1361754, and rs960603) and constructed haplotypes; additionally, we obtained normalized expression values for *PM20D1*. For each SNP, a Kruskall-Wallis test with a post-hoc Dunn's test was used to compare expression levels between genotypes. We found that expression of the gene differs significantly between genotypes for all SNPs except rs960603. We have previously observed that the CLM sample from Colombia in the 1000 genomes database is close in allele frequencies and LD patterns to the Costa Rican population. We used 1000 genomes information for the 6 eQTL SNPs to analyze LD blocks and haplotype frequencies for the CLM and CEU populations. The CEU population has two main haplotypes for the region, the high expression GGTCAC haplotype and the low expression AACTGT haplotype. Meanwhile, the CLM population has three main haplotypes, two associated with higher *PM20D1* expression (GGTCAC and GGTTAT) and one with low expression (AACTGT). It is known from previous studies, that the low expression haplotype is also linked to higher promoter methylation of the gene. This low expression haplotype has been described as an AD-risk haplotype, because hypermethylation prevents protective overexpression of the gene when it is needed. A further exploration of the relationship between eQTL haplotypes and

the temporal fluctuations of methylation and expression levels could provide insights into AD risk.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6037T: Expression of Alzheimer's genetic risk genes in glial cells correlates with amyloid beta in human brain ⭐

Authors:

X. Zhong[1,2], Y. Tan[3,2], N. Cox[1,2], B. Li[3,2]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Genetics Inst., Nashville, TN, [3]Vanderbilt Univ., Nashville, TN

Abstract:

Background: Amyloid beta (Ab) accumulation is a hallmark of Alzheimer's disease (AD). It remains unclear about the genes, cell types and processes that are implicated in pathogenesis and clearance of Ab. We have previously identified a set of high-confidence risk genes (HRGs) for AD through an integrative analysis of AD GWAS. In this work, we set out to investigate the relationship between HRG expressions and amyloid burden in human brains. **Methods:** Using single-cell RNASeq data from >400 postmortem brains with various degrees of AD, we tested the association between HRG expression and amyloid-beta burden in 12 brain cell-types. We further adjusted for APOE-e4 status (e4-carriers vs. non-carriers) to examine whether such associations are independent of APOE-e4. **Results:** We detected a total of 93 HRG-Ab associations in 12 cell types, mostly in glial cells, with 30, 19, and 16 association in oligodendrocytes (Oli), microglia (Mic) and astrocytes (Ast), respectively. In contrast, only twelve HRG-Ab associations were detected in the otherwise abundant neurons. Notably, the HRG expression in glial cells are all positively associated with amyloid burden. After adjustment for apoeE4 status, all associations in the glial cells remained significant except for *APOE* and *CTSH* in microglia, suggesting a role of the HRGs independent of apoeE4. The top enriched pathways/processes (all P<1e-06) of the associated HRGs in glial cells include endocytosis in both Mic and Oli; amyloid precursor protein catabolic process in both Oli and Ast; efferocytosis, signaling by receptor tyrosine kinase and lymphocyte chemotaxis in Mic; and plasma membrane bounded cell projection morphogenesis, death receptor signaling and regulation of cell cycle process in Oli. Notably, genes such as *ABCA1, ADAM10 and ADAM17, CCNH, NCK2, PICALM, PLEKHA1* and *SORT1,* showed expression association with A in multiple glial cell types. **Conclusion:** Expression of AD genetic risk genes in glial cells are positively correlated with amyloid burden and are independent of APOE e4 status, suggesting a critical role of the genetic risk genes in glial cells in amyloid pathogenesis and clearance.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6038T: Functional analysis of the chr13q12.2 pancreatic cancer risk GWAS locus reveals allele-specific regulatory effects of rs9581943 on *PDX1* gene expression

Authors:

J. Hoskins, T. Christensen, D. Eiser, E. Char, M. Mobaraki, I. Collins, A. Jermusyk, A. O'Brien, K. Connelly, J. Zhong, L. Amundadottir; Natl. Cancer Inst., Rockville, MD

Abstract:

The largest published genome-wide association study (GWAS) for pancreatic ductal adenocarcinoma (PDAC) risk has identified 23 independent association signals. Functional studies of these signals can reveal novel biological processes relevant to PDAC tumorigenesis and resilience. The risk signal at chromosome 13q12.2 is tagged by rs2297316 (OR = 1.17, P = 9.82 x 10^{-14}) and overlaps the *PDX1-AS1* gene body and *PDX1* promoter. There are 24 candidate functional variants highly correlated (r^2 > 0.95) with the tag SNP. Allele-specific regulatory activity was observed by massively parallel reporter assay (MPRA) for rs9581943 (P = 1.03 x 10^{-6}), a promoter SNP 139bp upstream of the *PDX1* transcription start site, which was validated by luciferase assays (P = 0.0004-0.0786). Electromobility shift assays (EMSAs) revealed allele-preferential protein binding to each rs9581943 allele. *In silico* transcription factor (TF) binding prediction suggested E-box domain TFs likely bind the G allele while BHLH family TFs bind the A allele (the risk-increasing allele) and proteomic analysis of oligonucleotide pulldown of nuclear extract proteins for the two alleles is underway to identify relevant allele-preferential TFs. The GTEx consortium identified each of these candidate functional variants as significant expression quantitative trait loci (eQTL) with *PDX1* in both pancreas (P = 2.45 x 10^{-6}) and transverse colon (P = 1.62 x 10^{-5}) tissues. To test whether rs9581943 has allele-specific regulatory effects on PDX1 expression *in situ*, we performed CRISPR-Cas9 single-base editing of the PANC 05.04 heterozygous pancreatic cell line to GG and AA genotypes. *PDX1* expression in these clones recapitulated the GTEx eQTL where the risk allele (A) dosage correlates with lower *PDX1* expression (P = 2.20 x 10^{-7}). *PDX1* encodes a TF critical for pancreas development and homeostasis, and mutation of *PDX1* can cause maturity onset diabetes of the young (MODY) and is therefore involved in processes highly relevant to pancreatic tumorigenesis. Tetracycline-inducible pancreatic cell lines were generated to explore phenotypic consequences of *PDX1* over expression. Time course assays showed induction of *PDX1* over expression slowed proliferation and increased apoptosis but did not affect cell cycle phase distributions. Single cell

RNA-seq of pancreas tissue is ongoing to determine transcriptional patterns associated with *PDX1* expression in a cell resolved manner. In summary, our study has implicated rs9581943 as a functional variant that likely affects PDAC risk through allele-specific regulatory effects on *PDX1* expression, thereby influencing pancreatic cell differentiation and homeostasis.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6039T: Functional assessment of Dishevelled pathogenic variants in Robinow syndrome

Authors:

R. Sinha Roy[1], C. Zhang[2], J. Mazzeu[3,4], J. White[2], W-L. Charng[2], J. Gustafson[1], Z. Dardas[2], R. Sutton[2,5], H-K. Lee[6,7,8], J. Lupski[2,5,9,10], C. Carvalho[1]; [1]Pacific Northwest Res. Inst., Seattle, WA, [2]Dept. of Molecular and Human Genetics, Baylor Coll. of Med. (BCM), Houston, TX, [3]Univ. of Brasilia, Faculty of Med., Brasilia,70910-900, Brazil, [4]Robinow Syndrome Fndn., Anoka, MN, [5]Texas Children's Hosp., Houston, TX, [6]Dept. of Pediatrics, Section of Neurology, BCM, Houston, WA, [7]Jan and Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [8]Dept. of NeuroSci., BCM, Houston, TX, [9]Human Genome Sequencing Ctr., BCM, Houston, TX, [10]Dept. of Pediatrics, BCM, Houston, TX

Abstract:

The highly conserved Wingless (Wnt) signaling is associated with embryonic development, adult tissue remodeling and cancer cell biology. Perturbation of WNT signaling leads to developmental disorders such as Robinow syndrome (RS; MIM#180700). It is a genetically heterogenous rare disorder that is characterized by skeletal dysplasia and distinctive facial and physical characteristics. Seven genes coding for proteins that converge on the noncanonical WNT/planar cell polarity (PCP) signaling pathway: *DVL1, DVL2, DVL3, FZD2, NXN, ROR2,* and *WNT5A*, have been associated with RS. Frameshifting variants in DVL genes are all predicted to generate a novel C-terminus which is the most frequent cause of dominant RS, comprising 33% RS cases without *ROR2* variants. We conducted functional studies of three DVL gene paralogues to investigate the underlying mechanism of these heterozygous pathogenic frameshift variant alleles using an in vitro cellular transfection system. DVL homologs polymerize into dynamic super-complexes containing DVL1, DVL2, and DVL3, shown as DVL puncta, and change to even cellular distribution when DVL polymerization decreases. Accumulation of DVL molecules occurs with stimulation of WNT signaling. Immunohistochemistry experiments revealed that overexpressed DVL frameshifting alleles do not respond to the cellular localization changes induced by WNT ligands as observed in the wild-type (WT) alleles. Interestingly, the

altered WNT-stimulus induced localization changes of the mutant DVL supramolecular complex, shows a distinct pattern from short version of DVLs which suggest that the functional effects are not due to lack of the C-terminal tail. Moreover, we provide evidence that these DVL frameshift alleles in *DVL1/DVL3* (but not *DVL2*) interfere with CSNK1E-induced phosphorylation which indicates a potential molecular mechanism for the localization response. To test whether the DVL alleles change the WNT canonical activity, TOPFlash assays were performed and show that both WT and mutant *DVL1* and *DVL3* could activate WNT canonical signaling pathway. Frameshifting variants alone exhibited lower canonical activity whereas a combination of frameshifting variants and WT showed comparable level as WT alone, but they are unable to respond to WNT stimulus which is consistent with the immunochemistry results. Our results provide further understanding of the cellular consequences of frameshifting variants in *DVL1, DVL2*, and *DVL3* paralogues, and offers insights into the effects of such alleles on WNT signaling, and the perturbations thereof, that may lead to developmental phenotypes observed in RS.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6040T: Functional insights into the divergent role of *PIGF* in adipogenesis and osteogenesis

Authors:

I. Khrystoforova[1], M. Romero-Lado[1], G. Carrasquilla[1], A. Jafari[2], T. Kilpeläinen[1]; [1]Univ. of Copenhagen, Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Copenhagen, Denmark, [2]Univ. of Copenhagen, Dept. of Cellular and Molecular Med., Copenhagen, Denmark

Abstract:

Genome-wide association studies have identified multiple adiposity loci by examining associations with body fat percentage. Notably, a body fat percentage locus near the phosphatidylinositol glycan anchor biosynthesis class F (*PIGF*) gene shows an association not only with increased fat mass but also with reduced fat-free mass. This dual influence could amplify its impact on overall health. We aimed to gain functional insight into this divergent effect of the *PIGF* locus on body composition. We found that the *PIGF* locus is associated with reduced heel bone mineral density, suggesting that its influence on fat-free mass may stem from a reduction in bone mass. This was supported by findings in heterozygous *Pigf* knockout mice, which showed reduced lean body mass and bone mineral content alongside increased total fat mass. Given that bone mass is regulated by the commitment of bone marrow-derived mesenchymal stem cells (BM-MSCs) towards either bone-forming osteoblasts or fat-storing adipocytes, we hypothesized that *PIGF* may play a role in this process. To test our hypothesis,

we performed gene expression analyses and loss-of-function studies in human BM-MSC cultures established using expression of the telomerase reverse transcriptase gene (BM-MSC TERT4). RNAseq data showed distinct expression levels of *PIGF* between the adipogenic and osteogenic pathways, with adipocyte differentiation showing increased *PIGF* expression in contrast to its more stable expression during osteoblast differentiation. siRNA-mediated knockdown of *PIGF* resulted in enhanced early osteoblast differentiation, evidenced by increased alkaline phosphatase activity and increased formation of mineralized matrix, assessed by Alizarin Red staining. However, the formation of lipid-filled mature adipocytes was impaired upon *PIGF* knockdown in adipogenic cultures. These findings suggest that *PIGF*'s discordant influences on fat mass and fat-free mass may be explained by a regulatory role in early MSC commitment, where *PIGF* inhibition favors osteoblast differentiation of BM-MSC, at the expense of inhibited differentiation towards adipogenic lineage in vitro. Our study sheds light on the role of *PIGF* in bone and fat homeostasis, offering new perspectives on body composition regulation.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6041T: Genetic Editing of Archaic Alleles into Human Stem Cell Models: NOTO and Cortical Development

Authors:

T. Stephan, S. Salama, R. Green; Univ. of California, Santa Cruz, Santa Cruz, CA

Abstract:

While *Homo sapiens* are the sole survivors of our species, humans >40kya shared close relationships with Neanderthals and Denisovans: extinct hominins that live on in our genome through repeated interbreeding. These introgression events introduced genetic variation over generations, accounting for ~4% of the human genome. When looking at species-specific variation, there are ~600 non-synonymous single nucleotide polymorphisms (SNPs) that have become fixed in modern populations. A majority of these derived mutations are enriched in genes involved in brain development and pattern formation.
I hypothesize that by reverting genes that are fixed in modern humans to the ancestral alleles in stem cell models, we may better understand human evolution and the rise of human-specific phenotypes and morphology.
First, I developed a population-aware pipeline to identify and prioritize SNPs to be introduced in human inducible pluripotent stem cells (hiPSCs) using genome editing. For my pilot study, I focused on NOTO: the notochord homeobox gene. NOTO is required for node morphogenesis; regulates axial versus paraxial cell fate; regulates embryonic stem cell differentiation and

progenitor cell proliferation; and is a transcriptional regulator downstream of FOXA2 and Brachyury. Given that Neanderthals had elongated braincases compared to humans, I hypothesize that the mutation in NOTO might affect brain size and neuronal populations. CRISPR-Cas9 editing was to create ancestral (An) iPSCs with T|T at chr:2:73,438,011_hg19, while the unedited (Hu) line is A|A. Next, I used a cortical induction protocol to create cortical organoids: a 3-dimensional model made from stem cells that recapitulate neuronal development by using specific media and culturing techniques. The organoids were grown to day 70 and data - bulk RNA; whole organoids for snRNA; and fixed organoids for immunohistochemistry - were collected at D18, D35, and D70 timepoints to assay any differences in development and/or function. After D70, the organoids were given to a team to study electrophysiological activity using multielectrode arrays. Preliminary microscopy indicates differences in expansion rates and metabolic activity between An and Hu organoids, despite otherwise identical editing (the Hu line was mock transfected) and culturing techniques.
The An/Hu NOTO cortical organoid pilot serves as a model to investigate the effects of SNPs on early brain development in a human genetic background. This system can be generalized to investigate functional and morphological effects of SNPs in other cases wherein a human genetic model (iPSCs) would be preferred over non-human (e.g., mouse models).

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6042T: Genetic insights in FCGR2/3 locus: Role in susceptibility to Kawasaki Disease

Authors:

V. Joshi, K. Arora, R. K. Pilania, A. Rawat, S. Singh, D. Suri; Post Graduate Inst. of Med. Ed. and Res. (PGIMER), Chandigarh, Chandigarh, India

Abstract:

Background/Objectives: Kawasaki Disease (KD) is childhood vasculitis that primarily affects the coronary arteries. Intravenous immunoglobulin (IVIg) is of care for patients with KD. The constant region of IgG mediates its effector functions through Fc gamma receptors ($F_c\gamma Rs$). Genome Wide Association Studies have identified FCGR as a candidate gene for susceptibility to KD. High homology in this locus makes it difficult to study. Functional relevant Copy Number Region Variations (CNRs), polymorphisms, and chimeric genes have been reported.This study aims to investigate the significance of genetic associations in FCGR2/3 locus and its implications on response to IVIg among Indian patients with KD.**Methods:**Patients with KD (n=70), diagnosed based on American Heart Association (AHA) criteria 2017, and healthy controls (n=24) were

enrolled for screening FCGR2/3 locus using MLPA assay, MRC Holland.**Results:**CNRs were found in 22/70 (31.4%) patients. None of the controls tested revealed these variations. Odds Ratio (OR) showed that there is causal association of CNRs with KD. Odds of development of Coronary Artery Aneurysms (CAAs) (1.27 (0.47,3.27)) and IVIg resistance (1.65 (0.55,4.67)) were greater in patients with CNRs than without CNRs. Most common polymorphism (rs1801274) affecting the binding affinity of IgG had odds ratio of 1.65 (0.58, 4.55) with respect to development of CAAs.**Conclusion:** Patients with KD have CNRs in FCGR2/3 locus. Effect of CNRs correlates with development of CAAs and treatment resistance in these patients. These variants can provide better insights and help understand pathogenesis of KD and response to IVIg.
Grants: N/A

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6043T: Genetic risk for obesity predicts weight loss outcome in bariatric surgery through open chromatin variants regulating adipocyte gene expression

Authors:

S. Lee[1], A. Kar[1], S. Heinonen[2], M. Alvarez[1], S. Rajkumar[1], K. M. Garske[1], U. Säiläkivi[2], T. Saarinen[2], D. Kaminska[3], V. Männistö[3], M. Laakso[3], J. Pihlajamäki[3], A. Juuti[2], B. Balliu[1], K. H. Pietiläinen[2], P. Pajukanta[1]; [1]UCLA, Los Angeles, CA, [2]Univ. of Helsinki, Helsinki, Finland, [3]Univ. of Eastern Finland, Kuopio, Finland

Abstract:

Genetic susceptibility to obesity attributes to severe obesity, which is commonly managed by bariatric surgery. However, the effect of genetic obesity risk on the magnitude of weight loss is not well understood. We hypothesized that cell-type and context -specific genetic regulation of weight loss responding genes in subcutaneous adipose tissue (SAT) contributes to the inter-individual variability in the weight loss outcome. To this end, we longitudinally profiled SAT biopsies from 33 Finns from the RYSA cohort at 2 pre- and 2 post-bariatric surgery time points using single nucleus RNA-seq. We first used a pseudobulk approach to identify weight loss responsive genes and found 4,003 (FDR<0.05) differentially expressed genes (DEGs) by weight loss in adipocytes. Next, we performed cell-type level *cis*-eQTL mapping of genes expressed in adipocytes and used a multivariate adaptive shrinkage software mashr, which identified 7,217 genes with adipocyte *cis*-eQTL SNPs (local false sign rate<0.05) including 2,222 (56% of) weight loss DEGs. Among the 7,217 genes with *cis*-eQTL SNPs, 1,712 (23%) were shared across all 4 time points and 634-1,004 (9-14%) were specific to one time point, indicating both shared and

context-specific regulatory effects in adipocytes during weight loss. To further investigate the contribution of genetic risk for obesity on the weight loss outcome, we used the *cis*-eQTL SNPs of the weight loss DEGs to construct polygenic risk scores (PRS) for obesity and apply these PRSs to predict the change in BMI (dBMI) between pre- and post-surgery time points as follows. We first refined the *cis*-eQTL SNPs to those landing in open chromatin regions of adipocytes from ATAC-seq as we observed stronger heritability enrichment for BMI from SNPs landing in adipocyte ATAC peaks (h^2 enrichment=5.82, $p=6.22x10^{-9}$) when compared to all other SNPs (h^2enrichment =0.78, $p=6.18x10^{-9}$) in the UK Biobank. We further performed adipocyte co-expression network analysis using hdWGCNA to identify 16 adipocyte *cis*-eQTL SNPs in ATAC peaks regulating expression of 16 co-expressed weight loss DEGs. These 16 SNPs were used to construct a network PRS for obesity in an independent Finnish bariatric surgery cohort (n=200). We observed a significant enrichment of variance explained (R^2) in dBMI (R^2=2.98%; $p_{perm10000}$=0.02) by this network PRS, which performed comparably to the genome-wide PRS (R^2=4.24%; p_{R^2}=0.002). In summary, we discover significant contribution of genetic obesity risk on the weight loss outcome through regulation of 16 adipocyte weight loss genes, discovering them as weight loss magnitude regulators and indicating that the achieved weight loss is under genetic control.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6044T: Genome-wide association study identifies genetic variants associated with induced pluripotent stem cell growth rates

Authors:

B. N. Lee[1], H. J. Taylor[1,2,3], F. Cipriani[4], N. Narisu[1], C. C. Robertson[1,5], A. J. Swift[1], N. Sinha[1], T. Yan[1], L. L. Bonnycastle[1], N. Dale[4], D. Paull[4], A. Butt[4], H. Parsaud[4], NYSCF Global Stem Cell Array Team, M. R. Erdos[1], S. Noggle[4], F. S. Collins[1], D. L. Taylor[1]; [1]Ctr. for Precision Hlth.Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]British Heart Fndn. Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [3]Heart and Lung Res. Inst., Univ. of Cambridge, Cambridge, United Kingdom, [4]New York Stem Cell Fndn. Res. Inst., New York, NY, [5]Dept. of Computational Med. and Bioinformatics, Ann Arbor, MI

Abstract:

The differentiation of induced pluripotent stem cells (iPSCs) to diverse cell types is central to ongoing efforts to understand the effects of disease-associated genetic variants. Recent studies use an iPSC cohort design to explore the effects of common and rare variants on molecular

phenotypes (e.g., gene expression) and cellular phenotypes (e.g., morphology). As such studies become increasingly frequent, it is essential to understand how genetic factors may influence growth rates prior to differentiation.
In this study, we measured iPSC growth rates over multiple days with replication across 477 iPSC lines of European ancestry. To measure growth rates, we tracked confluence daily using a Celigo Image Cytometer. After fitting curves to summarize the measured growth patterns, we found the growth rates to be highly reproducible ($r_{Pearson} > 0.99$) across replicates. We conducted a common variant genetic association study using SAIGE including a genetic relatedness matrix (GRM) as a random effect. We identified variants associated with several growth metrics ($P < 5 \times 10^{-8}$). For example, one association at rs2582981 with mean growth rate has previously been reported to be associated with *CTNNBIP1*, a gene known to regulate cell proliferation. To explore if iPSC growth rates may be relevant to disease, we calculated bivariate genetic correlations between genetic regulators of iPSCs and genetic summary statistics of disease phenotypes via linkage disequilibrium score regression. We found evidence of association with uterine leiomyoma ($r_g > 0.75$), suggesting that genetic effects associated with cell growth rate may influence the development of this condition. Collectively, we anticipate that our findings will help control for differences in growth rates, particularly in the design of future, large-scale differentiation studies.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6045T: Genomic deletions of candidate enhancers at QT interval associated loci

Authors:

A. Asim[1], K. Vickery[2], A. Kapoor[1]; [1]UTHlth.Houston, Houston, TX, [2]UT Hlth.Houston, Houston, TX

Abstract:

Electrocardiographic QT interval (QTi), an index of ventricular de- and re-polarization, is a clinically significant, heritable quantitative trait. Prolongation or shortening of QTi is associated with an increased risk for cardiac arrhythmias and sudden cardiac death. Genome-wide association studies (GWAS) of QTi have identified trait-associated common noncoding variants at dozens of loci, including SCN5A and NOS1AP. Following a cis-regulatory mechanistic hypothesis underlying GWAS signals in the noncoding genome, we performed luciferase assay-based enhancer screens among associated variants, and identified five and six enhancer variants at the SCN5A and NOS1AP GWAS loci, respectively. Here, we aim to employ CRISPR/Cas9-based genome editing to create isogenic iPSC-CMs with deletions of these

candidate enhancers to assess their endogenous cis-regulatory activities. We designed paired guide RNAs (gRNAs), one upstream and one downstream of the candidate enhancer, to introduce double-strand breaks and deletion of the intervening sequence by nonhomologous end joining-mediated repair. gRNAs were designed using five tools, and gRNAs highly recommended by at least three gRNA-designing tools were selected. We used Bowtie2 alignment to ensure the selected gRNAs have a single hit in the human genome. To test the efficacy of our designed gRNAs, we evaluated them for introducing genomic deletions in 293T and AC16 cells. Protospacers were cloned via the oligonucleotide annealing method into the BbsI site of pX330 vector designed for dual expression of Cas9 and single gRNA. Cells were harvested 48h post-transfection of paired gRNAs constructs, followed by genomic DNA isolation and evaluation of deletion by PCR. Deletion bands were detected for all 11 targeted candidate enhancers in both cell lines, indicating the high efficacy of our designed gRNAs. Next, for targeted deletion of multiple candidate enhancers together at a locus, we generated gRNA expression arrays carrying all paired gRNAs (5 and 6 pairs for SCN5A and NOS1AP locus, respectively). Individual gRNA-containing cassettes were amplified by PCR to introduce different 4-base overhangs (post BsmBI digestion) for sequential ligation of multiple gRNA cassettes and cloning into a single array plasmid. In summary, we have assessed the efficacy of designed gRNAs to induce the desired deletions in two cell lines and have prepared constructs carrying multiple gRNAs that will be used for iPSC-CM-based experiments. Our goal is to create isogenic iPSCs-CM with these deletions and to assess their impacts on gene expression and cellular electrophysiology.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6046T: GSAP: A Comprehensive Pipeline Integrating Genomics and Structural Biology for Gene and Variant Impact Analysis

Authors:

V. Magalhaes Borges[1], N. Fagundes Borges Teruel[2], J. Kim[1], P. E. Dickson[1], L. Yongke[1], A. Mazur[1], J. Denvir[1], A. Q. Nato[1]; [1]Marshall Univ., Huntington, WV, [2]Univ. of Montreal, Montreal, QC, Canada

Abstract:

Visually informed understanding of interactions of genetic variants or protein active binding sites in three-dimensional space is critical for elucidating functional implications in various traits. Leveraging recent advancements in structural biology, notably AlphaFold, we developed a Genomic Structural Analysis Pipeline (GSAP) written in Python/R to assess the structural

consequences of disease-associated genetic variants or active sites. GSAP is mainly employed to perform mutation-driven analyses by: (1) evaluating allele-specific impacts of nonsynonymous variants on protein structure to predict potential effects on stability, folding, and interactions and (2) quantifying interactions of residues flanking the mutation to facilitate new insights regarding the etiology of the disease. GSAP also performs active site-driven analysis of protein-protein interactions or protein-ligand interactions. GSAP streamlines access to public datasets (e.g., GWAS Catalog, Protein Data Bank) and employs high-throughput scriptable methods to score interactions between wild-type and mutated proteins using MODELLER for homology modeling and Surfaces to quantify per-residue contributions to molecular interactions. We conducted various mutation-driven analyses, exemplified by: (1) investigating S1559N mutation on the HEATR1 gene using MODELLER to model the mutation, Surfaces to assess HEATR1-RNA interactions and NRGTEN to evaluate dynamic effects, results that when compared to those from the wild-type structure suggest a more unfavorable interaction with RNA and a less accessible/stable post-A1 state, possibly affecting blood pressure, (2) examining the impact of the R46S missense mutation in the Cidec gene, linked to obesity in TALLYHO/JngJ mice, revealing a significant reduction in dimer stability in the mutant Cidec protein, and suggesting disrupted interactions in obesity etiology, and (3) investigating variants in Fam53b, a QTL candidate gene for intravenous cocaine self-administration in mice, and a GWAS candidate for human cocaine dependence, along with Cyfip2. For active site-driven analysis, we used GSAP to study the interplay between Acyl-CoA oxidase 1 (ACOX1) and catalase (CAT) in ethanol metabolism. We performed docking using HDOCK and evaluated interactions between ACOX1 and CAT using Surfaces. HDOCK's top results offered a larger docking interface, suggesting a chain of docking complexes exploiting the symmetric nature of CAT and ACOX1 multimers. GSAP’s ability to integrate genomics and structural biology offers valuable insights into disease mechanisms, guiding therapeutic target exploration and drug design.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6047T: Heritability of transcript splicing and the development of splicing genetic scores for biobank-scale TWAS analyses

Authors:

A. Tokolyi[1], D. A. Knowles[2]; [1]New York Genome Ctr., New York, NY, [2]New York Genome Ctr. & Columbia Univ., New York, NY

Abstract:

Splicing quantitative trait loci (sQTLs) have been best estimated when located proximal to the constituent gene (*cis*-sQTLs). With the expansion of cohort sizes, our capacity to accurately estimate sQTLs positioned farther from the gene (*trans*-sQTLs) has markedly improved. However, we presently lack in-depth assessments of the cis and genome-wide heritability of transcript splicing. This knowledge would allow us to develop and utilize risk scores for transcript splicing in population-level biobanks, for comprehensive association with linked electronic health records.

Using previously derived sQTL data from the INTERVAL cohort comprising 4,732 healthy blood donors sampled across England, we investigate both the *cis* and genome-wide heritability of transcript splicing. We identify subsets of splicing events predominantly under genetic control, alongside modest yet significant levels of genetic control on splicing in many more sQTL-genes (sGenes) in whole blood. With these estimates and the previously computed sQTLs, we derive polygenic risk scores for transcript splicing phenotypes using Bayesian Ridge Regression. We validate their accuracy both on a withheld subset of the INTERVAL cohort, as well as using external whole blood GTEX data (n=670), revealing high concordance between internally and externally derived correlation.

Leveraging trait-level genetic covariance, we impute over 22,000 transcript splicing phenotypes (derived from ~6,000 sGenes) in individuals of European ancestry from the UK Biobank (n=312,940), as well as in individuals of diverse ancestry in the All of Us cohort (n=255,052). These predicted transcript splicing phenotypes are then used to perform a splicing transcriptome-wide association study, from which we uncover both known as well as novel associations between transcript splicing and health outcomes.

This study represents the first biobank-scale genetic estimation of transcript splicing using the current largest single-cohort sQTL study (INTERVAL), serving as a useful resource for future study of the heritability of transcript splicing, as well as the utilization of smaller targeted studies to impute various transcriptional phenotypes in biobanks of diverse populations for overlap with health trait data.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6048T: *HGSNAT, KDM6B, LMNA* and *WFS1* genes pathophysiological heterogeneity for neurodevelopmental disorders reveals novel, rare and ultra-rare variants in four unrelated probands of consanguineous families.

Authors:

B. Mudassir[1], Z. Agha[2]; [1]COMSATS Univ., Islamabad, Pakistan, [2]Univ. at Buffalo, New York, NY

Abstract:

Neurodevelopmental disorders of genetic aetiology are a highly diverse set of congenital recurrent complications triggered by irregularities in the basic tenets of brain development. An exceptionally well-organized neural development cascade governs the formation of human brain. This multifactorial process includes neuronal organization, proliferation, and maturation, without any malformation or disruption events impacting the structure and function of the central nervous system, failing which leads to various cognitive diseases. We present whole exome sequencing analysis and expression characteristics of the probands from four unrelated Pakistani consanguineous families with facial dysmorphism, neurodevelopmental, ophthalmic, auditory, verbal, psychiatric, behavioural, dental, and skeletal manifestations otherwise unexplained by clinical spectrum. Objectives of the study were to disclose the genetic disruption associated with the expression of abnormal neurodevelopmental syndromes. Whole Exome Sequencing identifies a novel, bi-allelic, missense variant in ***HGSNAT*** gene [NM_152419.3: c.1411G>A (p. Glu471Lys) exon 14] for proband family E-1 and a rare, bi-allelic, non-frameshift variant in ***KDM6B*** gene [NM_001348716.2: c.786_791dupACCACC (p. Pro263_Pro264du) exon 10] for proband family E-2 while a novel, mono-allelic, missense variant in ***LMNA*** gene [NM_170707.4: c. 1328 A>G (p. Glu443Gly) exon 8] for proband family E-3 and an ultra-rare, mono-allelic, missense variant in ***WFS1*** gene [NM_006005.3: c.2131G>A (p. Asp711Asn) exon 8] for proband family E-4. Protein modelling shows conformation and size modifications in mutated residues causing damage to the conserved domains expressed as neurocognitive pathology. The current study broadens the distinctly cultural and genetically inbred pool of the Pakistani population for harmful mutations, contributing to the ever-expanding phenotypic palette. The greatest aspirations are molecular genetic profiling and personalised treatment for individuals with complex neurological symptoms, which will allow for improved medication, management of daily activities, and a longer life duration.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6049T: Human Inner Ear Organoids Validate Coding and Enhancer Variants in Cochlear Malformations

Authors:

M. Zafeer[1], C. Abad[1], M. Ramzan[1], G. Bademci[1,2], M. C. Robayo[1], D. Duman[3], A. Mutlu[4,5], S. Seyhan[6], M. Kalcioglu[4,5], S. Fitoz[7], S. Guo[1], A. Griswold[1], D. Dykxhoorn[1,8], K. Walz[1,8], M. Tekin[1,8]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami,

FL, [2]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, Miami, FL, [3]Dept. of Audiology, Ankara Univ., Ankara, Turkey, [4]Faculty of Med., Dept. of Otorhinolaryngology, Istanbul Medeniyet Univ., Istanbul, Turkey, [5]Otorhinolaryngology Clinic of Goztepe Prof. Dr. Suleyman Yalcin City Hosp., Istanbul, Turkey, [6]Lab. of Genetics, Mem. Şişli Hosp., Istanbul, Turkey, [7]Dept. of Diagnostic Radiology, Ankara Univ. Sch. of Med., Ankara, Turkey, [8]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract:

Cochlear malformations are diagnosed in approximately 25% of children with congenital deafness. Most patients with cochlear malformations remain etiologically undiagnosed due to insufficient knowledge about underlying genes or the inability to make conclusive interpretations of identified genetic variants. To assess the utility of human inner ear organoids for functional validation of identified DNA variants, we first knocked out the two recognized genes, *FGF3* and *GREB1L,* in control iPSC lines via CRISPR/Cas9. Upon differentiation of iPSCs towards inner ear organoids, we observed significant developmental aberrations in knockout lines compared to their isogenic controls. In the same control cell line, we subsequently knocked in two novel missense variants in these genes and a truncating variant in *PBXIP1*, which we detected in patients with cochlear malformations. Variants in *PBXIP1* are not known to cause human phenotypes. These lines bearing the patient-specific variants resulted in severe developmental defects in the organoids, comparable to those observed with knockouts. To further utilize inner ear organoids, we used another iPSC cell line derived from a cochlear aplasia patient with a non-coding deletion on chromosome 8, previously linked with *GDF6* expression. Via ATAC-Seq in inner ear organoids, we identified two previously unknown active sites mapping to the deleted region. We showed that mice with a deletion in the orthologous non-coding genomic region are deaf and have cochlear aplasia. We are further refining these otic-lineage-specific enhancers in organoids by showing their relationship for controlling *GDF6* expression. Each organoid is initially grown in 96-well plates, which allows for multiple comparisons between different variants and their isogenic controls. Our study shows the utility of human inner ear organoids in rapidly validating coding and non-coding DNA variants in hereditary deafness. The insights gained from studying these genes and their regulatory sequences in cochlear development hold promise for developing new treatment strategies.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6050T: Human synonymous codon variation is constrained within multivalent coding regions

Authors:

R. Dore[1], J. Ule[1,2]; [1]UK Dementia Res. Inst. at King's Coll. London, The Wohl, London, United Kingdom, [2]The Francis Crick Inst., London, United Kingdom

Abstract:

Cellular homeostasis hinges on precise protein dosage control, particularly for proteins harboring disordered low complexity domains (LCDs). Recently, a new regulatory mechanism, termed 'interstasis', was reported which facilitates mutual dosage regulation of LCD-containing proteins (Faraway, et al). Interstasis leverages multivalent CDS regions encoding LCDs, which stems from the repetitive amino acid arrangement, similarity of codons for biophysically akin amino acids, and conserved synonymous codon biases. Interstasis involves specific RNA-binding proteins that bind to multivalent RNA regions that enable negative feedback regulation of mRNAs encoding specific sets of LCD-containing proteins (Faraway, et al).

To investigate the extent to which RNA multivalency can constrain human genetic variation in coding regions, we utilised common human polymorphisms from dbSNP. We assessed the nucleotide-level constraint of synonymous mutations compared to non-synonymous mutations for each codon (‘relative synonymous constraint’). We found that the more CDS regions are multivalent, the greater the relative synonymous constraint. Since the CDS regions encoding LCDs are the most multivalent, their relative synonymous constraint is also very high.

The six codons representing arginine are particularly variable and can be affected by both first and third base "wobble". Interestingly, human genetic variation in arginine codons has the propensity to promote RNA multivalency by preference for synonymous variation of arginine codons that maintain multivalency of the LCD-encoding RNA, such that the choice of arginine codons mimics the codons of other amino acids.

We also sought to understand the role of CDS multivalency in interpreting disease-related genetic variation. We used the Catalogue Of Somatic Mutations In Cancer (Cosmic) to find increased relative synonymous variation in multivalent regions (compared to dbSNP). Moreover, we were able to identify transcripts showing the highest enrichment of multivalency-disrupting synonymous mutations that could be prioritised for further functional validation.

Overall, our project has identified synonymous constraint patterns that are associated with the need for high multivalency of RNA encoding LCDs. This opens new opportunities for future investigations to explore the extent to which these constraints contribute to interstasis to control the dosage of LCD-containing proteins.

References:

Faraway, R., et al. (2023). Mutual homeostasis of charged proteins. BioRxiv, 2023.08.21.554177.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6051T: Identifying cell state-specific genetic effects on gene expression in human microglia

Authors:

B. Jang[1], R. Kubler[2], H-H. Won[3], T. Raj[2]; [1]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Sungkyunkwan Univ, Seoul, Korea, Republic of

Abstract:

Background: Research into neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) has made significant progress through genome-wide association studies (GWAS), identifying over a hundred genetic loci associated with the disease. Human microglia have emerged as a crucial cell type in understanding the underlying mechanisms of neurodegeneration. However, despite these findings, the causal variants, genes, and pathways specific to microglia states underlying disease pathogenesis remain largely elusive, hindering a comprehensive mechanistic understanding of the disease.
Methods: To address this challenge, we conducted a large-scale meta-analysis of human brain nuclei data from Fujita, Mathys, Gabitto, and Bryois single-nucleus RNA-seq datasets, encompassing a total of about 5,845,812 brain nuclei from 1,083 donors, of which 260,441 were from microglia. Factor analysis was performed using Spectra to identify the microglia states. We conducted the eQTL mapping with the mmQTL program and colocalization analysis. We performed state eQTL (st-QTL) in the microglia dataset, which considers different microglial cell states at single-cell resolution.
Results: Our meta-analysis of human microglia datasets, incorporating single-nuclei resolutions, identified 5,217 cis-eQTLs at a false discovery rate (FDR) <0.05. Among these, 55 and 54 AD risk loci and PD were colocalized with eQTL signals of our meta-analysis (PP.H4 > 0.7). Moreover, our state-eQTL analysis identified additional state-eQTLs colocalizing with AD or PD GWAS that are not identified in the pseudobulk analysis. Many of these genes are specifically expressed in specific microglia states, including disease-associated microglia), lipid metabolism, phagocytosis, or homeostatic state.
Conclusion: Our study represents the most extensive human brain transcriptomic data collection. By identifying specific eQTLs, prioritizing functional variants and genes, and considering microglial cell states, we have taken a significant step forward in understanding the underlying mechanisms of AD and PD.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6052T: Identifying risk variants in the 3' untranslated regions affecting Alzheimer's Disease by integrative analysis of eQTL, microRNA, and protein binding profiling

Authors:

J. Baik[1], S. Yang[1], X. Liu[2], X. Ning[3], D. Kim[1], L. Shen[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]RIKEN, Yokohama, Japan, [3]Ohio State Univ., Columbus, OH

Abstract:

Background: Alzheimer's Disease (AD) is a neurodegenerative disorder that affects millions worldwide. While AD pathology is characterized by amyloid-beta plaques and tau tangles, the genetic architecture of AD remains to be fully understood. The 3' untranslated region (3' UTR) plays an important role in gene expression regulation via microRNA (miRNA) and RNA-binding protein (RBP) interactions, potentially implicating them in AD pathology. Large-scale genome-wide association studies (GWAS) have identified many genetic variants associated with AD, including those in the 3' UTR regions. However, whether these variants are truly causal, and the underlying genetic mechanisms remain largely unexplored. **Methods**: We propose an integrative analysis of potentially causal 3' UTR SNPs by exploring their impact on miRNA and RBP interactions. Initially, we utilized ROSMAP's cis-eQTL results to identify SNPs from our initial list of 3' UTR SNPs. These SNPs were associated with significant changes in gene expression levels. By leveraging ROSMAP's miRNA and proteomic data, we identified SNPs located in the binding regions of expressed miRNA and RBPs. Additionally, we conducted conservation analysis on these 3' UTR SNPs. **Results**: From our initial curated list of 1,756 AD-associated SNPs, we found 20 unique 3' UTR SNPs that significantly altered the expression level of the gene in which they were located. Three SNPs were found in potential miRNA binding regions (rs1136001, rs13205911, rs9767105). For RBP binding regions, all but one SNP were located in the potential RBP's binding region, indicating a disruption in normal gene expression regulation. Our conservation analysis showed that one SNP, rs9767105, had high phyloP and phastCons scores, indicating its highly conserved nature. A deeper analysis of rs9767105, located in the 3' UTR of LATS1, suggested that LATS1 gene expression dysregulation could be a potential factor in Aβ42-mediated neurodegeneration. **Conclusion**: We conducted an integrative analysis to identify risk loci among 3' UTR SNPs for AD. Our analysis highlighted the rs9767105 SNP of the LATS1 gene from 20 candidate 3' UTR SNPs for AD, potentially through Aβ42-mediated neurodegeneration.

This study may provide new insights for causal analysis and functional annotation of AD-associated 3' UTR SNPs.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6054T: Integrating Machine Learning and Functional Genomics to Prioritize CVD-associated Variants Altering Cardiac TF Function ⭐

Authors:

E. G. Peña-Martínez[1], D. Pomales-Matos[1], J. Messon-Bird[1], L. Sanabria-Alberto[1], A. Rivera-Madera[2], J. Medina-Feliciano[1], A. Barreiro-Rosario[1], J. Rivera-Del Valle[1], J. M. Rodríguez-Ríos[1], R. Velazquez-Roig[1], J. A. Rodríguez-Martínez[1]; [1]Univ. of Puerto Rico - Rio Piedras, San Juan, PR, [2]Univ. of Puerto Rico - Cayey, Cayey, PR

Abstract:

Genome-wide association studies have mapped over 90% of cardiovascular disease (CVD)-associated variants within the non-coding genome. Non-coding variants regulatory regions of the genome, such as promoters and enhancers, can alter the function of tissue-specific transcription factors (TFs) and gene expression. However, due to the overwhelming number of GWAS single nucleotide polymorphisms (SNPs) (>500,000), prioritizing variants for in vitro analysis remains challenging. In this work, we implemented a computational approach that considers support vector machine (SVM)-based TF binding site classification and cardiac expression quantitative trait loci (eQTL) analysis to identify and prioritize potential CVD-causing SNPs. First, we identified 1,535 CVD-associated SNPs within DNase I hypersensitivity genomic footprints (DGF) in cardiac tissue and enhancers that are active during heart development. We then performed a linkage disequilibrium (LD) expansion in multiple populations (EUR, AFR, SAS, EAS, and AMR, R^2=0.80) and filtered for variants with genotype-dependent gene expression in cardiac tissue (based on eQTL analysis), resulting in 14,218 SNPs. Afterward, we used human cardiomyocyte chromatin immunoprecipitation sequencing (ChIP-seq) data from NKX2-5, GATA4, and TBX5, three cardiac TFs that regulate heart development, to train a large-scale gapped k-mer SVM model to identify CVD-associated SNPs that altered cardiac TF binding. The model was tested by scoring ~520,000 human cardiac DGF within heart enhancers and generated binding motif logos. We observed binding motifs like those previously described for NKX2-5, GATA4, and TBX5. We tested the model by choosing nine SNPs predicted to alter NKX2-5 (rs59310144, rs6715570, and rs61872084), GATA4 (rs1506537, 56992000, rs2491506, and rs2301249), and TBX5- (rs7612445 and rs7790964) DNA binding activity and in eQTL with genes in cardiac tissue. All nine variants altered cardiac TF-DNA binding and gene expression levels in

cardiac tissue eQTLs, which were proportional to the changes in TF-DNA binding activity. To further evaluate the potential of CVD-associated SNPs to drive disease phenotypes, we performed a dual luciferase reporter assay to evaluate changes in gene expression. All changes were proportional to those observed in cardiac tissue, further suggesting that GATA4 dysregulation can be further studied as a causal mechanism. In summary, we present a comprehensive analysis comprised of in silico, in vitro, and cellular evaluation of CVD-associated SNPs predicted to alter cardiac TF function.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6055T: Integration of proteogenomics in cerebrospinal fluid and plasma with multiple GWAS identifies tissue-specific mechanisms dysregulated in Alzheimer's and other neurological traits ⭐

Authors:

D. Western[1], C. Yang[2], J. Timsina[1], P. Gorijala[2], M. Liu[1], W. Brock[1], A. Ruiz[3], M. Fernandez[3], M. Marquié[3], M. Boada[3], I. de Rojas[4], J. Rutledge[5], H. Oh[5], E. N. Wilson[5], Y. Le Guen[5], M. D. Greicius[5], T. Wyss-Coray[5], I. Alvarez[6], M. Aguilar[6], P. Pastor[7], C. Robins[8], D. J. Pulford[9], Y. Sung[1], C. Cruchaga[1]; [1]Washington Univ. in St. Louis, Saint Louis, MO, [2]Washington Univ. in St. Louis, St. Louis, MO, [3]Univ.t Internacional de Catalunya, Barcelona, Spain, [4]Ace Alzheimer Ctr. Barcelona, Barcelona, Spain, [5]Stanford Univ., Stanford, CA, [6]Univ. Hosp. Mutua Terrassa, Terrassa, Spain, [7]Univ. Hosp. Trias i Pujol and the Germans Trias i Pujol Res. Inst., Badalona, Spain, [8]GSK, Upper Providence, PA, [9]GSK, Stevenage, United Kingdom

Abstract:

While plasma proteogenomic (pQTL) studies have identified thousands of genetic regulators of protein levels, the relevance of their involvement in brain-relevant traits is unclear due to the blood-brain barrier. Cerebrospinal fluid (CSF), another biosample obtainable from living individuals, surrounds the brain and spinal cord and therefore may be more effective at identifying brain-relevant proteogenomic regulation. However, no large-scale analyses have compared the pQTL landscape between plasma and CSF, clouding our understanding of the relevance of these tissues to neuro-psychiatric traits. Here, we integrated plasma and CSF proteogenomics with GWAS to prioritize proteins relevant to 14 neurological traits and their mechanisms of action. We generated the largest-to-date CSF proteogenomic dataset (6,361 proteins; 3,506 samples) and a plasma pQTL dataset (6,164 proteins; 2,338 samples; 743 samples from overlapping individuals). We compared the magnitude and direction of the pQTLs and performed colocalization between them. To identify potentially causal proteins for 14

neurodegenerative and psychiatric traits, we performed proteome-wide association study (PWAS), colocalization, and Mendelian Randomization (MR). Finally, we performed pathway analysis to identify dysregulated mechanisms underlying these diseases. We found 2,477 pQTLs in CSF and 2,761 pQTLs in plasma. While we observed a moderate correlation of z-scores (R=0.65) between CSF and plasma pQTLs, only 25.7% (1,225/4,767) colocalized. We identified 190 protein-disease associations in CSF and 185 in plasma across 14 traits (via at least two of PWAS, COLOC, and MR). Of these, only 57 were shared across tissues. For Alzheimer's disease (AD), we identified 50 total associated proteins, of which 13 were significant in both tissues (i.e., TREM2, CR1, ACE), 25 in CSF only (i.e., CLN5, FCGR3B, IL34), and 12 in plasma only (i.e., PLG, CD55, AXIN2). While most proteins were disease specific (78.7% in CSF, 73.4% in plasma), we observed substantial overlap between schizophrenia and bipolar disorder, with nine proteins shared in CSF and twelve in plasma. Our study created the largest CSF proteogenomic resource and analyzed the overlap with a similar plasma pQTL dataset. We identified low levels of colocalization between CSF and plasma, indicating substantial pQTL tissue-specificity. While integration of these resources with neurological traits led to similar numbers of associated proteins, only 30% of them were shared across tissues. This suggests that plasma and CSF may be capturing distinct disease mechanisms, emphasizing the need for a multi-tissue approach to target discovery.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6056T: Integrative omics reveals the genetic and molecular landscape of immune responses and comorbidities in people living with HIV

Authors:

N. van Unen[1], J. Botey-Bataller[1,2], M. Blaauw[2], W. Vos[2], L. van Eekeren[2], V. Matzaraki[2], A. Verbon[3], N. Nguyen[1], A. Groenendijk[3], X. Jiang[1], M. Gupta[1], J. Dos Santos[2], M. Cleophas[2], C-J. Xu[1], L. Joosten[2], M. Netea[2], A. van der Ven[2], Y. Li[1,2]; [1]Ctr. for Individualised Infection Med., Hannover, Germany, [2]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [3]Erasmus Med. Ctr., Rotterdam, Netherlands

Abstract:

Persistent inflammation in people living with HIV (PLHIV) using antiretroviral treatment (ART) increases susceptibility to non-AIDS comorbidities. Here, we systematically profiled 1342 virally suppressed PLHIV of European ancestry across five omics layers and immune function. Integrating epigenomics, transcriptomics, proteomics, metabolomics, and immune responses

resulted in a set of latent factors linked to cardiovascular diseases, presence of carotid plaque and COPD in PLHIV. Mapping inter-individual differences at four omics layers to genetic variation identified 7,925 molecular QTLs, illustrating a common genetic regulation in PLHIV compared to healthy individuals. Mendelian randomisation analysis uncovered the host-genetic-driven changes in baseline molecules causally related with immune response upon stimulation. Lastly, harnessing all integration approaches revealed that the inflammasome, genetically regulated by the NLRP12 locus, contributes to systemic inflammation across multiple molecular layers. Our unique catalogue of genetic and molecular determinants of immune function in PLHIV, highlight molecular pathways driving inter-individual variation in immune response and comorbidities.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6057T: Interpretable GWAS by linking clinical phenotypes to quantifiable immune repertoire components

Authors:

Y. Tan[1], L. Wang[2], H. Zhang[3], M. Pan[1], D. Liu[4], X. Zhan[5], B. Li[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Penn State Coll. of Med., Hershey, PA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Penn State Univ., Hershey, PA, [5]Univ. of Texas Southwestern Med. Ctr., Dallas, TX

Abstract:

Bridging the gap between genotype and phenotype in GWAS studies is challenging. A multitude of genetic variants have been associated with immune-related diseases, including cancer, yet the interpretability of most variants remains low. Here, we investigate the quantitative components in the T cell receptor (TCR) repertoire to interpret the genetic associations of diverse human diseases. We first developed a statistical model to predict the TCR components using variants in the TRB and HLA loci. Applying this model to over 300,000 individuals in the UK Biobank data, we identified 2,309 associations between TCR abundances and various immune diseases. TCR clusters predicted to be pathogenic for autoimmune diseases were significantly enriched for autoantigen-specificity. Moreover, four TCR clusters were associated with better outcomes in distinct cancers, where conventional GWAS cannot identify any significant locus. Collectively, our results highlight the integral role of adaptive immune responses in explaining the associations between genotype and phenotype.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6058T: Investigating the impact of non-coding structural variants downstream of epilepsy-associated gene, *FOXG1* on chromatin architecture and gene regulation

Authors:

A. Ramamurthy, L. Aluru, M. Bandouil, E. Yoon, J. Calhoun, G. Carvill; Northwestern Univ., Chicago, IL

Abstract:

Structural variants (SVs) vary widely in size and include multiple chromosomal modifications. Epilepsy is a heterogenous neurological disorder characterized by recurrent seizures. Up to 12% of genetic epilepsies can be explained by SVs that disrupt coding regions causing gene dosage defects. SVs in non-coding (nc-) regions can perturb *cis*-regulatory elements (CREs) and/or topologically associated domain (TAD) boundaries important for genome architecture and gene expression. Whether nc-SVs cause epilepsy, and their putative pathogenic mechanisms are unknown. This study aims to define the molecular mechanisms by which nc-SVs influence epilepsy susceptibility. We identified deletions in the 14q12 genomic region downstream of *FOXG1* in individuals with epilepsy and related neurodevelopmental disorders (NDDs). This gene encodes a transcription factor essential for cerebral corticogenesis, and *FOXG1* haploinsufficiency is associated with epilepsy/NDDs. To test whether these 14q12 deletions alter *FOXG1* expression, we chose the minimum region overlapping these deletions (14q12del), ~0.1 Mb in size and ~0.94 Mb downstream of *FOXG1*. We utilized HAP1s, a haploid line expressing *FOXG1*, as their haploid nature facilitates efficient genome editing and easier interpretation of results. Using CRISPR editing, we created stable HAP1 lines harboring 14q12del. *FOXG1* mRNA expression was significantly reduced by 50% in 14q12del HAP1s relative to our isogenic controls but did not preclude *FOXG1* expression completely. To further explain this phenomenon, we examined chromatin interactions of *FOXG1* using Unique Molecular Identifiers (UMI)-4C. Long range interactions of *FOXG1* in wildtype HAP1s were indeed altered in the presence of 14q12del. Using Roadmap Epigenomics and in-house data, we observed active CREs overlapping *FOXG1*'s long range contacts. We are now further evaluating the regulatory activity of native and altered contact loci in HAP1s using Massively Parallel Reporter Assays (MPRAs). We hypothesize that disruption of multiple CREs is required to completely ablate *FOXG1* expression, and that the number of disrupted CREs may underpin genotype-phenotype relationships. This study will pioneer explorations of molecular mechanisms of epilepsy/NDD related nc-SVs. Moreover, the regulatory landscape of NDD genes is not fully understood; our study exemplifies the use of novel tools such as UMI4C and MPRAs to define FOXG1's CREs. A comprehensive understanding of nc-SVs will enable the identification of

misregulated genes that are potential targets of pharmacological agents and aid the development of therapy for those individuals with genetic epilepsies.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6059T: Investigating the mechanisms of cell-type-specific alternative polyadenylation

Authors:

A. Caputo[1], J. Liu[1], E. J. Wagner[2], A. Schaffer[3]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Dept. of Biochemistry and Biophysics, Univ. of Rochester Sch. of Med. and Dentistry, Rochester, NY, [3]Case Western Reserve Univ., Moreland Hills, OH

Abstract:

A necessary step for the functional maturity of nearly all eukaryotic pre-mRNAs is 3' cleavage and polyadenylation. Interestingly, over 70% of human transcripts have more than one site at which polyadenylation can occur, a phenomenon known as alternative polyadenylation (APA). APA produces mRNA isoforms that differ in stability, subcellular localization, translational efficiency, or even the encoded protein sequence. Poly(a) sites (PASs) are selected and processed by dynamic interactions of proteins known as the cleavage and polyadenylation (CPA) complex. Mutations in these proteins cause neurological disorders and are associated with poor cancer prognosis. Because PASs are present throughout gene bodies, aberrant APA caused by mutations of the CPA complex is thought to produce functionally diverse outcomes at the cellular level. Indeed, specialized RNA sequencing technologies have revealed dramatic alterations in PAS usage and significant impacts on cell viability when CPA complex proteins are perturbed.

Interestingly, PAS patterns across identical gene sets vary widely by cell type. Exploring the link between cell-type-specific APA and CPA complex activity, we have identified distinct, cell-type-specific responses to genetic knockout of the canonical CPA constituent, CLP1. CLP1 is dispensable for human embryonic stem cell (hESC) viability but is essential for the viability of hESC-derived motor neurons. A potential explanation for these observations is that the composition of the CPA complex varies between cell types. To test this hypothesis, we will first identify differences in CPA complex composition between three isogenic cell types: hESCs, motor neurons, and skeletal muscle. Next, we will determine whether APA differences observed between cell types are due to variation in CPA complex composition. To supplement these findings, we will also explore the connection between selective vulnerability of neurons to CLP1 depletion and the composition of the CPA complex across different cell types.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6060T: KIR associations with plasma protein levels and phenotypes in the UK Biobank

Authors:

C. Adams[1], D. Roe[1], C. Benner[1], H. A. Hejase[1], R. K. Pendergrass[1], A. Mahajan[1], C. Hammer[2], M. I. McCarthy[1]; [1]Genentech, South San Francisco, CA, [2]Altos Labs, Redwood City, CA

Abstract:

Killer-cell Immunoglobulin-like Receptor (KIR) genes encode receptors on Natural Killer (NK) cells and some T-cell subsets. Variability in KIR genotypes is known to influence susceptibility to and severity of various diseases, including infections, autoimmunity, and cancer. However, there is limited insight into how genetic variation in KIR genotypes impact disease phenotypes and protein levels in large studies. The recently released UK Biobank WGS data provides a unique opportunity to investigate understudied genetic variation such as KIR. We performed KIR gene in silico genotyping using KIR Probe Interpretation (KPI) for all UK Biobank participants with WGS data. KPI predicts the presence or absence of each KIR gene with greater than 99% accuracy using previously discovered 25mer gene markers. We then performed analyses to detect associations between 12 KIR genes, 2,940 plasma protein levels (PPLs) (n=44,462), and 1,500 phenotypes (n=347,383) in UKB participants of European ancestry. Within major genetic ancestry groups, frequencies of our KIR genotypes were consistent with published data. For KIR genes with measured PPLs, gene presence was strongly associated with increases in their corresponding protein levels (*KIR2DL2*, *KIR2DL3*, *KIR2DS4*, and *KIR3DL1*, all p<1e-56). We also observed significant associations between KIR gene presence and proteins from genes located near KIR in the Leukocyte Receptor Complex (*FCAR*, *LAIR2*, and *LILRB4*) and proteins involved in NK cell activity (*GZMH*, *KLRB1*, *KLRD1*, and *KLRF1*). We found no evidence for association between individual KIR genotypes and disease phenotypes after Bonferroni adjustment for multiple tests (n=1500 phecodes×12 KIR genes=18,000 tests). The strongest associations were between KIR2DS1 and Schizophrenia (OR=1.25, p=9.2e-5); and between KIR2DS3 and hemoglobinopathies (OR=2.19, p=6.6e-5). Furthermore, given the important role of KIR receptor and HLA class I ligand binding in the regulation of NK cell signaling, we also evaluated statistical interaction between HLA class I ligands and their corresponding KIR receptors (e.g. Bw4×*KIR3DL1*) for disease phenotypes. While several significant interactions were identified, sensitivity analyses indicated these were largely driven by underlying HLA-phenotype associations rather than HLA-KIR binding. This is the first study to report on KIR gene presence/absence genotyping and associations between KIR genotypes, PPLs, and disease in

UKB participants with WGS data. Our lack of evidence implicating KIR presence/absence genotypes in disease highlights the need for analysis of KIR at the allele level.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6061T: Loss of EPHB4 function leads to hyperproliferative phenotypes in model systems of central conducting lymphatic anomaly ⭐

Authors:

M. Battig[1], M. March[1], L. Matsuoka[1], S. Sheppard[2], C. Kao[1], C. Seiler[1], D. Li[1], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]NIH, Rockville, MD

Abstract:

Central conducting lymphatic anomaly (CCLA) is one of many rare and progressive disorders of the lymphatic system. It is characterized by the dilation and dysfunction of the thoracic duct and cisterna chyli, which cause retrograde flux or abnormal drainage of lymphatic fluid and the manifestation of clinical symptoms such as chylothorax, chylous ascites, effusions, and/or lymphedema. To understand the etiology of this disorder, we have identified several variants in *EPHB4*, which encodes ephrin type-B receptor 4 and regulates important lymphangiogenic processes. We began investigating the functional impact of these variants using an *in vitro* spheroid sprouting assay with transduced human dermal lymphatic endothelial cells. A hyperproliferative phenotype, characterized by an increase in the number and length of capillary-like features extending outward from the spheroids, was observed with many of these variants. Moreover, this phenotype was exacerbated by VEGF-C, a potent inducer of lymphangiogenesis, and was unaffected by ephrin B2, the ligand of EPHB4. To determine the functional effects of these variants *in vivo*, ephb4 was knocked down in zebrafish with morpholinos, which demonstrated lymphatic vessel misbranching and developmental deformities similar to the lymphatic presentation observed in patients. Further analyses revealed that loss of EPHB4 function and subsequent mTOR overactivation led to the observed phenotypes. Importantly, the phenotype could be rescued in both the spheroid-sprouting and the zebrafish models using mTOR inhibitors, including rapamycin and OSI-027, a dual mTORC1 and mTORC2 inhibitor. These findings offer new insight into how *EPHB4* variants give rise to CCLA, and how targeting mTOR activation may be therapeutic for patients with CCLA.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6062T: Metabolome-wide association of carotid intima media thickness identifies FDX1 as a determinant of cholesterol metabolism and cardiovascular risk in Asian populations.

Authors:

N. Sadhu[1], R. Dalan[1,2], P. Jain[1], C. Lee[3,4], L. Pakkiri[3], K. Tay[1], T. Mina[1], D. Low[1], Y. Min[3], M. Ackers-Johnson[3,4], T. Thi[3], V. Kota[3], Y. Shi[3], Y. Liu[3], H. Yu[3], K. Wong[5], M. Lam[6,7], R. Sarangarajan[5], J. Lee[1,7], E. Lee[1,8], J. Ngeow[1,9], P. Tan[10,11], C. Cheung[1,10], C. Drum[3], R. Foo[3,4], G. Michelotti[5], H. Yu[3], P. A. Sheridan[5], M. Loh[1,12,13], J. C. Chambers[1,11,12]; [1]Nanyang Technological Univ., Singapore, Singapore, [2]Tan Tock Seng Hosp., Singapore, Singapore, [3]Natl. Univ. of Singapore, Singapore, Singapore, [4]Natl. Univ. Hlth.System, Singapore, Singapore, [5]Metabolon, Durham, NC, [6]Nanyang Technological Univ., Sinagpore, Singapore, [7]Inst. of Mental Hlth., Singapore, Singapore, [8]Natl. Hlth.care Group Polyclinic, Singapore, Singapore, [9]Natl. Cancer Ctr., Singapore, Singapore, [10]Agency for Sci., Technology and Res., Singapore, Singapore, [11]Precision Hlth.Res., Singapore, Singapore, [12]Imperial Coll. London, London, United Kingdom, [13]Natl. Skin Ctr., Singapore, Singapore

Abstract:

Background and Objective: The burden of cardiovascular disease (CVD), and CVD-related mortality is rising in the Asia-Pacific region, in contrast to its falling mortality rates in Europe and North America. With the objective of exploring the metabolic signature of atherosclerosis, and discovering new insights into the underlying pathways influencing CVD risk in Asian populations, we performed a metabolome-wide study of carotid intima media thickness (cIMT), among 8,124 participants from the Health for Life in Singapore (HELIOS) study (53±12 years, female=59%, Chinese=67%, Malay=14%, Indian=19%). **Methods and Results:** cIMT, a marker of atherosclerosis, was assessed in the carotid bulb and distal common carotid artery using 2D/3D carotid ultrasonography. 883 metabolites were quantified by untargeted mass spectrometry on Metabolon's Global Discovery Panel. We investigated the relationship between metabolites and cIMT using linear regressions, while controlling for age, sex, ethnicity, batch and CVD risk factors (BMI, blood pressure, total cholesterol, smoking and diabetes status). We found that plasma concentrations of 3beta-hydroxy-5-cholestenoate (3BH5C), a cholesterol metabolite, associated inversely with cIMT (Beta[SE]=-0.013[0.002]). Genetic instruments supported a causal relationship of 3BH5C metabolic pathways on cardiovascular risk wherein increased plasma concentrations of 3BH5C associated with decreased risk of coronary artery disease. We noted a 5-6 fold higher effect size in Asians (OR_{GSMR}[95% CI]=0.89[0.87-0.92], OR_{IVW}[95% CI]=0.86[0.80-0.92]) compared to Europeans (OR_{IVW}[95% CI] = 0.98[0.96-0.99]). Colocalization analyses using whole blood cis-expression quantitative trait loci (QTL) data from HELIOS (N=1,228) and GTEx

(N=670), and protein QTL data from the UK Biobank (N=49,748), revealed the presence of a shared causal variant between 3BH5C plasma levels and ferredoxin-1 (FDX1) expression, a protein essential for sterol and bile acid synthesis. We then functionally validated *FDX1* as a key regulator of 3BH5C synthesis in the acidic pathway of cholesterol metabolism, in both hepatocytes and macrophages, through knockout and overexpression models. We also demonstrated that *FDX1* modified cholesterol efflux in aortic smooth muscle cells. **Significance:** This study advances our understanding of the metabolic basis for atherosclerosis, highlighting population specific differences, and identifies *FDX1* as potential therapeutic target for the prevention of CVD.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6063T: Metabolomic profiling of the entire UK Biobank enables interpretation of Mendelian and complex disease mutations ⭐

Authors:

L. Jostins-Dean[1], K. Schut[2], S. Kerminen[2], P. Wurtz[2], J. Barrett[2], Nightingale Health Biobank Collaborative Group; [1]Nightingale Hlth., Oxford, United Kingdom, [2]Nightingale Hlth., Helsinki, Finland

Abstract:

We have recently completed the generation of metabolomic profiles for the entire UK Biobank. Here we describe the use of this resource to understand and interpret rare pathogenic disease-causing variants. Our first aim was to investigate the impact of known pathogenic mutations, including mutations that cause autosomal dominant or recessive diseases or substantially increase the risk of common diseases, on human metabolism. Our second aim was to investigate whether metabolomic profiles can aid interpretation of candidate pathogenic variants.

We quantified 249 metabolite parameters from 490,830 UK Biobank participants. After quality control and combining with exome sequences we analysed 469,835 participants. Based on high-confidence pathogenic mutations in ClinVar, we identified 76 autosomal dominant diseases with ≥5 carriers and 84 autosomal recessive diseases with ≥50 carriers. We also analysed 207 missense variants from the GWAS Catalog that had a large impact (OR > 2) on the risk of 119 complex diseases. We tested for association between each metabolic parameter and mutation status in a linear regression. For familial hypercholesterolemia, we also fitted a logistic predictive model to identify pathogenic mutations, using LDL levels, statin use and 17 principal components of the metabolic parameters.

We identified ≥1 metabolite association at FDR < 0.05 for 17 diseases. These included expected associations, including dyslipidemia in familial hypercholesterolemia (FH) and hyperglycemia in monogenic diabetes. They also included novel associations, such as a decrease in serum concentration of fatty acids in carriers of dominant Noonan Syndrome 1 mutations in *PTPN11*. Our model for classifying pathogenic FH mutations had higher accuracy than a model including LDL and statin use alone (AUC of 0.868 vs 0.828). This model identified 8 FH variants in the gene LDLR marked as uncertain significance in ClinVar with significant (Bonferroni-corrected p < 0.05) evidence of predicted pathogenicity.
These results show how combined exome sequencing and large-scale metabolomic profiling in population biobanks can identify new mechanisms of metabolic dysregulation in genetic disease and improve our models for interpreting pathogenic mutations.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6064T: Modeling heterogeneity in single-cell perturbation states enhances detection of response eQTLs

Authors:

C. Valencia[1], A. Nathan[2], J. Kang[3], L. Rumker[4], H. Lee[5], S. Raychaudhuri[6]; [1]Harvard Med. Sch., Cambridge, MA, [2]Harvard Univ., Boston, MA, [3]Harvard Med. Sch. / Brigham & Women s Hospita, Cambridge, MA, [4]Harvard Med. Sch., Boston, MA, [5]Brigham and Women's Hosp. and Harvard Med. Sch., Cambridge, MA, [6]Brigham and Women s Hosp., Brookline, MA

Abstract:

Studying expression quantitative trait loci (eQTL) aims to improve the interpretation of GWAS results by linking genetic variants to expression levels of specific genes. However, eQTLs can be context-dependent. One way to understand context-dependent eQTLs is by studying response eQTLs (reQTL), eQTLs whose magnitudes change in response to perturbation such as infection or cytokine stimulation. Previous work has helped to link GWAS loci with reQTLs in conditions such as autoimmune diseases.Typically, such studies model the effect of perturbation as discrete response. However, perturbation experiments usually affect perturbed cells heterogeneously. Thus, standard model may be underpowered to find response QTLs. Single cell data enables us to assess heterogeneous responses to perturbation. Here, we demonstrated that modeling of per-cell perturbation state enhances power to detect reQTLs. We use public single-cell peripheral blood mononuclear cell (PBMC) data, to study the effect of perturbations with *Influenza A* virus (IAV), *Candida albicans* (CA), *Pseudomonas aeruginosa* (PA), and *Mycobacterium tuberculosis* (MTB) on gene regulation. For each perturbation experiment,

we defined the continuous perturbation state of each cell using a penalized logistic regression. To find reQTLs, we then used a Poisson mixed effects model of gene expression in single cells as a function of genotype and the genotype interactions with the discrete and continuous perturbation states, accounting for confounders and batch. We found 6 times more reQTLs compared to a bulk approach, on average across perturbations. Furthermore, our model detected 36.9% more reQTLs by accounting for single cell heterogeneity compared to the standard single cell discrete-reQTL model. In some cases, these reQTLs pointed to perturbation-specific disease relevance, such as the decrease in the eQTL effect of rs11721168 for the gene *PXK* in PBMCs exposed to IAV. This eQTL colocalized with a risk locus for systemic lupus erythematosus (SLE). Furthermore, we found that on average of 25% reQTLs have cell-type-specific effects. For example, in IAV the increase of the eQTL effect of rs10774671 for the gene *OAS1* was stronger in CD4+T and B cells. Similarly, in all four perturbation experiments, the reQTL effect for the gene *RPS26* was stronger in B cells. Our work provides a general model for more powerful reQTLs identification and underscores the value of modeling cell-level variation.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6065T: Multi-ancestry brain pQTL fine-mapping and integration with GWAS of 21 neurologic and psychiatric conditions ⭐

Authors:

T. Wingo[1], Y. Liu[2], S. Vattathil[3], E. Gerasimov[4], S. Ravindran[2], J. Liu[1], A. Shantaraman[1], S. Fatemeh[2], W. Erming[5], Z. Bin[5], R. Joseph[6], M. Younkin[6], E-T. Nilufer[6], P. De Jager[7], F. Edward[2], D. Doung[8], M. Epstein[9], D. Cutler[10], A. Levey[2], D. Bennett[11], N. Seyfried[2], A. Wingo[4]; [1]Univ. of California, Davis, Davis, CA, [2]Emory Univ., Atlanta, GA, [3]Univ. of California, Davis, Athens, GA, [4]Univ. of California, Davis, Sacramento, CA, [5]Mt Sinai, New York, NY, [6]Mayo Clinic, Jacksonville, FL, [7]Columbia Univ Med Ctr, New York, NY, [8]Emory Univ., Atlanta, CA, [9]EMORY Univ., Atlanta, GA, [10]Emory Univ. Sch. of Med., Atlanta, GA, [11]Rush Univ., Chicago, IL

Abstract:

To understand the shared and unique genetic control of brain protein expression among different ancestries and their ramifications on disease, we performed protein quantitative trait loci (pQTL) mapping in 1362 multi-ancestry brain proteomes. We first identified pQTLs in each population separately (African American, Hispanic/Latin American, and Non-Hispanic White) and found up to 40,870 candidate pQTLs in 7,513 brain proteins per population. Then, using multi-ancestry pQTL fine-mapping, we identified 2,190 causal pQTLs at PIP >50%. We found that the vast majority of the causal pQTLs (95.2%) were shared across the three populations at 75%

confidence level. These multi-ancestry shared causal pQTLs were highly enriched in promoter regions and exonic sites (non-synonymous, synonymous, and UTRs). Next, using these multi-ancestry shared causal pQTLs as instrument variables, we performed Mendelian randomization to resolve GWAS signals to brain proteins that contribute to the pathogenesis of 21 neurologic and psychiatric conditions using population-matched pQTL and GWAS (21 traits in European ancestry, 10 in African ancestry, and 4 in Hispanic participants). We found 126 causal pQTL-protein pairs consistent with a causal role in these conditions. Remarkably, 29% of the causal pQTLs in these pairs were coding variants (22% non-synonymous and 7% in the UTRs). Since these coding causal pQTLs are shared across ancestries, they lay an important foundation for creating new molecular models for these polygenic disorders that are broadly relevant for individuals of diverse ancestries.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6066T: Multiple target proteins with altered-recycling result in varieties of pathological cellular processes, establishing the complex phenotypes of Ritscher-Schinzel syndrome 3

Authors:

Y. Nishio[1,2,3], K. Kato[1,2,3,4], A. Yoshida[1], K. Sawamura[5], M. Matsushita[5], H. Oishi[6], T. Ogi[3], P. Cullen[4], S. Saitoh[1]; [1]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [3]Dept. of Genetics, Res. Inst. of Environmental Med., Nagoya Univ., Nagoya, Japan, [4]Sch. of Biochemistry, BioMed. Sci. Building, Univ. of Bristol, BS8 1TD, Bristol, United Kingdom, [5]Dept. of Orthopedic Surgery, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [6]Dept. of Comparative and Experimental Med., Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan

Abstract:

Background: VPS35L forms a heterotrimer with VPS29 and VPS26C, known as the retriever complex, which plays pivotal roles in the endosomal recycling system. The retriever complex promotes the recycling of selective membranous proteins, specifically those with the NPxY motif, back to the cell surface, instead of their degradation in lysosomes. Reduced expression of certain membranous proteins results in pathological cellular processes depending on these functions. Previously, we established *VPS35L* as the causative gene for Ritscher-Schinzel syndrome 3 (RSS3) but the pathophysiological mechanisms remain to be elucidated. **Methods:** Surface proteome and murine analysis were performed to specify

causative surface proteins and demonstrate the pathophysiological impacts on the development of central nervous and skeletal systems. **Results:** We analyzed two lines of conditional knockout mice (*Vps35l f/f;Nestin-Cre* and *Vps35l f/f;Prx1-Cre* mice as neural crest- and mesenchymal stem cell (MSC)-specific lines, respectively). For neurological manifestations (e.g., intellectual disability), neural crest-specific cKO mice were analyzed, which exhibited failure to thrive, behavioral abnormalities, and lethality around postnatal day 14 to 21 (P14 to 21). Brain weight was significantly decreased, although no overt morphological abnormalities were observed. Immunoblot analysis for the synaptosome fraction of the hippocampus at P7 indicated decreased excitatory synaptic components, despite unchanged inhibitory components, suggesting that an unbalanced excitatory/inhibitory ratio (E/I ratio) would contributes to neurological phenotypes. On the molecular level, surface proteome analysis indicated the reduced surface expression of the SLITRK family, an important molecule for both excitatory and inhibitory synaptogenesis, suggesting that altered recycling of these proteins is one of the pathogeneses. In terms of skeletal phenotypes, MSC-specific cKO mice showed reduced body weight and shortened hindlimbs at P70. Transcriptome analysis of embryonic tibial chondrocytes indicated altered MAPK and WNT pathways, and insufficient endochondral ossification and extracellular matrix formation. This result was consistent with the altered surface proteins observed in proteome analysis, which we confirmed as one of the molecular backgrounds for skeletal phenotypes. **Conclusion:** We specified some of the target proteins with altered recycling, contributing to the pathogenesis of RSS3. Step-by-step delineation of target proteins deepens the understanding of the pathophysiology and would contribute to the development of therapeutic approaches.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6067T: Multi-study pQTL analysis of circulating aptamer-based proteomics in multi-ancestry TOPMed Cohorts

Authors:

C. Debban[1], U. Tahir[2], K. A. Pratte[3], J. Brody[4], M. Lee[5], C. Guo[3], A. Hill[3], J. Nicholas[6], D. Katz[7], B. Yu[8], J. Wilson[9], J. G. Wilson[10], H. Lin[11], K. Kechris[12], S. Gharib[13], S. Rich[1], K. Taylor[14], J. Rotter[15], P. Ganz[16], R. F. Dubin[17], R. Deo[18], D. Meyers[19], B. Psaty[13], M. Cho[20], S. London[21], R. Gerszten[10], L. Raffield[22], R. Bowler[3], A. Manichaikul[23]; [1]Univ. of Virginia, Charlottesville, VA, [2]BIDMC, Lexington, MA, [3]Natl. Jewish Hlth., Denver, CO, [4]Univ of Washington, Seattle, WA, [5]NIH/NIEHS, Rancho Palos Verdes, CA, [6]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [7]Stanford Univ., Stanford, CA, [8]UTHlth.Houston, Houston, TX, [9]Univ Mississippi Med Ctr, Jackson, MS, [10]Beth Israel Deaconess Med. Ctr., Boston, MA, [11]UMass Chan Med. Sch., Worcester, MA, [12]Univ. of Colorado

Anschutz Med. Campus, Aurora, CO, [13]Univ. of Washington, Seattle, WA, [14]The Lundquist Inst. for BioMed. Innovation, 1124 W. Carson Street, Torrance, CA 90502, Torrance, CA, [15]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [16]Univ. of California San Francisco, San Francisco, CA, [17]Univ. of Texas Southwestern, Dallas, TX, [18]Univ. of Pennsylvania, Philadelphia, PA, [19]Univ. of Arizona, Tucson, AZ, [20]Brigham and Women s Hosp., Duxbury, MA, [21]NIEHS, Research Triangle Park, NC, [22]UNC - Chapel Hill, CHAPEL HILL, NC, [23]Univ Virginia, Charlottesville, VA

Abstract:

Integration of genome-wide association study (GWAS) with gene expression quantitative trait loci (eQTL) has proven a valuable approach as a first step to identifying molecular mechanisms underlying GWAS signals. However, many GWAS loci do not show evidence of colocalization with eQTLs. Motivated by the hypothesis that high-throughput proteomics can complement eQTLs for enhanced interpretation of GWAS signals, we assembled the largest multi-ancestry pQTL resource to date by combining SomaScan proteomics aptamer-based data obtained from four community-based cohorts (Cardiovascular Health Study [CHS], Framingham Heart Study [FHS], Jackson Heart Study [JHS], Multi-Ethnic Study of Atherosclerosis [MESA]; total n=11,280), two smoking-enriched cohorts (COPDGene and SPIROMICS; total n=7076), and one asthma-enriched cohort (Agricultural Lung Health Study [ALHS]; n=1834). The proteomics measures span multiple ancestries, representing European Americans (EUR; n=10,842) and African Americans (AFA; n=5,508) with 1300-7000 protein aptamer measures per sample, depending on SomaScan version. Whole genome sequence (WGS) data with a minor allele count greater than 20 were obtained from the TOPMed program (CHS, FHS, JHS, MESA, SPIROMICS, and COPDGene) and from genome-wide imputation to the TOPMed reference panel for ALHS to perform pQTL mapping. Using these pooled samples and adjusting for age, sex, study, PCs of ancestry, and PCs of hidden variance, we detect significant *cis*-pQTLs for 2923 proteins based on a significance threshold of $q < 0.05$. In addition, we recapitulate well-described variant-expression associations such as the known *SERPINA1* S and Z alleles for alpha-1 antitrypsin levels (AAT) levels. In sex-stratified analysis, pQTLs are largely concordant between the sexes, with only 9.1% of pQTLs significant in only one sex. In ancestry-stratified analysis, we identify 1863 proteins with significant pQTL in both European- and African-ancestry groups, 789 proteins with significant pQTL in European-ancestry only, and 331 proteins with significant pQTL in African-ancestry only. Only 14.9% of pQTLs significant in both ancestry strata shared the same lead variant, but 72.5% have overlapping credible sets after statistical fine-mapping, indicating shared causal loci. Future analyses include *trans*-pQTL analysis and colocalization of pQTLs with GWAS to determine evidence of heterogeneity by strata. This pQTL mapping effort leverages high-throughput proteomics to integrate multi-ancestry ‘omics data to expand the repertoire of protein-associated variants and identify putative molecular mechanisms underlying GWAS signals.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6068T: Non-linear genetic regulation of the blood plasma proteome

Authors:

J. Gräf[1,2,3], A. I. Sigurdsson[1,2,3], Z. Yang[4], K. Ravn[5,2], J. Meisner[1,2,6], R. Thielemann[7,2,3], H. Webel[1,2], R. Smit[1,3,8], B. Vilhjalmsson[9,3,10], B. Neale[11,3,12,13], A. Ganna[14,12,13,15], T. Hansen[16], R. Loos[7,3,8], S. Rasmussen[1,2,3]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Copenhagen, Denmark, [2]Novo Nordisk Fndn. Ctr. for Protein Res., Copenhagen, Denmark, [3]Novo Nordisk Fndn. Ctr. for Genomic Mechanisms of Disease, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Univ. of Helsinki, Helsinki, Finland, [5]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, D, Copenhagen, Denmark, [6]Mental Hlth.Ctr. Copenhagen, Copenhagen, Denmark, [7]Univ. of Copenhagen, Copenhagen, Denmark, [8]Icahn Sch. of Med. at Mount Sinai, New York, NY, [9]Natl. Ctr. for Register-based Res., Aarhus BSS, Aarhus, Denmark, [10]Bioinformatics Res. Ctr., Aarhus Univ., Aarhus, Denmark, [11]Massachusetts Gen. Hosp., Boston, MA, [12]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA, [13]Analytic and Translational Genetics Unit, ATGU, Massachusetts Gen. Hosp., Boston, MA, [14]Inst. for Molecular Med. Finland, Helsinki, Finland, [15]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [16]Univ. of Copenhagen, København N, Copenhagen, Denmark

Abstract:

Although thousands of genetic variants are linked to human traits and diseases, the underlying mechanisms influencing these traits remain largely unexplored. One important aspect is to understand how proteins are regulated by the genome by identifying protein quantitative trait loci (pQTLs). Beyond this, there is a need to understand the role of complex genetics effects that regulate plasma proteins and protein biomarkers. Therefore, we applied EIR, a deep learning framework developed in our group to genetics and protemics data from the UK Biobank. From the cohort of 57,200 individuals and 2922 measured blood plasma proteins, we identified 171 proteins that were regulated by non-linear effects including non-linear covariates, genetic dominance and epistasis. We uncovered a novel epistatic interaction between the *ABO* and *FUT3* loci, and demonstrated dominance effects of the *ABO* locus on plasma levels of pathogen recognition receptors CD209 and CLEC4M. These findings could improve our understanding of the *ABO* locus as a regulator of the susceptibility to infectious diseases. We replicated these findings in the FinnGen project and could confirm dominance effects in the *ABO* locus acting on plasma levels of CD209 and CELC4M as well as epistatic interactions

between *ABO* and *FUT3*. Our approach presents the first systematic, large-scale attempt to identify complex effects of plasma protein levels.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6069T: Optical Pooled Phenotyping at Nucleotide Resolution ⭐

Authors:

S. Pendyala[1], N. Bradley[1], H-J. Kim[1], S. Fayer[1], K. Partington[1], K-C. Yang[2], L. Starita[1], **D. Fowler**[1]; [1]Univ. of Washington, Seattle, WA, [2]Inst. for Stem Cell and Regenerative Med., Univ. of Washington, Sch. of Med., Seattle, WA

Abstract:

Multiplexed Assays of Variant Effect have revolutionized single nucleotide genomic screening at scale. However, they have been restricted to single, general phenotypes like growth or protein abundance in simple cell line models. Current MAVEs exclude the majority of Mendelian disease-causing genes which are non-essential, cell type-restricted or drive diverse cell phenotypes. For example, the variable expressivity of genotypes in *LMNA* manifests clinical disease in diverse tissues, from cardiomyopathy to muscular dystrophy, lipodystrophy, and progeria. Because the map from genotype to phenotype in this gene is many-to-many, an assay that associates thousands of genotypes to hundreds of cellular phenotypes in disease-relevant model systems is needed. Therefore, we built Variant In-Situ sequencing (VIS-seq), a generalizable and scalable assay that maps genotypes to diverse phenotypes in disease-relevant model systems using *in-situ* sequencing to optically phenotype protein variant libraries. VIS-seq employs dual promoters to sequence an abundantly-expressed, barcoded circular RNA instead of the transgene RNA, improving reads per cell by an order of magnitude. VIS-seq enables measurement of localization of variant protein, other proteins via immunofluorescence, and cell morphology via four-dye Cell Painting in individual cells. VIS-seq can genotype and phenotype ~4 million cells per experiment and has a cost per cell two orders of magnitude lower than scRNA-seq. We used VIS-seq to measure the localization of ~1,700 coiled-coil *LMNA* missense variants in U2 OS cells. We extracted cell image profiles for each variant and clustered these variant-level profiles, enabling us to segregate aggregating control variants and perform unbiased discovery of new cellular phenotypes associated with different lamin-associated clinical phenotypes. We have also combined VIS-seq with a silencing-free transgene expression system to perform in-situ variant genotyping and phenotyping in iPSC and specialized cell types. We cloned GFP-tagged libraries of single variants in *LMNA*, *RPS19*, *PTEN*, and *CFTR* and expressed them in iPSCs. We are now applying VIS-seq to optically phenotype >3,000 missense

variants for pathogenic mislocalization in *RPS19* in iPSC, *PTEN* in iPS-derived neurons, and *LMNA* additionally in iPS-derived cardiomyocytes to compare variant effects in two cell types. By enabling pooled variant expression and *in-situ* sequencing in diverse cell types, VIS-seq expands Multiplexed Assays of Variant Effect to the broad set of genes with pre-existing optical phenotypes like mislocalization or cellular dysfunction.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6070T: Pervasive human mitochondrial-nuclear interactions and their implications for disease identified across >600,000 individuals

Authors:

R. Gupta[1], G. Chau[2], T. J. Durham[2], D. Howrigan[2,3], B. Neale, M.[3], V. K. Mootha[3]; [1]Harvard Med. Sch. / Broad Inst., Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Massachusetts Gen. Hosp., Boston, MA

Abstract:

The human mitochondrial DNA (mtDNA) is a small, circular, maternally inherited genome. Given that mtDNA is present in many copies per cell, mtDNA mutations are often characterized by heteroplasmy, where a mixture of alleles are found in an individual. All 13 proteins encoded by the mtDNA are required for oxidative phosphorylation, and mtDNA mutations can lead to a host of devastating maternally-inherited rare diseases such as mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS). All known proteins involved in mtDNA synthesis and maintenance, however, are encoded by the nuclear DNA (nucDNA). Mutations in nucDNA-encoded genes responsible for mtDNA maintenance can produce Mendelian disorders characterized by accrual of mtDNA deletions and mtDNA depletion.
Less is known about how common nucDNA variation can influence mtDNA to alter the risk of common disease. In the general population, per-cell mtDNA copy number (mtCN) declines with age while mtDNA mutations accumulate, however the mechanisms of these processes are not understood. In a recent study, we identified for the first time that common human nucDNA variation can influence mtDNA heteroplasmy, however, sample sizes precluded a comprehensive assessment of implications for common disease.
Here, we construct a new call-set of mtDNA variation and copy number across >600,000 individuals across 6 ancestry groups in UK Biobank (UKB) and AllofUs (AoU), expanding on previous sample sizes by over 100%. We perform GWAS for three types of phenotypes: (1) mtCN, (2) heteroplasmy of individual common mtDNA variants, and (3) an age-adjusted count of heteroplasmic mtDNA SNVs (which we have shown are typically somatic), identifying new

interactions between common nucDNA variants and mtDNA phenotypes. We systematically compare the genetic architectures of these phenotypes using genetic correlations, identifying shared mechanisms behind variation in different mtDNA regions. We also compare the nuclear genetic architecture of mtCN with a metabolomic predictor of MELAS (MitoScore), with our results highlighting the heterogeneity in molecular mechanisms that can perturb aspects of mitochondrial biology. Finally, we perform pheWAS across UKB using nucDNA variants durably associated with multiple mtDNA heteroplasmies.
Our work expands the repertoire of mito-nuclear interactions identified in humans with implications for our understanding of the basic mechanisms governing mtDNA maintenance and transmission as well as the impact of mtDNA on common human disease, shedding light on the mechanisms of age-related changes in mtDNA sequence and abundance.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6071T: Predicting missense mutation effects with biophysics-informed protein language model

Authors:

C. Hou, Y. Shen; Columbia Univ., New York, NY

Abstract:

Understanding the effects of missense mutations on protein function is critical for insights into disease mechanisms and protein engineering. Previous computational methods leverage evolutionary conservation and static structures to predict mutation impacts. However, proteins function through the intricate folding of linear sequences into ensembles of structures. Many pathogenic mutations do not lead to significant changes in the static structure but can alter protein conformational equilibria and dynamics. Studying the dynamics typically requires laborious experiments or intensive computational simulations, making proteome-wide mutation studies impractical. In recent years, Protein Language Models have emerged as novel tools for studying and engineering proteins by learning specific properties from protein sequences. Here, we introduce SeqDance, a novel protein language model pretrained on physicochemical features derived from over 50,000 normal mode analyses and molecular dynamics simulations. Biophysics-informed pretraining enables SeqDance to capture the structural dynamics of proteins from sequence. We showed that SeqDance can augment traditional methods in predicting the effects of missense mutations in mutational scanning experiments. We also fine-tuned SeqDance to predict effect of all possible missense mutations in human. Overall, SeqDance offers a complementary perspective to traditional evolution or static structure-based

methods, enabling a better and unbiased understanding of protein fitness landscapes, contributing to identifying pathogenic missense mutations and uncovering previously unknown disease-causing genes.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6072T: Predicting the functional impact of missense variants using a multi-gene classification model

Authors:

S. Padigepati[1], D. A. Stafford[1], M. R. Silvis[2], K. Jamieson[2], L. Fresard[1], K. Nykamp[1], J. Reuter[1]; [1]Invitae Corp., San Francisco, CA, [2]Epic Bio, South San Francisco, CA

Abstract:

One of the major challenges in genetic testing is resolving variants of unknown significance (VUS). To address this challenge, Multiplexed Assays of Variant Effect (MAVE) have been developed to characterize the functional impact of variants at scale. However, existing MAVE approaches are gene-specific, and the limited number of variant annotations in most genes constrains the training and validation of classification models when trained in a single-gene context. We hypothesize that genes related to a set of clinically overlapping disorders caused by disruption of a common biological pathway will produce similar benign and pathogenic variant effects and can be effectively modeled together. To test this hypothesis, compatible data would need to be collected and integrated for multiple genes in a common pathway, enabling the training of a multi-gene classification model.
Single cell RNA-seq (scRNA-seq) coupled with machine learning has recently emerged as a novel MAVE strategy, and gene expression-based measurements could serve as the foundation of multigene models. Therefore, we generated scRNA-seq data for variants from six genes in the RAS/MAPK signaling pathway: *SOS1*, *KRAS*, *RAF1*, *BRAF*, *MAP2K2* and *SPRED1*. For each gene, HEK293 cells were engineered to express a single exogenous variant copy using a landing pad approach. After identifying the variants in cells and removing low-quality data, we generated a high-quality dataset that consists of scRNA-seq data for 696 variants (~88 cells/variant), of which 511 variants were missense. We trained a variant prediction model on this full dataset, and the model exhibited high performance (~0.96 AUROC) when evaluated on established pathogenic and benign variants compiled from Clinvar as well as our internal variant database. Importantly, the performance of this model was comparable to single-gene models (for *BRAF*, *RAF1* and *SPRED1*) or significantly higher (for *KRAS*, *MAP2K2* and *SOS1*), demonstrating the value of combining multiple genes within a single pathway into one model.

This work opens up the possibility of providing high-quality pathogenicity predictions for genes with limited variants of known significance, if they are members of a well-conserved molecular pathway. In addition, we propose that multi-gene models of this kind can provide a deeper, mechanistic understanding of pathogenic variation and novel insights into therapeutic targets for specific variants within the pathway.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6073T: Predicting, understanding, and rescuing pathogenic vasopressin 2 receptor variants

Authors:

T. Mighell, B. Lehner; Ctr. for Genomic Regulation, Barcelona, Spain

Abstract:

Vasopressin receptor 2 (V2R) is a G-protein coupled receptor (GPCR) with a critical role in water homeostasis. Without v2R function the kidneys cannot concentrate urine, leading to nephrogenic diabetes insipidus (NDI). Individuals with NDI suffer chronic dehydration, with few treatment options, that can lead to severe clinical outcomes. Hundreds of V2R variants are linked to NDI; among the few with functional characterization, the main mechanism is decreased plasma membrane expression (PME). Interestingly, some hydrophobic V2R antagonists cross the plasma membrane, bind to, and stabilize nascent receptors, promoting proper membrane trafficking; these are known as pharmacological chaperones (pharmacoperones). In many cases pharmacoperone-rescued variants retain some signaling capacity at the membrane. Despite the clear promise of pharmacoperone therapy for NDI, it remains underexplored because for the vast majority of V2R variants, the PME and capacity for pharmacoperone rescue is unknown.

To address these questions, we established a framework to measure the PME of thousands of V2R variants in human cells in parallel. We developed Scalable and Uniform Nicking (SUNi) mutagenesis to generate a library encoding 93% of all single amino acid changes. HA-tagged V2R was used so that cells expressing receptor variants could be fluorescently stained and binned via cell sorting. Variants in each bin were sequenced and counted to calculate PME for all variants.

We measured PME for 7,216 variants with high reproducibility (r = 0.9). Mutation effects are strongest in transmembrane helices, and are related to hydrophobicity of the substituted amino acid. Glycosylation sites stand out with strong effects in unstructured loops, and mutations to cysteine are damaging in extracellular loops as they likely disrupt a key disulfide bond. As

reported, the majority of NDI alleles have a PME defect (56% null, 21% hypomorphic). Classifying pathogenic and benign variants with PME achieves an AUROC of 0.84, suggesting PME explains a large fraction of pathogenicity for V2R. Next, we measured PME after treatment with a known V2R pharmacoperone, tolvaptan. Strikingly, the vast majority of variants were rescued, including most NDI alleles (7% null, 41% hypomorphic after treatment). The few sites where variants were not rescued were glycosylation sites or the predicted tolvaptan binding site. Key future studies will ask if rescued variants retain signaling activity and if pharmacoperone rescue is as general among other GPCRs and protein classes. This work supports renewed interest in pharmacoperones and provides a framework for quickly assessing efficacy for all variants in a gene.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6074T: PreMode predicts mode-of-action of missense variants by deep graph representation learning of protein sequence and structural context

Authors:

G. Zhong[1], Y. Zhao[1], D. Zhuang[1], W. Chung[2], Y. Shen[1]; [1]Columbia Univ., New York, NY, [2]Boston Children's Hosp. and Harvard Med. Sch., Boston, MA

Abstract:

Accurate prediction of functional impact of missense variants is fundamentally important for disease gene discovery, clinical genetic analysis and therapeutic strategies. Previous efforts have focused on predicting a binary pathogenicity classification, but the functional impact of missense variants is multi-dimensional. Pathogenic missense variants in the same gene may act through different modes of action and result in distinct clinical conditions that require different treatments. The mechanisms of the functional impact depend on the functions of the encoded protein, therefore, predicting mode-of-action requires components that are tuned to the functions specific to each protein or protein families. Molecularly, the effect of a missense variant is about the change in biochemical properties of a protein, such as enzyme activity, stability, and protein-protein interactions. Genetically, one common categorization of the overall outcome of molecular effects is loss- and gain-of-function (LoF/GoF) variants. Few efforts have been made to study mode-of-action in scale because of lacking data and effective models. In this work, we tackled this challenge with two efforts. First, we curated the largest-to-date mode-of-action labeled missense variants datasets. The data includes 41,081 missense variants in 8 genes with multi-dimensional measurements of different biochemical properties by deep

mutational scan experiments, 2043 GoF and 7889 LoF missense variants in ~1300 genes. We found that the information useful for predicting pathogenicity can be useful for mode-of-action prediction but need additional training with gene or domain- specific data. Second, we developed a new method, PreMode, that learns a general representation of sequence variation within protein structural context using SE(3)-equivariant graph neural networks on AlphaFold2-predicted protein structures and protein language model (ESM2) embeddings. We first pre-train PreMode on large labeled pathogenic/benign variant databases, then performed transfer learning for mode-of-action predictions. For pathogenicity predictions, PreMode reached state-of-the-art performance comparable with AlphaMissense. For mode-of-action predictions, PreMode outperformed previously published methods, including LoGoFunc, FunCion, Augmented EVE and ESM1b. Additionally, We performed in silico mutagenesis using PreMode in a kinase and found its prediction of G/LoF variants is consistent with inactive-active conformation transition energy changes. Finally, we show that PreMode enables efficient study design of deep mutational scans and optimization in protein engineering.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6075T: Quantifying the effects of rare genomic variants affecting cell types on Autism spectrum disorders across time thanks to brain single-cell burden analysis

Authors:

T. Renne[1], A. Dubuc[2], C. Poulain[3], M. Jean-Louis[4], Z. Xie[4], V-R. Bourque[1], G. Huguet[5], T. Nowakowski[6], S. Jacquemont[7]; [1]Université de Montréal, Montreal, QC, Canada, [2]ENS Lyon, Lyon, France, [3]Montréal Univ., Montreal, QC, Canada, [4]CHU Sainte Justine, Montreal, QC, Canada, [5]CHU Ste justine, Montréal, QC, Canada, [6]UCSF, San Francisco, CA, [7]CHU Sainte justine Res. Ctr., Montreal, QC, Canada

Abstract:

Autism spectrum disorder (ASD) is a neurodevelopmental disorder diagnosed in 2.7% of the population. While the heritability of ASD is high, most of the genetic contribution remains unidentified and underlying mechanisms are still elusive. Gene-disrupting Copy Number Variants (CNVs) and Single Nucleotide Variants (SNVs) contributing to autism have mainly been identified through an excess of de novo events. These ASD-associated variants were found enriched in maturing neurons. However, such results may reflect biological functions tightly related to genetic constraint rather than ASD. Our understanding of rare gene-disrupting inherited variants and corresponding biological functions remains limited. In this study, we aim

to develop a single-cell functional burden association test to quantify the association of gene-disrupting variants with liability for ASD and characterize the cell types implicated across brain development. This method improves the statistical power by aggregating genes disrupted by one type of variant and assigned to one function.Using single-nuclei RNAseq data of prefrontal cortices collected across 6 developmental stages, from fetal to adult, we defined gene sets as the top decile of over-expressed genes in each cell type. We tested genetic association for SNVs and CNVs in 37k autistic individuals and 500k controls. The burden of variants disrupting specific gene sets was associated with the liability for ASD as well as autism-related traits using regression models. Among the 19 gene-sets assigned to cell types, 10 were significantly associated with ASD when disrupted by either SNVs or CNVs. Genes disrupted by SNVs and assigned to oligodendrocytes and maturing neurons increase liability for ASD. Among these SNVs, we found that the ASD risk is mostly associated with Frame shifts and stop gained variants in oligodendrocytes and maturing neurons, respectively. Analyses across developmental stages showed that most excitatory neurons subclasses have an impact on ASD during early childhood, while the effects of oligodendrocyte-associate genes were evenly distributed across brain development. Remarkably, we found no correlation between the effect sizes on ASD of CNVs and SNVs, suggesting that cell-type functions showed distinct sensitivities to classes of variants.This single-cell burden association study shows that, beyond known ASD genes, a broad spectrum of variants contributes to the liability for ASD. Cell type functions also differ between ASD liability and ASD-related traits liability. This whole study will help us detangle the biological pathways impacted by rare variants and leading to ASD.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6076T: Revealing disease mechanisms driven by BAF chromatin remodeling complexes using computational structural and functional genomics

Authors:

M. Zimmermann[1], N. Haque[1], J. Wagenknecht[1], Y-I. Chi[2], F. Palau[3], J. Hoenicka[3], A. Mathison[1], G. Lomberk[1], R. Urrutia[1]; [1]Med. Coll. of Wisconsin, Milwaukee, WI, [2]Genomic Sci. and Precision Med. Ctr., Wauwatosa, WI, [3]Sant Joan de Déu Children's Hosp., Barcelona, Spain

Abstract:

BAF is the SWI/SNF family chromatin remodeler that facilitates heterochromatin-euchromatin state interconversion. BAF sub-complex heterogeneity has been described and interconversion is necessary for human body development. We study the BAF complex, beginning with its

catalytic unit, SMARC4A. SMARCA4 is an ATPase, DNA-binding protein, histone reader, and more, with each function amenable to computation and laboratory study. SMARCA4 is frequently mutated in germline neurodevelopmental syndromes like Coffin-Siris (CSS), and cancer, with 2237 distinct missense variants observed. Yet, 97% are VUS. Most are rare or one-of-a-kind. Thus, new approaches are needed to enhance the information available to each patient. We take a proteogenomic approach and use computational tools to calculate how genetic variants alter their encoded protein structures, dynamics, and effects on the BAF complex. Using calculations derived from cryo-EM and protein structure, such as rigidity and flexibility analyses and active site rearrangements enable new means to score the damaging potential of genetic variations. Structural analysis complements traditional genomics scores and protein language models. Our goal is to interpret how mutations affect the enzyme functionally. Then, how SMARCA4 fits within its cognate chromatin remodeling complex, and how genetic variation in SMARCA4 may also change the complex. This scale in the number of mutations characterized and structural context is key for improving the interpretation of novel and ultra-rare genomic variants. The mechanistic readouts our approach generates, indicate specific testable hypotheses for functional genomics validation, including changes to protein stability, histone binding capacity, and helicase processivity. Our study is trailblazing and practical for enhancing the field. Similar approaches can be applied to help interpret human genetic variation underlying many congenital and malignant conditions, such as we demonstrate for CSS.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6077T: Saturation genome editing based functional evaluation and clinical classification of BRCA2 single nucleotide variants ⭐

Authors:

F. Couch[1], H. Huang[1], T. Pesaran[2], R. Karam[3], S. Yadav[1], D. M. Susan[4], M. N. Alvaro[5], N. Boddicker[1], W. Chen[1], M. E. Richardson[6], C. Hu[1]; [1]Mayo Clinic, Rochester, MN, [2]Ambry Genetics, New York, NY, [3]Ambry Genetics, Aliso Viejo, CA, [4]Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA, [5]H. Lee Moffitt Cancer Ctr., Tampa, FL, [6]Ambry Genetics, Phippsburg, ME

Abstract:

Germline *BRCA2* loss-of function (LOF) variants identified by clinical genetic testing predispose to breast, ovarian, prostate and pancreatic cancer. However, variants of uncertain significance (VUS) (n>5000) limit the clinical use of testing results. Thus, there is an urgent need for

functional characterization and clinical classification of all *BRCA2* variants. Here we report on comprehensive saturation genome editing-based functional characterization of 99% of all possible single nucleotide variants (SNVs) in the BRCA2 DNA Binding Domain hotspot for pathogenic missense variants that is encoded by exons 15 to 26. The assay was based on deep sequence analysis of HAP1 haploid cells endogenously targeted using a CRISPR/cas9 knockin approach. A total of 6959 SNVs were characterized for effects on cell survival. The assay was validated relative to nonsense and synonymous variants, ClinVar pathogenic/likely pathogenic and benign/likely benign variants and homology directed repair cell based DNA repair assay results, all of which showed >94% sensitivity and specificity. Variants were assigned posterior probabilities of pathogenicity using a VarCall two component Bayesian mixture model and were further grouped according to ACMG strength of evidence under the PS3/BS3 rule resulting in Benign Strong (n=5430), Benign Moderate (n=190), Benign Supporting (n=61), VUS (122), Pathogenic Strong (n=1021), Pathogenic Moderate (n=88), and Pathogenic Supporting (n=47). Breast cancer case-control association studies showed that pooled SNVs encoding functionally pathogenic missense variants were associated with increased risks of breast (odds ratio (OR)3.81, 95%CI: 2.88-5.07) and ovarian cancer (OR 5.93, 95%CI: 4.12-8.52). The functional data were also combined with other sources of information in the ClinGen BRCA1/2 VCEP ACMG/AMP-classification model. A total of 785 SNVs, including 261 missense SNVs, were classified as pathogenic or likely pathogenic, while 5566 SNVs, including 3786 missense SNVs, were classified as benign or likely benign. These classified variants can now be used for risk assessment and clinical care of variant carriers.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6078T: Sex-stratified Association of Variants in the Serotonin 1A Receptor Gene with Acute Crisis Pain among African American Sickle Cell Disease Patients

Authors:

Y. He[1], N. Sadhu[2], Y. Kashyap[2], Y. Yao[3], D. Wilkie[3], R. Molokie[2], **Z. Wang**[2]; [1]Midwestern Univ., Downes Grove, IL, [2]Univ. of Illinois Chicago, Chicago, IL, [3]Univ. of Florida, Gainesville, FL

Abstract:

Patients with sickle cell disease (SCD) suffer from pain in their daily lives. Both the acute and chronic pain phenotypes of this disease exhibit high variability, making pain management a challenge. The underlying reasons for the phenotypic variability are poorly understood. Since serotonergic neurotransmission is an integral part of pain signaling, we aimed to explore the

role of variants in the 5-HT1A receptor gene (*HTR1A*) on acute and chronic pain variability in SCD. Four variants (rs6449693, rs878567, rs6294, and rs10042486) in *HTR1A* were genotyped in a cohort of 131 self-reported African Americans with SCD. Acute and chronic pain were measured by the acute care utilization due to crisis and non-crisis pain scores from the McGill Pain Questionnaire, respectively. Association analyses were performed for three genetic models (additive, dominant, and recessive) and sex-stratified effects. We found three variants (rs6449693, rs6294, and rs10042486) in *HTR1A* showed significant association ($p<0.05$) with crisis pain in both the additive and dominant models. While the G allele of rs6449693 and the C allele of rs10042486 associated with lower acute crisis pain, the T allele of rs6294 associated with increased acute crisis pain. Sex-stratified analyses revealed that the associations of these three variants with acute pain were significant only in males, and not in females. Furthermore, the A allele rs878567 that did not reach statistical significance in the overall cohort, showed a significant association with lower crisis pain in males. As the first study to explore the role of *HTR1A* variants in sickle cell pain, we identified that four variants across the gene are associated with acute crisis pain in SCD in a sex-stratified manner.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6079T: Shift augmentation for improved indel scoring in DNA sequence-based machine learning models

Authors:

A. Korsakova, D. Srivastava, D. Kelley; Calico Life Sci., South San Francisco, CA

Abstract:

Insertions and deletions (indels) make up a significant proportion of all inherited variants. Along with single base substitutions (SBS), indels may affect transcriptional and epigenomic states of the cells. Deep learning models trained to predict a variety of biochemical assays, such as ATAC-seq, from the underlying DNA sequence have become important tools in deciphering the effects of such variants. Such models score the effect of the variant by comparing predictions for the reference and alternative alleles, enabling interpretation of its likely functional consequences. Recent large-scale models Enformer and Borzoi have demonstrated state-of-the-art performance identifying causal eQTL variants. However, most analyses so far have focused on SBS, and few have considered indels due to technical challenges.
We use the Borzoi model to demonstrate that indels introduce technical variance to predictions due to shifted boundaries of the convolutional network blocks. Upon introducing an alternative indel allele, an input sequence shift occurs on the right or left side from the indel. This, in turn,

interferes with scoring metrics used to evaluate variant effects. Metrics that compare predictions between output bins, as compared to simpler aggregations like summing, are especially sensitive to these shifts. We propose strategies to alleviate this technical variance and bring the predicted score distribution of indels closer to that of SBS. Additionally, we show that proposed techniques are effective for sequential track predictors and peak predictors. Evaluated on indel eQTLs, indel shift augmentation strategies boost the performance of a classifier trained to distinguish between causal and non-causal variants that are associated with expression changes. Therefore, these strategies may be used to expand the applications of deep learning models to score and prioritize regulatory function-modifying indels in QTL and GWAS studies.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6080T: Single-nuclear profiling of chromatin accessibility and gene expression in human liver reveals genetic effects on cell-type gene regulation

Authors:

A. Alkhawaja[1], K. W. Currin[1], H. J. Perrin[1], S. Vadlamudi[1], A. S. Etheridge[1,2], G. H. Cannon[3], C. W. Anderson[3], A. H. Moxley[1], E. G. Schuetz[4], F. Innocenti[2,5], T. S. Furey[1], K. L. Mohlke[1]; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [2]Eshelman Sch. of Pharmacy, Univ. of North Carolina, Chapel Hill, NC, [3]Advanced Analytics Core, Univ. of North Carolina, Chapel Hill, NC, [4]Dept. of Pharmaceutical Sci., St. Jude Children's Res. Hosp., Memphis, TN, [5]Lineberger Comprehensive Cancer Ctr., Univ. of North Carolina, Chapel Hill, NC

Abstract:

The cell types in which GWAS variants function remain largely unknown. While bulk tissue studies identified variant effects on gene expression and chromatin accessibility (e/caQTL) that colocalize with GWAS signals, they do not pinpoint the cell type of action, and variant effects in less abundant cell types may have been missed. We jointly profiled single-nucleus gene expression and chromatin accessibility on 40 human livers and compared results to large bulk liver studies. We identified 68,393 nuclei and detected expression for 17,607 genes, of which 966 were not detected in the most common cell type, hepatocytes. While gene expression in liver cell types is well characterized, cell-type chromatin accessibility is not well defined. We detected accessibility for 306,706 chromatin regions (peaks), of which 105,369 were not detected in hepatocytes. GWAS heritability enrichment linked peaks to cell-type functions including liver enzymes in hepatocytes and cholangiocytes, and blood pressure in mesenchymal cells. We defined 17,147 marker peaks that had higher accessibility in one cell type compared to

every other; these peaks were near genes enriched for relevant cell-type processes (lipid homeostasis in hepatocytes, extracellular matrix organization in mesenchymal cells), and transcription factor motifs in these peaks aligned with cell identities (HNF4A in hepatocytes, ZEB in cholangiocytes). Matching cell-type peaks to bulk liver ATAC-seq peaks from 138 individuals identified 70,884 peaks missed in the bulk analysis, most (51,717) detected only in non-hepatocytes. We next predicted cell types for 35,361 bulk liver caQTL; we mapped 3,112 to a single non-hepatocyte cell type and predicted that 715 of them regulate 686 genes. In the cell-type data, we identified 67 eQTL and 1,885 caQTL, including 12 eQTL and 5 caQTL detected only in non-hepatocytes. For instance, at a GWAS signal for plasma levels of gamma-glutamyl transferase where a bulk eQTL for *CD22* and a bulk caQTL were detected, we observed an eQTL in liver sinusoidal endothelial cells (LSEC) for which the lead (rs10411704) is within a peak in LSEC. Second, at a GWAS signal for blood cell traits and a bulk eQTL for *ITGAD*, we observed an eQTL in Kupffer cells for which the lead (rs8050500) is within a peak detected only in Kupffer cells. Third, at a GWAS signal for glucuronide of piperine, implicated in elevated blood pressure, we observed a caQTL in LSEC. Overall, this study provides a high-resolution map of chromatin accessible regions in liver cell types, shows that single-nucleus profiling captures peaks and genes missed in bulk, and offers insights into genetic mechanisms underlying cell-type gene regulation.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6081T: Small but potentially clinically-relevant contribution of CNVs to idiopathic nephrotic syndrome

Authors:

Y. Gupta[1,2], T. Lim[1,3], J. Ke[1], A. Mitrotti[4], G. Jin[1], M. Verbitsky[1], I. Pisani[5], M. Bodria[6], K. Mucha[7], J. Reiterova[8], M. Saraga[9], L. Zibar[10,11], D. Milosevic[12], V. Tasic[13], A. Simoes e Silva[14], W. Morello[15], G. Montini[16,17], Y. Caliskan[18], U. Maggiore[19], E. Fiaccadori[5], L. Gesualdo[20], G. Ghiggeri[6], M. Kretzler[21], M. Saleem[22], M. G. Sampson[23], A. Gharavi[1], M. Pollak[24], F. Hildebrandt[25], S. Sanna-Cherchi[1]; [1]Columbia Univ., New York, NY, [2]Univ. of Lübeck, Lübeck, Germany, [3]Univ. of Turin, Turin, Italy, [4]Univ. of Bari, Bari, Italy, [5]Univ. of Parma, Parma, Italy, [6]IRCCS Inst. Giannina Gaslini, Genoa, Italy, [7]Med. Univ. of Warsaw, Warsaw, Poland, [8]Charles Univ., Prague, Czech Republic, [9]Univ. of Split, Split, Croatia, [10]Univ. Hosp. Merkur, Zagreb, Croatia, [11]Univ. Josip Juraj Strossmayer in Osijek, Osijek, Croatia, [12]Univ. of Zagreb, Zagreb, Croatia, [13]Univ. Children's Hosp., Skopje, North Macedonia, The Republic of, [14]Federal Univ. of Minas Gerais, Belo Horizonte, Brazil, [15]Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico di Milano, Milan, Italy, [16]Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico di Milano, Milan,

Italy, [17]Univ. of Milan, Milan, Italy, [18]Saint Louis Univ., Saint Louis, MO, [19]Univ. Hosp. of Parma, Parma, Italy, [20]Univ. of Bari Aldo Moro, Bari, Italy, [21]Univ. of Michigan, Ann Arbor, MI, [22]Univ. of Bristol, Bristol, United Kingdom, [23]Boston Children's Hosp., Boston, MA, [24]Beth Israel Deaconess, Boston, MA, [25]Boston Children's Hosp., Boston, MA

Abstract:

Background: Idiopathic Nephrotic Syndrome (INS) is a heterogenous trait characterized by heavy protein loss in the urine, low serum albumin, edema, and is a frequent cause of kidney failure. The utility of genetic diagnostics for rare variants in Mendelian INS-associated genes has been reported in patients with various kidney diseases. However, genetic studies have overwhelmingly focused on the contribution of single nucleotide variants, leaving the role of rare Copy Number Variations (CNV) in INS largely unexplored. **Methods:** We conducted DNA microarray (CMA) genotyping on a multi-ancestry cohort of 3,133 INS patients. The cases were partitioned into 5 categories: Steroid Resistant Pediatric (SRP, n = 747), Steroid Resistant Adult (SRA, n = 1,017), Steroid Sensitive Pediatric (SSP, n =705), Steroid Sensitive Adult (SSA, n = 450), and Other (n = 214), comprised cases with an unknown age of disease onset or unknown initial response to immunotherapy. We called CNVs using PennCNV. Variants were annotated against the CNV start and stop boundaries corresponding to 221 known Genomic Disorders CNVs (GD-CNVs) and a curated list of 126 genes that, when mutated, are known to cause INS or a phenocopy of it. **Results:** Our analysis revealed that 3.03% of INS cases (95 of 3,133) carry candidate likely pathogenic CNVs, 51 duplications and 44 deletions. 0.86% of INS cases (27 of 3,133) carry a known pathogenic genomic disorder GD-CNVs. However, there was no significant enrichment of GD-CNV in INS cases when compared against a reference population control (134 of 21,498, 0.68%) suggesting no overall significant contribution at population level, and only a potential role at individual level. While the majority of these GD-CNVs likely represent the background mutational rate in the general population, other GD-CNVs such as the chr17q12 RCAD deletion and chr2q13 homozygous deletion in *NPHP1* are of direct diagnostic value. 0.7% (22 of 3,133) individuals carry candidate CNVs that encompassed INS-associated genes such as *INF2* (12), *APOL1* (5), *NPHS1*, *COPA*, *CFH*, *CEP83*, *ARHGAP24* (17 SR, 5 SS). Lastly, 1.47% of INS cases (46 of 3,133) carry large CNVs (>500Kb) that were either rare (<0.005%) or absent in the reference control population. There was no significant difference of CNV burden across INS phenotypic categories. **Conclusion:** In this large multiethnic study, we characterized the rare CNV landscape of INS. The CNV diagnostic rate for INS was relatively low, (3%), but relevant since these variants add to the overall genetic diagnostic workup and have direct implication in clinical management for INS and risk stratification for extra-urinary complications associated to these variants.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6082T: Splice rescue of protein-truncating variants is associated with reduced selection pressure and attenuated disease phenotype

Authors:

X. Cheng[1], F. Liu[1], **J. Yang**[1], Y. Gai[1], L. Zhao[1], W. Wen[1], Q. Li[1], G. Li[1], K. Xu[1], Z. Zhao[1], H. Zhao[1], Y. Niu[1], N. Chen[1], Z. Wu[1], J. Zhang[1], S. Zhao[2], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Baylor Coll. of Med., Houston, TX

Abstract:

The deleterious effects of putative loss-of-function variants could be masked by various genetic compensation mechanisms, presenting challenges in variant interpretation and the understanding of genotype-phenotype correlations. This study explores a splice rescue mechanism where protein-truncating variants (PTVs) introduce a cryptic splice site, leading to an in-frame splice-out of the variant and the preservation of protein function. We retrieved all nonsense and frameshift variants (n=346,433) from 125,748 exome sequencing (ES) and 15,708 genome sequencing (GS) data in the gnomAD database (v2.1.1) and predicted their splicing effects using SpliceAI. We found that PTVs associated with splice rescue constituted approximately 0.5% of all PTVs and 20.1% of PTVs anticipated to affect splicing. These variants were associated with reduced selection pressure, evidenced by higher population maximum allele frequency ($p=0.043$) and lower proportion of singleton ($p<0.001$) compared to those variants predicted to cause out-of-frame splicing changes (not rescued). Among previously reported PTVs, we highlighted a nonsense variant in the *APC* gene (NM_000038.6: c.288T>A[p.Tyr96Ter]), which is associated with attenuated familial adenomatous polyposis (AFAP) and is predicted to undergo splice rescue. A minigene assay validated the splice rescue transcript, and the application of antisense oligonucleotides (ASO) was shown to promote the rescued transcript levels in the *APC* mutant cell line. Further *in vitro* experiments revealed that *APC* mutant cells exhibited similar biological functions to those of wild-type cells. Notably, we observed that β-catenin levels remained stable despite a reduction in APC protein levels, suggesting a threshold effect of APC protein on downstream β-catenin in human cells. These observations enhance our understanding of splice rescue in modulating disease severity. Although rare, splice rescue variants warrant careful interpretation, particularly in avoiding overdiagnosis in patients with mild clinical presentations. For diseases linked to the splice rescue mechanism, our study also underscores the feasibility of developing novel therapeutic approaches, such as ASO inhibition or pathway targeting, thereby advancing precision medicine and the effective management of genetic disorders.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6083T: Systematic targeting and inhibition of bone mineral density GWAS loci in single cells identifies novel regulatory networks for human bone formation ⭐

Authors:

J. Morris[1], L. D. Mesner[2], E. A. Farber[2], A. C. Mello[3], S. Ghatan[3], T. Lappalainen[3], N. E. Sanjana[3], C. R. Farber[2]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Univ. of Virginia, Charlottesville, VA, [3]New York Genome Ctr., New York, NY

Abstract:

Bone mineral density (BMD) is a heritable, causal risk factor for osteoporosis, a debilitating age-related disease. Genome-wide association studies (GWAS) of BMD have identified over 500 loci (20% heritability explained) and known osteoporosis drug targets. However, 90% of BMD GWAS variants are noncoding, with unknown regulatory effects in *cis* and *trans*, making identifying causal genes a significant challenge. We previously developed an approach to discover target genes for noncoding loci at high-throughput, Systematic Targeting and Inhibition of Noncoding GWAS loci with single-cell sequencing (STING-seq). STING-seq leverages biobank-scale GWAS data, massively parallel CRISPR screens, and single-cell sequencing to identify target genes in *cis* and *trans*-regulatory networks. *Trans*-networks were often mediated by transcription factors (TFs) and enriched for polygenic contributions to complex traits. Here, we applied STING-seq to BMD GWAS through the lens of human bone formation. We focused on the fetal human osteoblast (hFOB) model to identify noncoding loci that influence genes and *trans*-networks across osteogenic differentiation.

We targeted identifying *trans*-networks by reanalyzing our previous BMD GWAS that identified 518 loci from 426,824 UK Biobank participants. We performed finemapping and integrated ATAC-seq data from hFOBs across early osteogenic differentiation (days 0, 2, and 4 post-differentiation). We targeted 361 candidate *cis*-regulatory elements (CREs) harboring fine-mapped noncoding variants (138 loci), where at least one of the three closest genes were TFs, with CRISPRi. Additionally, we targeted 27 TFs with coding BMD GWAS variants. We performed a pilot study in hFOBs at differentiation day 0 and identified target genes for 17% of the 138 loci, identifying 28 genes in total (12 TFs) and multiple *trans*-networks. Three were from direct TF perturbation (*ING3*, *UBP1*, *ZBTB4*) and one from an intronic GWAS-CRE (rs1734116) for *CHD4*. *CHD4* is a chromatin remodeler and loss in humans results in neurodevelopmental disorders that present with skeletal abnormalities, suggesting that these abnormalities may be mediated by the *CHD4 trans*-network in osteoblasts.

In summary, we performed a pilot BMD STING-seq screen in undifferentiated hFOBs, discovered

target genes at 17% of targeted noncoding GWAS loci, and identified a *CHD4 trans*-network in osteoblasts. This enables the full BMD STING-seq screen we are performing across early osteogenic differentiation in hFOBs to learn if target genes and *trans*-networks are impacted by changes in cell state or if additional genes can be discovered.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6084T: Telomere length in Colombian Centenarians (COOLCEN cohort)

Authors:

O. Gomez[1,2], J. E. Gallo[3,4], I. P. Torres[3], I. D. Lozada-Martinez[5,6], J. C. Salazar[7], J. M. Anaya[5]; [1]Soluciones Moleculares SAS, Medellin, Colombia, [2]Sch. of Med., Univ. de Antioquia, Medellin, Colombia, [3]Med. Sci. Res. group, Sch. of life Sci.s and Med., Univ. EIA, Medellin, Colombia, [4]ASRT Inc., Atlanta, GA, [5]Hlth.Res. and Innovation Ctr. at Coosalud EPS, Cartagena, Colombia, [6]Univ. de la Costa, Barranquilla, Colombia, [7]Univ. Natl. de Colombia, Medellin, Colombia

Abstract:

Background: Telomere length has been associated with aging and overall health. Centenarians, individuals who live to or beyond 100 years, provide a unique model for studying the mechanisms underlying extreme longevity. One of the key factors implicated in aging and longevity is telomere length. The COOLCEN cohort is currently comprised of centenarians in Colombia, South America. To date, this cohort is the only of admixed Latino individuals that has been characterized with an extensive phenotypic questionnaire of more than 300 clinically related parameters. This study aims to analyze leukocyte telomere length (LTL) in centenarians and investigate its correlation with quality of life (QoL) and health status, specifically focusing on dependency, based on the Barthel Index (BI) score.**Methodology:** The study analyzed a sample of 116 centenarians from the COOLCEN cohort. LTL was measured using monochromatic multiplex qPCR (MMQPCR) to estimate absolute leukocyte telomere length (aLTL). Statistical analyses were conducted to examine the correlation between aLTL and clinical variables, including quality of life (QoL), and dependency (BI < 80 vs. BI > 90). **Results:** The median of the aLTL results was 9.5 kb (IQR: 5.1), with a minimum and maximum of 3.0 kb and 19.2 kb, respectively. A significant correlation between aLTL and QoL was observed (r_s = 0.663, p = 0.001) in male centenarians. Significant difference in aLTL was observed between dependent (n=77, 9.55 ± 3.37) and independent centenarians (n=20, 12.61 ± 4.42, p = 0.0087). Furthermore, aLTL varied among geographic regions from Colombia (p=0.0062).**Conclusion:** Telomere length in

Colombian centenarians is associated with dependency and QoL and varies across different geographic regions (ancestry). The telomere length in the COOLCEN cohort is higher than that observed in previous studies in centenarians from other populations. These findings underscore the potential role of aLTL as a biomarker for health and longevity in centenarians.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6085T: Tissue specific modifiers of stenosis in Williams-Beuren syndrome

Authors:

D. Liu[1], A. Globus[1], N. Raja[2], M. Levin[2], E. Biamino[3], M. Bedeschi[4], M. Digilio[5], G. Squeo[6], R. Villa[4], S. Osgood[2], J. Freeman[1], G. Merla[7], B. Pober[8], C. Mervis[9], A. Roberts[10], C. Morris[11], L. Osborne[12], B. Kozel[13]; [1]Natl. Heart, Lung, and Blood Inst., Bethesda, MD, [2]NHLBI, Bethesda, MD, [3]Univ. of Turin, Turin, Italy, [4]Fondazione IRCCS Ca Granda Ospedale Maggiore Policlinico Med. Genetic Unit, Milan, Italy, [5]Bambino Gesù Children's Hosp., Rome, Italy, [6]Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, [7]IRCCS Casa Sollievo Della Soff, San Giovanni Ro, Italy, [8]Mass Gen Children's Hosp, Boston, MA, [9]Univ. of Louisville, Louisville, MD, [10]BCH, Boston, MA, [11]Univ Nevada Sch Med, Las Vegas, NV, [12]Univ Toronto, Toronto, ON, Canada, [13]NIH, Bethesda, MD

Abstract:

Introduction: Stenoses of the supravalvar aorta (SVAS) and pulmonary arteries (PAS) are common manifestations of Williams-Beuren syndrome (WS). The presence, progression, and severity of stenosis at these sites vary. **Methods:** Among 473 people with WS evaluated, 334 had sufficient data to extract details on trajectory and severity of both their SVAS and PAS. Phenotypic outcomes were calculated and then, using extreme phenotyping and a non-synonymous variant prioritization strategy, we identified enriched variants in 3 comparisons: (1) those with bilateral outflow tract repair (n=29), (2) those requiring SVAS but not PAS repair (n=46) and (3) those requiring repair for PAS but not SVAS (n=26), all *vs* the no stenosis group (n=57). Prioritized variants were used for pathway enrichment. **Results:** This study confirmed that 69% of PAS resolves by mid-childhood. Only 3.7% of infants with initially mild PAS require surgery later. Children requiring bilateral outflow tract repair were at higher risk of other cardiovascular interventions ($p<0.01$) than those with SVAS or PAS only. That said, the severity scores for SVAS and PAS were non-concordant 46% of the time, suggesting tissue specific modifiers. Modifier studies showed enrichment for extracellular matrix (ECM), sensory, and immune pathway genes in all 3 stenosis groups, however the specific ECM genes and immune

sub-pathways were vascular location specific. Variants in other pathways, like PIK3-AKT and metabolism of lipids, were identified in only one locational analysis. When evaluated at the variant level, some variants (e.g. rs267561 in IGTA9) were more frequent (range 9-21% more) in surgical patients than in those with no stenosis. The common coding *ELN* variant rs2071307 showed increased frequency in people requiring only SVAS repair (62%), while those with the other stenosis outcomes exhibited variant frequencies similar to the general population (39%). **Conclusion:** This study describes the interplay between SVAS and PAS in people with WS. We show that people requiring bilateral outflow tract repair are at increased risk of additional cardiovascular procedures, but that right and left sided lesions are uncorrelated in severity 46% of the time. Modifier studies show that variation within ECM pathway and immune genes is enriched in people with extreme stenosis outcomes, but that the enriched genes within those pathways vary by vascular location. Of interest is the finding that *ELN* variant rs2071307 is more frequent in people with SVAS. This finding suggests a paradigm where seemingly benign structural changes in the elastin monomer may influence mechanical properties of resulting fibers, impacting phenotype.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6088T: Unproductive splicing QTLs implicate post-transcriptional regulatory variants in Alzheimer's Disease

Authors:

C. Dai[1], R. Feng[2], G. Wang[2], Y. Li[1]; [1]Univ. of Chicago, Chicago, IL, [2]Columbia Univ., New York, NY

Abstract:

Genetic variants associated with Alzheimer's disease (AD) are often associated with inter-individual differences in alternative splicing (AS) (Raj et al., 2019). However, it remains unclear exactly how differences in splicing impacts gene function in AD. We recently discovered that a huge number of AS events generate unproductive transcripts, i.e. transcripts with a premature termination codon (Fair et al., 2023). As unproductive transcripts are rapidly degraded by nonsense-mediated decay, the coupling between alternative splicing and nonsense-mediated decay (AS-NMD) represent a critical post-transcriptional regulatory pathway that impacts mRNA expression levels (Fair et al., 2023). Thus, we hypothesized that many AD-associated variants may increase unproductive splicing at risk genes, leading to dysregulation of gene expression levels. Here, we performed the largest analysis to date of AS-NMD in the GTEx and ROSMAP datasets to study the impact of genetic variants on unproductive splicing, and downstream effects on gene expression levels and disease risk. These data represent the largest dataset of

prefrontal cortex RNA-seq samples from AD patients (n = 322) and non-AD controls (n = 526) to date. To study unproductive splicing in these samples, we developed LeafCutter2, extending many functionality of LeafCutter, a method that we previously designed for annotation-free analysis of RNA splicing. We show that LeafCutter2 effectively distinguishes productive from unproductive splicing events, and allows us to directly map the impact of AS on gene expression levels. Notably, we find that differences in unproductive splicing across tissues were globally anti-correlated with differences in gene expression levels, supporting a pervasive role of unproductive splicing on tissue-specific regulation of gene expression levels. We used LeafCutter2 to map 185,065 u-sQTLs (unproductive splicing QTLs) across our datasets, and found that u-sQTLs showed strong enrichment in eQTLs, and GWAS loci. Statistical integration of eQTLs, histone QTLs, and sQTLs along with AD GWAS summary statistics, pinpointed 17 AD risk loci that impact unproductive splicing. Among these, we highlight two high-confidence loci that function through AS-NMD (TSPAN14 and CASS4). Both risk loci colocalize with an unproductive sQTL and an eQTL. TSPAN14 is a known AD risk gene with potential of pathological splicing variations, while CASS4 is a promising novel AD gene. In conclusion, we demonstrate that AS-NMD is a prevalent mechanism linking AD risk loci to gene expression, and present LeafCutter2 as an effective method to study AS-NMD from short-read RNA-seq data.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6089T: Unraveling Rare Diseases: The NIH Undiagnosed Diseases Program's Integrated Approach to Genetic Diagnosis and Treatment

Authors:

M. Malicdan[1], M. Morimoto[2], E. Burke[1], E. Macnamara[3], B. Pusey[4], J. Fu[5], Y. Huang[3], V. Maduro[6], L. Wolfe[7], P. D'Souza[1], D. Novacic[3], C. Tifft[3], C. Toro[1], D. Adams[1], W. Gahl[8]; [1]NIH, Bethesda, MD, [2]NIH, SILVER SPRING, MD, [3]NIH, Bethesda, MD, [4]Natl Human Genome Res. Inst., Bethesda, MD, [5]NHGRI, Rockville, MD, [6]NIH, Rockville, MD, [7]NIH - NHGRI, Bethesda, MD, [8]NHGRI (NIH), Bethesda, MD

Abstract:

The National Institutes of Health's Undiagnosed Diseases Program (NIHUDP) is a pioneering initiative dedicated to solving the mysteries of rare and undiagnosed diseases through advanced biomedical research and clinical practice. This program exemplifies the power of translational research, bridging the gap between scientific discoveries and patient care. The NIHUDP aims to diagnose and find potential treatments for patients with prolonged diagnostic journeys. This work outlines the NIHUDP's integrated research strategy, designed to systematically investigate

the genetic underpinnings of disease. Central to our methodology is the exhaustive use of computational tools, including genetic database information, variant prediction software, and algorithms that assess variant pathogenicity. Following computational studies and extensive feasibility evaluation, selected variants undergo rigorous experimental testing, including in vitro assays using patient-derived cells or engineered cell lines to study protein activity, and in vivo studies in model organisms (e.g., mice, drosophila, worm, and zebrafish) to observe phenotypic outcomes. These experiments confirm the functional significance of variants, linking genetic findings to disease manifestation. Recognizing the interdisciplinary nature of diagnosing undiagnosed diseases, this framework fosters collaboration among NIHUDP, geneticists, clinicians, bioinformaticians, molecular and cell biologists, and other specialists. This collaborative environment ensures that findings are interpreted within the broader context of patient symptoms and disease mechanisms, facilitating the translation of research into clinical insights. A key facet of this approach is the continuous reevaluation of undiagnosed cases. As new genetic data emerge and technologies advance, previously uninterpretable findings can be revisited, allowing for iterative refinement of diagnoses and exploration of novel therapeutic avenues, accelerating the journey from research hypothesis to bedside application. Our efforts have led to notable discoveries in novel gene-disease associations (~30 genes) and expansion of disease phenotypes, providing critical insights into rare disease pathology and offering diagnoses to over 400 patients since 2008. These successes highlight the potential of our framework to significantly impact the field of rare and undiagnosed diseases. The challenges and unexpected outcomes we've encountered have been invaluable in refining our methodologies and driving innovation within the NIHUDP, emphasizing the iterative and adaptive nature of translational research.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6090T: Unraveling the role of WNT2B in ocular development

Authors:

A. Verma[1], L. Bien[1], L. Al-Abdi[2], F. Alkuraya[3], A. Slavotinek[1]; [1]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [2]Translational Genomics (Ctr. for Genomic Med.), King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [3]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia

Abstract:

Wnt signaling is an evolutionarily conserved pathway that is critically important for several developmental processes, including cell fate determination, cell migration, cell polarity, neural patterning, and organogenesis. Pathogenic *WNT2B* variants have previously been associated

with Congenital Diarrheas and Enteropathies (CODE) in humans. Several patients with CODE have demonstrated ocular findings, including coloboma and microcornea. We hypothesize that *WNT2B* plays a role in eye development independent of its role in intestinal homeostasis in CODE. We identified two patients with eye defects and novel *WNT2B* variants who had no known history of congenital diarrhea. A patient with a microphthalmia, anophthalmia and coloboma (MAC) phenotype was heterozygous for a missense variant in *WNT2B*(NM_024494.3):c.806A>G;p.(Tyr269Cys). Another patient with colobomatous microphthalmia was homozygous for *WNT2B*(NM_024494.3):c.757C>G;p.(Arg253Gly). We explored zebrafish as a model to gain insight into the role of *WNT2B* in eye development. At the protein level, zebrafish wnt2bb shares 78% identity with human WNT2B and putative pathogenic residues are conserved in zebrafish. We used CRISPR/Cas9 to generate three mutant alleles in *wnt2bb* and created a zebrafish model to investigate loss of function (LOF) for *wnt2bb*. We observed a mild eye phenotype, with microphthalmia in F1 embryos that were compound heterozygotes for LOF variants in *wnt2bb*. We also injected wildtype *wnt2bb* mRNA to examine whether there are ocular phenotypes resulting from gain of function for this gene. Injections of wildtype mRNA showed normal eyes in 49/85 (57.6%) surviving, non-deformed larvae, anophthalmia in 17 (20%) and microphthalmia in 19 (22.4%) larvae. The data from these two patients with *WNT2B* missense variants exhibiting a MAC phenotype, along with the ocular phenotypes observed in *wnt2bb* loss of function larvae further strengthen the association between *WNT2B* and eye development.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6091T: Utilizing large-scale whole genome sequencing data from the All of Us research program to discover novel genes and endogenous retroviral loci associated with female infertility and recurrent pregnancy loss

Authors:

E. Stricker, R. Zemet, I. B. Van Den Veyver, E. G. Atkinson; Baylor Coll. of Med., Houston, TX

Abstract:

Approximately 12% of reproductive-aged women struggle with failure to establish a clinical pregnancy with 2.5% of women trying to conceive experiencing recurrent pregnancy loss (RPL). Of these couples, 15% remain affected by "unexplained infertility." Our study aimed to identify novel maternal genes and endogenous retroviruses (ERVs) associated with infertility and RPL

through a comprehensive analysis of large-scale genome sequencing data from the All of Us Research Program.
We studied two groups of women enrolled in the All of Us Research Program, one with female infertility without children or clinical pregnancy (n = 1,840) and the other with RPL (n = 652). Each group was compared to corresponding controls, women with child or pregnancy history but no infertility (n = 38,829) or miscarriage history (n = 31,616) via rare variant burden analyses and common variant genome-wide association studies (GWAS). For the burden analysis, rare variants were aggregated and counted for each protein-coding human gene (18,006) and ERV (17,559), while GWAS evaluated each variant individually. Like protein-coding regions, ERV internal regions comprise 1.5% of the genome and therefore serve as added entities with coding potential. A catalog of ERV internal regions was obtained from the HERVd and the HERVd Atlas database. For both analyses, we restricted the genomic data to quality-controlled biallelic variants and included covariates accounting for the principal components. The analysis was performed separately for genetically inferred populations: African/African American, American Admixed/Latino, and European. Finally, statistical significance was defined using a Bonferroni-corrected p-value threshold to account for multiple statistical tests. Genes associated with infertility or RPL were assessed through functional annotation using gene ontology and gene set enrichment analysis.
The analysis revealed two genes (*MGMT*, *OR6N2*) and four ERVs (ERVL-B4_1p31.1, HERV16_8q24.11, ERVL-E_21q21.1, MSTB-int_5q31.1) that are significantly enriched for rare variants compared to controls. In particular, MSTB-int_5q31.1, which was enriched in rare variants in RPL for women with European ancestry, appeared to have relevant homeostatic functions, as has been shown to be expressed in various tissues and downregulated in cancers. Therefore, these findings offer new insights into the genetic factors contributing to reproductive challenges and pave the way for further research into their role in infertility and RPL using large-scale population genome data. The involvement of ERVs underscores the potential role of retroviral elements in reproductive pathology.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6092T: Variability of variant pathogenicity prediction accuracy across genes associated with orofacial clefts

Authors:

P. Wang[1], E. C. Liao[1,2], K. Wang[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

OBJECTIVE: The Orofacial Clefts (OFCs) are common human birth anomalies with strong genetic basis leading to frequent clinical testing. Accurate prediction of variant pathogenicity is crucial to understand pathophysiological mechanisms and inform clinical decisions. This study aims to build a variant catalog for OFC-associated genes and evaluate the accuracy of current bioinformatics tools to predict pathogenicity. **METHODS**: The list of genes related to OFCs are collected from various database and recent OFC genetic studies. Variants were collected from ClinVar (20240514) and gnomAD (v4.0). Pathogenicity prediction from various bioinformatics tools was performed using ANNOVAR with the dbnsfp47a annotation. **RESULTS**: Of all initial genes related to OFCs, 143 genes were selected based on UK PanelApp Green Panel and recent studies. After filtering out unconfident and conflicting variants, there were total of 1881 benign variants (BV), 3062 likely benign variants (LBV), 681 pathogenic variants (PV), and 1367 likely pathogenic variants (LPV) from 100 genes. Prediction from 23 categorical tools showed an overall lower accuracy for BV (0.662) and LBV (0.599), compared to PV (0.832) and LPV (0.812). This was further supported by the low accuracy (0.555) for all 22333 gnomAD common (assumed benign) variants. Among all OFC-associated genes, *WNT5A* showed the lowest average (across all tools) accuracy (0.420) for both BV and PV, followed by *COL11A2* (0.438) and *FLNA* (0.469). The best accuracies were for genes *SIX1* (1.00) and *PAX3* (0.992). The most prevalent genetic drivers of OFC are *IRF6* and *CTNND1*, where the accuracy was 0.630 and 0.587, respectively. Pathogenicity score from 54 quantitative tools were also examined, achieving an average ROC AUC score of 0.837 for all ClinVar BV and PV. *COL11A2* received a lowest AUC (0.639), followed by *EYA1* (0.654) and *CDH1* (0.663). Constrained genes showed higher accuracy (0.034 more) for PV but lower accuracy (0.044 less) for BV compared to unconstrained genes. Genes with experimental structures showed higher accuracy (0.072 more) for PV but similar accuracy for BV than genes without experimental structures. **CONCLUSION**: Pathogenicity predictions from bioinformatics tools showed relative low accuracy in OFC-related genes, especially in over-predicting BV as PV. There is substantial variability in prediction accuracy among different genes, suggesting the importance of gene-specific pathogenicity interpretation of variants. This study raises concerns about the clinical applicability of current predictions and underscore the important need for improved computational methods or experimental validations.

Session Title: Molecular Effects of Genetic Variation Poster Thursday Session

Board 6093T: VISTA: An integrated framework for structural variant discovery

Authors:

V. Sarwal[1], S. Lee[1], J. Yang[2], M. Chaisson[2], S. Sankararaman[1], S. Mangul[3]; [1]UCLA, Los Angeles, CA, [2]Univ. of Southern California, Los Angeles, CA, [3]USC, Los Angeles, CA

Abstract:

Structural variation (SV), refers to insertions, deletions, inversions, and duplications in human genomes. SVs are present in approximately 1.5% of the human genome, but this small subset of genetic variation has been implicated in the pathogenesis of psoriasis, Crohn's disease and other autoimmune disorders. Specialized computational methods often referred to as SV callers are capable of detecting structural variants directly from sequencing data. With advances in whole genome sequencing (WGS) technologies, a plethora of SV detection methods have been developed. However, dissecting SVs from WGS data remains a challenge, with the majority of SV detection methods prone to a high false-positive rate, and no existing method able to precisely detect a full range of SVs present in a sample. Previous studies have shown that none of the existing SV callers can maintain high accuracy across various SV lengths and genomic coverages. Here, we report an integrated structural variant calling framework, VISTA (Variant Identification and Structural Variant Analysis) that leverages the results of individual callers using a novel and robust filtering and merging algorithm. In contrast to existing consensus-based tools which ignore the length and coverage, VISTA overcomes this limitation by executing various combinations of top-performing callers based on variant length and genomic coverage to generate SV events with high accuracy. We evaluated the performance of VISTA on comprehensive gold-standard datasets across varying organisms and coverage. We benchmarked VISTA using the Genome-in-a-Bottle (GIAB) gold standard SV set, haplotype-resolved de novo assemblies from The Human Pangenome Reference Consortium (HPRC), along with an in-house PCR-validated mouse gold standard set. VISTA maintained the highest F1 score among top single as well as consensus-based tools measured using a comprehensive gold standard across both mouse and human genomes. VISTA also has an optimized mode, where the calls can be optimized for precision or recall. VISTA-optimized can attain 100% precision and the highest sensitivity among other variant callers. VISTA represents a significant advancement in structural variant calling, offering a robust and accurate framework that outperforms existing consensus-based tools and sets a new standard for SV detection in genomic research.

Session Title: Omics Technologies Poster Thursday Session

Board 1100T: A comparison between traditionally used variant detection techniques and long-read whole genome sequencing in a clinical setting

Authors:

F. Eiengård[1,2], A. Rohlin[3,4], P. Palmebäck[5], M. Nordling[6,7]; [1]Dept of Lab. Med., Univ. of Gothenburg, Gothenburg, Sweden, [2]Dept of Lab. of Precision Med., Linköping Univ. Hosp., Linköping, Sweden, [3]Dept of Lab. Med.,, Univ. of Gothenburg, Gothenburg, Sweden, [4]Dept of Clinical Genetics and Genomics, Sahlgrenska Univ. Hosp., Gothenburg, Sweden, [5]Dept of Clinical Genetics, Linköping Univ. Hosp., Linköping, Sweden, [6]Dept of Clinical Genetics, Linkoping Univ. Hosp., Linkoping, Sweden, [7]Dept of Cell and Neurobiology, Univ. of Linköping, Linköping, Sweden

Abstract:

Introduction: Presently, techniques used for detection of structural and copy-number variations in clinical settings are dominated by exome sequencing, MLPA (Multiplex Ligation-dependent Probe Amplification) and microarray, increasingly also short read whole genome sequencing (srWGS) have been applied. To better characterize certain complex genetic variations in the genome, multi-omics approaches have recently been introduced in variant-detection pipelines. The aim of this study was to compare samples which not gave satisfying results with previous techniques to long-read sequencing (lrWGS). Materials and methods: The study involves two index patients with Neurofibromatosis type 1 (Linköpings University Hospital, Linköping, Sweden) and three index patients with hereditary breast and colorectal cancer (Sahlgrenska University Hospital, Gothenburg, Sweden). Long read sequencing were performed on the sequencing instrument Revio (PacBio, Menlo Park, USA) for all the patients. Results: Improved characterization of the structural variants of all five samples were achieved with lrWGS. In one patient previously MLPA results indicated a deletion of one exon in the gene *NF1* but a closer look by lrWGS revealed the variant to be a complex translocation. A second patient with a presumed duplication including three exons in gene *ATR* turned out to be just one exon. Conclusions: Introduction of lrWGS means that more complex and time-consuming techniques used in the clinic today can be replaced by one that also, in most cases, allows for increased sensitivity. As lrWGS permits improved characterization by e.g. mapping of exact break points of CNVs, more adequate sizing of repeat expansions and also detection of methylation. Overall, lrWGS will bring several opportunities for improved characterization of variants and also the possibility of increased diagnostic yield. The study was supported by grants from Swedish Cancer Society, RFoU and ALF Grants, Region Östergötland.

Session Title: Omics Technologies Poster Thursday Session

Board 1101T: A curated research catalogue of structural variation detected by optical genome mapping

Authors:

A. Pang[1], A. Chitsazan[1], S. Shukor[1], A. Hastie[2], A. Chaubey[3]; [1]Bionano Genomics, Inc., San Diego, CA, [2]Bionano Genomics, San Diego, CA, [3]Bionano, San Diego, CA

Abstract:

Comprehensive identification of genomic variants is key to understanding the mechanism of disease etiology and the curation of these variants into databases is critical to enable better understanding of their association with disease and other phenotypes. Optical genome mapping (OGM) is a technique that can detect all classes of structural variation (SV) including those undetectable by traditional methods such as, karyotyping, FISH, microarray and sequencing, and is discovering many novel variants not reported in current variant databases. The aim of this study is to develop a critical resource consisting of those curated variants to enable researchers to differentiate polymorphic variants from rare variants, and to better and more efficiently assess variants for their association to disease. From a cohort of 809 research subjects enrolled in a research study on constitutional postnatal disorders, 19 were controls, 101 were family members, and 689 were cases with various phenotypes such as musculoskeletal and neurodevelopmental disorders, intellectual disabilities, and developmental delay. The genomes of these subjects were mapped with OGM at an average coverage of ~800 Gbp, and SVs were called by comparing the de novo assembly against the public human reference (GRCh38). The variants were classified by Board certified laboratory directors using ACMG classification criteria. In total, this data contains a nonredundant set of 32,436 insertions (median 1.4kbp) and 43,995 deletions (median 1.6kbp), 3,544 duplications (median 74.1kbp), and 1,060 inversions (median 138.8kbp), and >100 translocations. In total, 176 of these variants were classified as pathogenic and likely-pathogenic. The remainder were either filtered out as polymorphic, or classified as benign, likely-benign, or of unknown significance. This curated dataset, along with sample clinical indications, aids in analysis of variants detected in research, and may improve the understanding of constitutional genetic diseases.

Board 1102T: A graph-based method to characterize tandem repeats and methylation for long read genome sequencing ⭐

Authors:

M. Khan[1,2], S. Chaudhary[3], M. Mrzik[1], S. He[1,4,5], N. Hassan[1,2,6], C. E. Pearson[1,2], J. T. Simpson[3,2,5], R. K. C. Yuen[1,2]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]The Univ. of Toronto, Dept. of Molecular Genetics, Toronto, ON, Canada, [3]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [4]The Univ. of Toronto, Dept. of Mathematics, Toronto, ON, Canada, [5]The Univ. of Toronto, Dept. of Computer Sci., Toronto, ON, Canada, [6]The Univ. of Toronto, Dept. of Cell and Systems Biology, Toronto, ON, Canada

Abstract:

Over 7% of the human genome is comprised of tandem repeats (TRs) - stretches of DNA sequence repeated adjacent to one another. These repeats are polymorphic in sequence and length across the population, and tandem repeat expansions (TREs) beyond a pathogenic length are associated with over 60 neurological, neurodevelopmental, and neuromuscular disorders. Larger TREs lead to earlier onset, increased severity, and faster progression of disease. GC-rich TREs can become hypermethylated, a mechanism that is associated with diseases such as Fragile X Syndrome. Repeat sequence purity also modifies outcomes of disease; interruptions in repeat sequences reduce disease severity and result in later age of disease onset.
Repeat size, sequence and methylation are difficult to characterize using existing assays including PCR, RP-PCR, and short-read sequencing, primarily due to the difficulty of sequencing through large, complex repeats. These assays also do not provide methylation information, and only provide limited repeat purity information.
Long-read sequencing, such as Oxford Nanopore Technologies (ONT), overcomes these limitations by providing base-pair sequence resolution and methylation information. We have developed a novel tool to estimate tandem repeat size, methylation (adjacent to, and within the repeat) and sequence composition. Here, we demonstrate methods for tandem repeat characterization using targeted Cas9 sequencing, adaptive sampling, and whole-genome sequencing approaches. We have also benchmarked our tool against other existing repeat-sizing tools. Using targeted Cas9 sequencing, the estimated repeat size from our tool is strongly correlated with the known repeat size in repeat-disease patient samples (R=0.93, p = 2x10-2), and is up to 95.5% accurate at low read depth (5x reads). Characterizing genome-wide pathogenic tandem repeat sizes from Nanopore WGS and comparing to short-read WGS reveals a strong correlation between our tool and short-read sequencing estimates by ExpansionHunter (R = 0.93, p = 2x10-7).
Our work highlights the benefit of long-read sequencing for the characterization of TREs. We

suggest optimized parameters for analyzing TREs with Nanopore sequencing, and demonstrate the utility of our approach in sizing heterogenous and mosaic TREs. In addition to TREs, we also analyze sequence purity and differential methylation patterns at TR loci in patients with rare diseases, and in trio cohorts. This method can ultimately result in faster characterization of TREs, more accurate prognosis, and faster clinical diagnosis of tandem-repeat associated diseases.

Session Title: Omics Technologies Poster Thursday Session

Board 1103T: A Method for genome-wide analysis of short tandem repeats' impact on human gene expression using massively parallel reporter assays ⭐

Authors:

M. Zhang[1], Y. Hu[2]; [1]Shanghai Jiaotong Univ., Shanghai, China, [2]Shanghai Dynegene Technologies, Shanghai, China

Abstract:

Short tandem repeats (STRs) serve as significant contributors to genetic variation, with expression STRs (eSTRs) playing a pivotal role in regulating the expression levels of neighboring genes and thus contributing substantially to variations in human gene expression. Studies have demonstrated that STR variations are linked to Mendelian diseases, exemplified by Huntington's disease. The Massively Parallel Reporter Assay (MPRA) has been a prevalent technique for investigating genetic variations and gene regulation associated with diseases, primarily focusing on single nucleotide polymorphisms (SNPs) of genome-wide association study.However, there has been a notable lack of direct, genome-wide analysis concerning the influence of STRs on human gene expression. To address this gap, we have developed a method termed esMPRA, an abbreviation for eSTR-related Massively Parallel Reporter Assay, designed for the high-throughput validation of regulatory elements associated with eSTRs. The esMPRA platform, an enhancement of the existing MPRA workflow, integrates our novel ultra-long oligonucleotide synthesis platform and a dual fluorescence reporter assay. This integration allows for the simultaneous screening and validation of numerous STR loci to assess their influence on downstream gene expression.In our study, we employed esMPRA to screen a database of short tandem repeats and successfully validated multiple eSTRs associated with various genes, including AKAP7, CD151, ARHGAP27, ANKEF1, TBC1D10C, RNF38, among others. The eSTRs identified through this process have been shown to significantly enhance the expression of proximal genes. The development of esMPRA facilitates rapid and efficient screening and

validation of eSTR-enriched regulatory elements, which is expected to be invaluable for researchers aiming to elucidate the roles of a broader spectrum of STR loci in various diseases.

Session Title: Omics Technologies Poster Thursday Session

Board 1104T: A Modular Workflow with Minimal QC Steps Provides Improved Capture Performance for Intact gDNA/total RNA as well as Low Input FFPE.

Authors:

N. Liu[1], A. Janssen[2], M. Borns[2], S. Shiryaev[2], K. Chen[2], B. Hsue[2], B. Clark[2], J. Fox[2], N. Saremi[1], I. Jingga[1], K-s. Jeong[1], K. Win[1], G. Amparo[1], N. Mehendale[1], J. Ghosh[1], C. Hon[1], B. Arezi[2]; [1]Agilent Technologies, Santa Clara, CA, [2]Agilent Technologies, La Jolla, CA

Abstract:

The limited availability of genomic DNA as well as its poor quality and variation in tumor fraction from sample to sample, can influence how readily NGS is applied to clinical tumor samples. Furthermore, practical considerations such as turnaround time, workflow simplicity, pre-capture pooling capability and performance are important factors on whether NGS would be the method of choice for germline mutation detection. Here we address these challenges by developing a modular and flexible workflow solution which is optimized for a wide range of DNA and RNA input (10-200 ng), various sample types (intact, FFPE, and cf DNA samples), standard and molecular-barcoded adaptors (dsMBC or ssMBC), different shearing methods (mechanical vs. enzymatic), sequencing chemistries (2 x 100 bp, 2 x 150 bp, or 2X250bp), no pre-capture QC requirement, fast or overnight bait panel hybridization, and up to 384 unique dual sample indices to eliminate index hopping. Our enzymatic library preparation method combines fragmentation, end repair, and dA-tailing in a single step using a high volume of 50ul. This approach allows FFPE DNA samples with low concentrations to be utilized without a need for a vacuum concentrator, with an input capacity of up to **40 µl.** As we will demonstrate, fragmentation and library preparation experiments using three bacteria of varying GC content (30%-67%) shows high coverage uniformity and minimal bias across all GC content (Fold 80 of 1.12-1.13). Optional molecular barcoding at the ligation step allows PCR errors to be filtered out by making consensus calls using MBC information from both strands (duplex MBC), hybrid (duplex MBC where info is available), single strand (single MBC) or discard MBC information (no MBC), based on the application. While both duplex and single MBC improve the accuracy of low VAF detection compared to no MBC, duplex MBC enables the most effective error correction. Finally, our novel fast hybridization buffer formulation provides maximum complexity (3-5%

duplicates at 6Gb for 8 pooled exome V8 libraries) while maintaining high uniformity (Fold 80 of 1.4) and coverage (96% at 30X base coverage), specifically to address the issue of complexity loss when multiple libraries are pooled during bait hybridization.

Session Title: Omics Technologies Poster Thursday Session

Board 1105T: A multi-dimensional approach to map disease relationships challenges classical disease views

Authors:

L. Möbus[1], A. Serra[1], M. Fratello[1], A. Pavel[2], A. Federico[1], D. Greco[1]; [1]Tampere Univ., Tampere, Finland, [2]Technical Univ. of Denmark, Kongens, Denmark

Abstract:

This study introduces a novel approach to understanding diseases by mapping their relationships through a multi-dimensional framework. Traditional disease categorization relies on anatomical views and symptoms. In contrast, this method integrates genomic, chemical, and clinical data from public sources to uncover deeper mechanistic connections. We analyzed disease-to-disease similarities for 502 diseases using six data dimensions: disease-associated genes, pathways, symptoms, drugs, chemicals targeting genes, and chemicals mediating diseases. By merging these dimensions into a consensus disease map and hierarchical clustering, we identified 71 distinct disease clusters.
Our analysis revealed significant distinctions from the classical International Classification of Diseases (ICD) system, as organ-system or anatomical site dominance was evident in only a few clusters. Among cancerous disease clusters, a clear trend based on cell type of origin rather than anatomical site was observed, with notable differences between epithelial and soft-tissue cancers. For non-cancerous diseases, known associations between inflammatory and cardiometabolic diseases were confirmed, and further links to neurodegenerative diseases were observed.
These findings indicate the need for an updated classification system incorporating mechanistic insights, which could enhance drug development, improve understanding of comorbidity patterns, and aid in classifying rare diseases.
To evaluate the utility of the disease map in grouping diseases with their mechanistically closest counterparts, we introduced a seventh data dimension from curated transcriptomic studies and lists of differentially expressed genes. Gene expression data, whether low-dimensional (PCR) or high-dimensional (NGS), are frequently generated to understand diseases due to their affordability and interpretability. We are currently testing if differentially expressed genes can

group a disease with its counterpart in the multi-dimensional map. Results from this ongoing work will be presented at the final conference.
This multi-dimensional approach for disease classification offers a nuanced understanding of disease mechanisms and relationships. By going beyond traditional anatomical views, it could enhance drug development and our ability to address complex comorbidity links. The ongoing validation with transcriptomic data will refine this model, offering a promising avenue for precise medical interventions.

Session Title: Omics Technologies Poster Thursday Session

Board 1106T: A streamlined enzymatic method for whole genome methylation detection

Authors:

J. Sanford, V. Panchapakesa, V. Ponnaluri, D. J. Evanich, A. Erijman, R. Vaisvila, B. W. Langhorst, L. Williams; New England Biolabs, Ipswich, MA

Abstract:

Methylation is one of many factors that contributes to epigenetic regulation of gene expression. Gene regulation in the human genome is typically associated with cytosines methylated in CpG context, which are frequently clustered in CpG rich regions. Sodium bisulfite conversion was the traditional method used to distinguish 5-methylcytosines (5mC) and 5-hydroxymethylcytosines (5hmC) from cytosines. However, this chemically-based method is harsh, it damages DNA and introduces significant sequencing bias. Here we describe a streamlined and optimized enzyme-based conversion method that gives superior results. This protocol improves on a previously developed enzyme-based protocol that minimizes DNA damage therefore enabling longer insert sizes, lower duplication rates and more accurate quantification of methylation.
Here we demonstrate robust results for our streamlined protocol using DNA inputs ranging from 100 pg to 200 ng. Assessed methylation metrics for the libraries across the input range are consistent with earlier versions of our protocol, while also demonstrating improvements in library yield, consistent global methylation and evenness of GC bias representation. The increased library yields using this enhanced, streamlined protocol subsequently reduces sequencing duplications and costs.
Our method marks a direct improvement in enzyme driven methylation detection of 5mC and 5hmC in the human genome. Leveraging this enhanced streamlined protocol and associated improvements results in higher quality data and cost effective methylome analysis.

Session Title: Omics Technologies Poster Thursday Session

Board 1108T: An Integrative Approach to Estimate the Adverse Impacts of Butylparaben on Othologous Human Genes using *Daphnia magna*via a Two-step BLAST Strategy and Network Analysis

Authors:

Y. Seo[1], J. Lee[1], D. SHIN[1], H. Kim[1], D. Choi[2], J. Lee[3], E-M. Cho[4]; [1]Dongguk Univ., Goyang-si, Korea, Republic of, [2]Korea Conformity Lab., Seoul, Korea, Republic of, [3]Gachon Univ., Seongnam, Korea, Republic of, [4]Seokeong Univ., Seoul, Korea, Republic of

Abstract:

Butylparaben, an organic compound classified as an endocrine-disrupting chemical, is commonly used as a preservative in cosmetics, baby products, and deodorants. It enters aquatic ecosystems through wastewater and has been detected in various water sources, including drinking water, at low concentrations. While most research on butylparaben toxicity in humans focuses on acute toxicity, few studies have elucidated the underlying molecular mechanisms of butylparaben exposure in humans. Due to the challenges of conducting in vivo experiments on humans, we aim to explore the chronic toxic effects of low concentrations of butylparaben using the model organism *Daphnia magna* (*D. magna*). *D. magna*, a freshwater organism, is recommended by the OECD for chronic exposure experiments and allows for standardized testing (OECD TG 211). Our study revealed that butylparaben at a dose of 0.32 mg/L over 21 days significantly reduced *D. magna* fecundity. Moreover, we identified 441 differentially expressed genes through RNA sequencing of the whole body of *D. magna*. To estimate the adverse impacts of chronic exposure to low concentrations of butylparaben on humans, we employed a two-step BLAST strategy and identified 401 homologous human genes. Network analysis estimated that chronic exposure to low concentrations of butylparaben could be associated with various diseases, including male reproductive issues, embryonic development disorders, diabetes, and mental disorders. Our integrative approach identified the adverse effects of chronic butylparaben exposure in humans and provided insight into genes that should be considered as potential biomarkers for environmental diseases resulting from chronic exposure to low concentrations of chemicals.

Session Title: Omics Technologies Poster Thursday Session

Board 1109T: Assessment and comparative analysis of novel single cell RNASeq platforms

Authors:

C. Kao, H. Qu, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Profiling of whole transcriptome at the single-cell level (scRNASeq) has undergone explosive growth and usage in the research community in the past decade, predicated in large part to the accessibility of microfluidics-based technology capable of reliably sorting individual cells from a suspension culture each into a uniquely barcoded, gel bead-in-emulsion that then accurately and uniformly converts the RNA within each cell into libraries which are compatible with available next-gen sequencing (NGS) platforms. However, due to the fragility of RNA as well as the rapid changes that occur in cells following ex vivo extraction from its native environment, there is a limited time window available to preserve and/or process the cells before its transcriptome signature begin to deviate (e.g. from cell death). Thus, current workflows often require that the cells be harvested, prepared in single cell suspension, then rapidly inputted and processed in the microfluidics instrument (e.g. 10x Genomics Chromium Controller) to accurately capture a “snapshot” of the native transcriptome. But situations arise where it may be impractical to harvest and prepare cells/tissues and then immediately process on specialized single-cell instrumentation. So the prospects of “instrument-free” methods to generate scRNASeq libraries are attractive as potential alternatives that can relieve such bottlenecks and reduce costs. Previously, we had examined split-pool barcoding kits/assays to generate scRNASeq libraries. We now evaluate two new additional platforms: 1) the SimpleCell 3’ Gene Expression Assay from CS Genetics and 2) the HIVE Single Cell Solution from HoneyComb Biotechnologies. Both SimpleCell and HIVE are able to perform immediate, single-cell indexing and preservation of transcriptome information without the need for specialized instrumentation or prior cell fixation. Importantly, the initial assay intermediates for both platforms are stable for extended periods (weeks to months) for subsequent processing and completion of library preparation and sequencing. We compared full transcriptome data on peripheral blood mononuclear cells (PBMCs) at the single-cell level generated by SimpleCell, HIVE, and 10x Genomics Chromium 3’ assays and sequenced on the NovaSeq. We also assessed the impact of cryopreservation on transcriptome profiles using standard freezing media containing fetal calf/bovine serum (FBS/FCS) vs. non-serum alternatives. Advantages and limitations of each platform were evaluated in terms of data quality and pragmatic considerations (e.g. cost, ease of use). The impact of cryopreservation on transcriptional signatures was also assessed.

Session Title: Omics Technologies Poster Thursday Session

Board 1110T: Automated and scalable single technology telomere-to-telomere assembly of human genomes

Authors:

S. Mckenzie[1], G. Dodd[2], S. Juul[3], P. Rescheneder[2]; [1]Oxford Nanopore Technologies, Spokane, WA, [2]Oxford Nanopore Technologies, Oxford, United Kingdom, [3]Oxford Nanopore Technologies, New York, NY

Abstract:

Telomere-to-telomere (T2T) genome assemblies provide the most complete picture possible of an organism's genetic sequence. Such assemblies allow researchers to precisely characterize genetic differences and potential causes of disease in otherwise un-interrogatable repeat regions and complex or highly divergent structural variants. Despite the apparent advantages of T2T assembly for genetic characterization, adoption is not widespread, in large part due to the high costs and onerous wet and dry lab burden to generate and analyze the multiple data types required for current T2T assembly algorithms. Previously, these algorithms required high-depth sequencing from two separate technologies - near-perfect long reads and ultra-long reads - in order to build the T2T assembly backbone. In addition, some form of long-range phasing data, such as chromatin conformation capture or parental sequencing, is necessary for final haplotype reconstruction. Here, we show that improvements in Nanopore ultra-long read accuracy, combined with the new HERRO algorithm for pre-assembly read correction, allow for a T2T assembly backbone to be built with just two flowcells of Nanopore ultra-long reads using the Verkko assembler. Furthermore, when Nanopore Pore-C data is used for long-range phasing of the assembly backbone, we show that this method even outperforms the previous state-of-the-art multi-technology methods, with most human chromosomes assembled T2T with no manual intervention. Together, our results demonstrate a method for researchers to build near-perfect T2T assemblies with a fraction of the resources previously required.

Session Title: Omics Technologies Poster Thursday Session

Board 1111T: Automated Spatially Resolved Laser-Activated Cell Sorting (SLACS) for In-depth Analysis of Heterogeneity in Mouse Brain

Authors:

S. Lee, E. Kim, H. Jeongwoo, A. Lee; Meteor Biotech, Seoul, Korea, Republic of

Abstract:

Traditional staining methods like Hematoxylin and Eosin (H&E) and immunofluorescent (IF) staining, alongside next-generation sequencing (NGS), have significantly advanced our understanding of cellular structures and genomic landscapes. However, integrating spatial and molecular data to fully elucidate cellular behaviors within their native environments remains challenging. This highlights the need for technologies that bridge the gap between large-scale imaging and detailed molecular analysis.To address these limitations, we present Spatially Resolved Laser-Activated Cell Sorting (SLACS), a novel methodology that preserves spatial information while integrating spatial assays with molecular techniques such as DNA sequencing, RNA sequencing, and protein profiling. SLACS uses laser activation on a discharging layer-coated slide to selectively isolate cells from heterogeneous samples. These cells are then propelled into PCR tubes for further analysis.SLACS operates through sophisticated software that allows for both automated and manual control of cell sorting, enabling high precision and throughput. This method supports the isolation of various entities, including bacteria, chromosomes, and neural cells from complex biological matrices, demonstrating its versatility.We applied SLACS to analyze heterogeneous cell populations in the mouse brain. Using this advanced platform, we successfully isolated each cell types from regions of interest while preserving their spatial context. The laser activation allowed for the precise propulsion of cells into PCR tubes, enabling the identification and comparison of molecular profiles from microglial cells located in various regions. SLACS effectively maintained spatial integrity and offered high-throughput capabilities, validating the platform's reliability. These findings underscore the importance of spatial context in studying function and interactions within the brain.SLACS represents a significant advancement in spatial omics, bridging the gap between imaging and molecular analysis. This technology opens new avenues for exploring cellular interactions and functions within their native environments, offering a powerful tool for both basic research and therapeutic development. By providing detailed spatial and molecular insights, SLACS stands to significantly enhance our understanding of biology and pathology.

Session Title: Omics Technologies Poster Thursday Session

Board 1112T: Automation of time-consuming WES steps and use of LLMs

Authors:

A. Janakiraman[1], A. Pal[2], S. Sur[3], **R. Bettadapura**[4], R. Hariharan[4]; [1]Strand Life Sci. Pvt Ltd, Bengaluru, India, [2]Strand Life Sci. Pvt LTD, Bangalore, India, [3]Strand Life Sci. Pvt Ltd, Bangalore, India, [4]Strand Life Sci. Private Limited, San Francisco, CA

Abstract:

In order to keep pace with increasing WES workloads for rare disease diagnosis in our CAP-accredited laboratory, we have implemented an effective variant prioritization algorithm named festiVAR (fast estimation of variants for automated reporting). festiVAR takes as input ~40,000 SNVs and small InDels, and CNVs, obtained after sequencing at ~120-150x followed by alignment and variant calling. It then prioritizes variants taking into account HPO terms for the clinical indication,age of disease onset, presence of the gene in our repository of curated gene lists, the gene inheritance mode, the number of variants in the gene, the predicted effects of these variants, the allele frequencies of these variants and the presence of these variants in ClinVar. For benchmarking festiVAR, we used a set of 996 cases that were processed using a previous variant interpretation algorithm that on average shortlisted ~60 genes per case for manual assessment and reporting; these were manually assessed and selected variants were then reported. In 99.93% of all cases, festiVAR prioritized all reported variants within the top 25 genes.

To further improve performance, we augmented festiVAR with an LLM-based approach to shortlist the top 25 genes above, by directly using the free-text clinical notes for the case. Clinical notes as well as 1418 manually assessed gene-phenotype correlations (including 459 Strong and 959 Weak correlations) were used to fine-tune the GPT-3.5 Turbo LLM. A separate test set of 1590 genes assessed using this GPT model, resulted in an accuracy of 98.43% (498/506 Strong correlations labeled correctly, and 1067/1084 Weak correlations labeled correctly), indicating significant improvements compared to the base GPT-4 LLM which yielded an accuracy of 92%.

Next, the variants in these genes are assessed for most of the ACMG criteria in an automated manner to arrive at a tentative Pathogenicity label. A final step involving LLM based automation of the literature search for functional studies and segregation studies completes the overall analysis of the WES case in an efficient short turnaround. We have implemented a LangChain application to automate the search for scientific literature. We also implemented an assessment interface that displays the shortlisted ranked gene list, with known phenotypic information from HPO, OMIM and Clinvar displayed readily for efficient genotype-phenotype correlation, and all relevant variant information, enabling manual inspection of the entire automated workflow where needed.

festiVAR is now in production, and accompanied by a user interface, enabling our teams of variant interpreters to process hundreds of exomes every month.

Session Title: Omics Technologies Poster Thursday Session

Board 1113T: Benchmarking of Straglr for the detection of short-tandem repeats using high-coverage Cas9-enriched nanopore sequencing data

Authors:

V. Scholz, M. Lucas, V. Schönrock, V. Mayer, C. Dineiger, H. Erdmann, C. Heintz, A. Hallermayr, A. Benet-Pages, E. Holinski-Feder, T. Neuhann, A. Abicht; MGZ - Medizinisch Genetisches Zentrum, Munich, Germany

Abstract:

In recent years, several bioinformatics methods for detecting expansions of short-tandem repeats (STRs) from long-read data have been developed. Establishing these tools within a diagnostic setting and validating the detection of repeats remains challenging due to the individuality of the disease-causing loci presenting as variable repeat lengths, expansions up to several hundreds of bases, and polymorphisms of the STRs.
We integrated the Oxford nanopore Epi2Me Human variation workflow (wf-human-variation), including Straglr, into a pipeline for calling and reporting of STRs from high-coverage data based on CRISPR-Cas9 target enrichment of Oxford Nanopore long-read sequencing. We conducted comprehensive benchmarking using positive controls for known Ataxia loci, Coriell samples, and negative controls.
We calculated specificity, positive predictive value, the limit of detection, and minimum coverage needed to reliably detect known repeat expansions using Straglr. The Coriell Ashkenazi family trio was used to determine the correct detection of the zygosity of the repeat loci. Coriell samples and positive controls were used to assess the performance of detection of expansions in the *HTT* locus. Furthermore, the repeat pattern was assessed within the *RFC1* locus using the visualization of the wf-human-variation.
Here, we validated for the first time the wf-human-variation and Straglr with high-coverage targeted long-read data. We determined the best course of action for the individual disease-causing loci for routine diagnostics, depending on the accuracy and the cost of the method.

Session Title: Omics Technologies Poster Thursday Session

Board 1114T: Beyond genomics: using RNAseq to unlock the clinical relevance of splicing variation.

Authors:

J. Pinto Basto, A. Bertoli-Avella, M. Radefeldt, A-a. Ruslan, L. M. Pardo, S. Lemke, A. Leubauer, D. Polla, K. Kandaswamy, C. Beetz, P. Bauer; CENTOGENE GmbH, Rostock, Germany

Abstract:

Clinical exome and genome sequencing are proven first line diagnostic methods in rare diseases with diagnostic yields going up to 60%. Yet, many patients remain undiagnosed, even after genome sequencing. We aimed to assess the impact of splicing variants reported in our laboratory, to gain insight into their clinical relevance. A total of 108 consecutive individuals for whom 113 splicing variants had been reported were selected for RNA-sequencing (RNA-seq), considering the gene expression in blood. A protocol was developed to perform RNA extraction and sequencing using the same sample (dried blood sports) provided for the DNA analysis, including library preparation and bioinformatic pipeline analysis. Splicing in genes of interest was inspected using IGV with at least three unaffected individuals as controls. From the 113 variants, we confirmed an abnormal splicing in 64 variants (57%). In 15 variants (13%) there was no alteration in splicing observed. In the remaining 34 variants, no decision could be made on splicing effect because the low sample quality (21%), or due to insufficient number of reads (9%). The most common event leading to aberrant splicing was exon skipping, that we identified in 31 variants (50%). Other events included cryptic donor/acceptor sites (n=25; 38%), intron retention (n=4; 6%), and other complex events (n=4; 6%). In three patients, the pathologically reduced enzymatic activity (measured using the same DBS) served as additional confirmation of the abnormal splicing caused by the variants in *HEXA, GAA* and *GLA*. With the current work, we implemented an RNA-seq pipeline using the same sample provided for ES/GS, with the aim to deliver a more comprehensive and efficient genetic testing. In some cases, additional (biochemical) testing was performed, providing a multiomic approach for the diagnosis of patients with rare diseases.

Session Title: Omics Technologies Poster Thursday Session

Board 1116T: Brain-regional gene expression imputed from the blood transcriptome by *BrainGENIE* recapitulates dysregulation observed in the *postmortem* brain in Alzheimer's disease

Authors:

J. Hou[1], **S. Razavi**[1], S-J. Lin[2], J. Hess[1], C. Zhang[3], S. Glatt[1]; [1]SUNY Upstate Med. Univ., Syracuse, NY, [2]Univ. of California San Diego, La Jolla, CA, [3]Upstate Med. Univ., Syracuse, NY

Abstract:

Background To circumvent some of the major issues plaguing *postmortem* brain analyses, we developed the Brain Gene Expression and Network Imputation Engine (*BrainGENIE*), which is a

method for predicting gene-expression levels in the brain from peripheral blood that goes beyond assessing monotonic relationships of genes expressed in both tissues. In this study, we aim to use *BrainGENIE*-imputed gene expression to identify genes and gene-sets that are differentially expressed in the brains of living individuals with AD. **Methods** We used *BrainGENIE* to impute brain-regional gene-expression profiles across 10 brain regions, including amygdala, anterior cingulate cortex, cerebellum, dorsolateral prefrontal cortex (DLPFC), hippocampus, hypothalamus, caudate nucleus, nucleus accumbens, putamen, and substantia nigra. For each brain region, we identified AD-associated gene-expression changes by pooling imputed data from 777 AD cases and 779 cognitively unimpaired comparisons. **Results** We identified genes that were significantly differentially expressed (FDR-$p < 0.05$) in AD cases in nine brain regions (except for substantia nigra), with the number of genes ranging from six to 705. For cerebellum and DLPFC, differential gene-expression effect sizes estimated from imputed gene-expression profiles were significantly correlated with those estimated from directly measured *postmortem* gene-expression profiles (Spearman's r = 0.351, p = 3.5e-180; Spearman's r = 0.281, p = 1.5e-144; respectively). Among the AD-associated genes identified in directly measured *postmortem* brain data, two genes in cerebellum and six genes in DLPFC were significantly replicated in the *BrainGENIE*-imputed data (FDR-$p < 0.05$). Across 10 imputed brain regions, we found a total of 493 distinctly up-regulated and 326 distinctly down-regulated Reactome pathways in AD cases (FDR-$p < 0.05$). We found two up-regulated pathways shared between those identified from imputed and directly measured DLPFC gene-expression profiles (Fisher's exact test, Bonf-p = 0.015). We found an overlap of 24 down-regulated pathways between imputed and directly measured gene-expression profiles, including 12 in DLPFC (Fisher's exact test, Bonf-p = 6.4e-08) and 12 in cerebellum (Fisher's exact test, Bonf-p = 7.6e-12). **Conclusion** Our study demonstrated that *BrainGENIE*-imputed brain-regional gene-expression profiles of AD could parallel transcriptomic changes in the *postmortem* brain to some extent, and could help identify previously undetectable patterns of transcriptome abnormality in the brains of people actively suffering from AD.

Session Title: Omics Technologies Poster Thursday Session

Board 1117T: Causal inference of drug effects on metabolomics and proteomics in UK Biobank

Authors:

J. Moon, S. Lee, M. Kho, S. Lee, C. Kang, J. Lee, S. Jeong; Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

Accurate evaluation and prediction of drug response and adverse drug effects are crucial for targeted treatment and prevention. A large-scale population-based biobank, combined with electronic health records and extensive omics biomarkers, could provide real-world evidence to identify and predict drug response and adverse effects. However, longitudinal omics data is limited, and confounding factors of cross-sectional data often obscure the true effects of drugs when using biobank data. This study demonstrates that regression discontinuity in time (RDiT) can be used to infer drug effects on biomarkers and then identify metabolite and protein level changes affected by selected drugs in UK Biobank.
First, we evaluated the utility of the RDiT framework as a causal inference analysis tool of drug effects by testing the known target biomarkers (e.g., statins-LDL) or biomarkers of adverse drug effects (e.g., statins-bilirubin) in UK Biobank European ancestry participants. Next, we applied the RDiT framework to omics biomarkers (325 NMR metabolites and 2,923 proteins) with 24 commonly prescribed ATC level-4 drug categories such as statins, PPIs, ACE inhibitors, and biguanides. Lastly, we compared the results of RDiT with those from Mendelian Randomization analyses for drugs targeting specific genes.
Our results with the RDiT method confirmed the known drug responses - statins on LDL (FDR adj P=3.0E-37, Beta= -1.3), biguanides on glucose (FDR adj P=2.8E-03, Beta= -3.9) - and side effects (e.g., statins on bilirubin with FDR adjusted P=4.6E-10, Beta=0.4) on blood biomarkers after adjusting for age, sex, assessment center, BMI, and genetic principal components. Evaluating metabolites and proteins with the RDiT method identified 703 causal associations with NMR metabolites (lowest FDR adj P=1.8E-33 with statins on non-HDL-C, Beta= -1.3) and 84 with proteins (lowest FDR adj P=1.6E-17 with statins-PLA2G7, Beta= -0.5). After further adjustment of smoking status, alcohol consumption, years of education, socio-economic status, and other medication use, causal associations for 596 drug-metabolite and 68 drug-protein remained significant. Our Mendelian Randomization results also showed a significant causal association between statins and protein PLA2G7 (weighted median method P=2.9E-16) with an appropriate direction of effect.
This research presents the possibility of identifying omics biomarkers influenced by drugs using large-scale real-population data, which contributes to the real-world evaluation of the drugs' effects and better personalized therapeutical drug usage. Further research is warranted to validate suggested causal associations.

Session Title: Omics Technologies Poster Thursday Session

Board 1118T: Caution when using network partners for target identification in drug discovery

Authors:

D. Tan[1], Y. Chen[1], Y. Ilboudo[2], K. Liang[1], G. Butler-Laporte[3], B. Richards[4]; [1]McGill Univ., Montreal, QC, Canada, [2]Jewish Gen. Hosp., Montreal, QC, Canada, [3]Univ. of Oxford, Montreal, QC, Canada, [4]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada

Abstract:

Identifying novel, high-yield drug targets remains challenging and is often associated with a high failure rate. Recent data strongly suggests that using human genetic evidence to identify and validate drug targets increases the probability of drug development success. Two recent papers from Open Targets (Hinxton, UK) showed that the targets of about one-half of FDA-approved drugs had direct evidence from human genetics and even more successful drug targets could be identified by expanding to protein network partners, regardless of human genetic evidence, using the IntAct database. Network partners were defined as molecules in physical contact with each other at the same time. Expanding the list of targets to include such network partners increased the proportion of FDA-approved medicines whose targets had genetic evidence support to two-thirds from one-half. While more drug targets were identified using molecularly interacting proteins, the efficacy of including network partners to aid in target identification was not formally tested. To address this question, we identified positive control genes from a variety of sources that are known to influence disease risk, and tested whether adding IntAct-identified network partners increased the ability and precision to find these positive controls. We used the IntAct database to find molecular interacting proteins of genes identified by exome-wide association studies (ExWAS) and genome-wide association studies combined with a locus-to-gene mapping algorithm called the Effector Index. We evaluated the precision, sensitivity and specificity of including the interacting genes to the ExWAS and Effector Index selected genes in finding positive controls. The results showed that directly including network partners of ExWAS significant genes decreased precision by 6-fold, increased sensitivity by 5%, and maintained high specificity compared to ExWAS significant genes alone. For Effector Index prediction, with network partners, the precision also decreased by 6-fold, the sensitivity increased by 13%, and specificity remained above 95%. Taken together, while the Open Target studies have highlighted the potential of molecular interactions in identifying drug targets, their practical application is hindered by low precision. This makes it challenging to employ in real-world scenarios like the identification of new drug targets. Expanding genetically identified targets to include their network partners does not appear to increase the ability to identify positive control genes and such results should be interpreted with caution.

Session Title: Omics Technologies Poster Thursday Session

Board 1119T: Charting inflammasome-driven immune infiltration in homologous recombination deficient high grade serous ovarian cancer

Authors:

N. Gull[1], K. Dabke[2], B. Ben Cheikh[3], M. Diaz[4], P-C. Peng[4], N. Mazumdar[5], B. Karlan[6], B. Rimel[4], S. Gayther[7], S. Parker[4], M. Jones[8]; [1]Cedars Sinai Med. Ctr., Los Angeles, CA, [2]Ellison Inst. of Technology, Los Angeles, CA, [3]Akoya BioSci.s, Menlo Park, CA, [4]Cedars-Sinai Med. Ctr., Los Angeles, CA, [5]Cedars Sinai Med. Ctr., West Hollywood, CA, [6]David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [7]The Joe R. and Teresa Lozano Long Sch. of Med., San Antonio, TX, [8]Cedars Sinai Med Ctr, Los Angeles, CA

Abstract:

Despite great improvements in our understanding of epithelial ovarian cancer (EOC), this disease still ranks among the top five deadliest cancers in women. High grade serous ovarian cancer (HGSOC) accounts for more than 2/3 of all EOC cases. HGSOC are diagnosed at late stage and associated with poor prognoses. Genomic instability from homologous recombination deficiency (HRD) can modulate tumor immunogenicity and promote immune invasion, rendering these tumors more susceptible to immunotherapy. For example, *BRCA1* mutated tumors show an immunoreactive molecular subtype with intra-tumoral immune cell infiltration. These immune microenvironment alterations have been associated with activation of immune sensors and receptors associated with inflammasomes. Recently, inflammasomes have been suggested as potential therapeutic targets for epithelial cancers. By limiting inflammasome activation, tumor cells may be hindered in their ability to migrate and proliferate. We hypothesize that the inflammasome has a role in controlling the presence of tumor infiltrating lymphocytes in HRD tumors, increasing inflammation, and rendering these tumors more susceptible to therapeutics. We sought to establish the effects of HRD and chemotherapy on immune cell infiltration in HGSOC. We employed a 26-plex antibody panel to explore the single-cell spatial relationships of immune cells and HRD tumors in paired primary and recurrent HGSOC using Akoya's multiplex spatial technology. A total of 36.3 million cells were identified and classified into 19 cell subtypes from 23 tumors sections. It has previously been reported that HRD tumors demonstrate increased immunogenicity through activation of dsDNA anti-viral response pathways mediated through cGAS-STING. In our study, STING+ tumor cells were significantly higher in HRD tumors compared to homologous recombination proficient (HRP) tumors ($p= 0.049$). In recurrent tumors, Treg levels changed inversely based on HRD status: in HRD tumors, Tregs increased, while in HRP tumors, Tregs decreased. Consistent with previous observations from an image analysis of HGSOC, CD11c+ dendritic cells were significantly higher in HRP tumors compared to HRD tumors. In contrast, we saw a significant increase in M1 macrophage count in HRD tumors compared to HRP tumors. We previously characterized these

tumors with RNAseq, and found differential expression of GSDMC, a protein associated with inflammasome-mediated pyroptosis, between HRD and HRP tumors. Ongoing analysis of these spatial relationships is underway, but our preliminary data indicate a role of the inflammasome in mediating immune cell infiltration into HGSOC tumors.

Session Title: Omics Technologies Poster Thursday Session

Board 1120T: Circulating cell-free DNA as a novel genomic biomarker of immune and organ dysfunction during acute infection

Authors:

K. Cano-Gamez[1], E. Foss[1], M. Inoue[2,3], S. Hussainy[2,3], P. Maclean[1], H. Qin[1,4], C. Wainright[1], S. McKechnie[5], GAinS Investigators, C. Song[2,3], J. Knight[1,4,5]; [1]Ctr. for Human Genetics, Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [2]Ludwig Inst. for Cancer Res., Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [3]Target Discovery Inst., Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [4]Chinese Academy of Med. Sci. Oxford Inst., Univ. of Oxford, Oxford, United Kingdom, [5]John Radcliffe Hosp., Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom

Abstract:

Circulating cell-free DNA (cfDNA) is a well-studied biomarker in cancer, pregnancy, and transplantation. However, its role in acute infection remains poorly understood. Previous studies have shown that cfDNA increases in sepsis, is predictive of patient prognosis, and correlates with organ failure. This makes cfDNA a promising biomarker for patient monitoring. Here, we isolated cfDNA from 86 plasma samples, corresponding to 12 healthy controls and 46 patients with sepsis sampled repeatedly over time. The amount of DNA in circulation increased by a mean of 33.8-fold during sepsis. Further, cfDNA concentration was associated with patient outcome (e.g. likelihood of ICU admission and 28-day mortality), highlighting its translational potential.

We used TET-assisted pyridine borane sequencing (TAPS) to profile genome-wide methylation in 58 randomly selected cfDNA samples. EpiDISH, a computational deconvolution approach, was used to infer tissues of origin. Compositional estimates showed good agreement with blood cell counts and reflected immune cell turnover rates, with long-lived cells (e.g. T and B cells) showing lower contribution to cfDNA than short-lived cells (e.g. neutrophils and monocytes). Surprisingly, we observed no difference in the composition of cfDNA between patients and controls, challenging the idea that cfDNA increases solely due to higher rates of apoptosis, NETosis, or pyroptosis during sepsis. Instead, we observed a strong correlation between cfDNA

concentration and liver function tests (ALT, AST, albumin, and PT), suggesting liver dysfunction prevents clearance of cfDNA during sepsis. This was supported by associations between liver function and cfDNA fragmentation, with mononucleosomal particles and DNASE1L3 cutting sites increased during sepsis. These alterations tended to resolve over time, suggesting that cfDNA clearance fluctuates throughout disease.
Finally, we demonstrate that cfDNA carries information about gene activity, with promoters of active genes showing clear nucleosome footprints in cfDNA. We used mathematical models to infer the strength of nucleosome phasing, revealing a strong correlation between phasing and gene activity.
In conclusion, we observe a large increase in circulating cfDNA during sepsis, which could be explained by impaired DNA clearance in the liver. Furthermore, we show that cfDNA retains information about gene activity, which could enable novel assays that capture gene expression and cfDNA information in a single experiment. We believe cfDNA could be used to monitor liver function, cellular turnover, organ damage, and gene activity during infection.

Session Title: Omics Technologies Poster Thursday Session

Board 1121T: Clinical blended genome exome (cBGE) for translational genomics

Authors:

K. Larkin, M. DeFelice, J. Grimsby, M. Coole, C. Patterson, M. Gatzen, M. Gildea, C. Kachulis, S. Hofherr, N. Lennon; Broad Clinical Labs, Cambridge, MA

Abstract:

For precision medicine to be widely accessible, there is a need for robust and cost-effective next-generation sequencing methods that can support screening applications at scale with the delivery of actionable findings. Population-based studies typically have two requirements, one is to get genome-level single nucleotide variant information and the other is to identify mutations in a specific list of genes that can be returned to study participants. Historically, this would have been accomplished by performing two assays, a screening microarray, and a targeted gene sequencing panel. Clinical whole genome sequencing (WGS) can satisfy both requirements, but while WGS has become less expensive in recent years it can be cost-prohibitive for very large-scale population-based studies. To reduce the cost and complexity of running two independent tests, we developed the Blended Genome Exome (BGE), a novel test that marries the benefits of WGS for single nucleotide variant detection, utilizing imputation, with deep whole exome sequencing (WES) to identify mutations for actionable return of results.

This approach combines an exome library and a whole genome library into a single tube for sequencing. Initially launched in a research setting, BGE yields a 35-45x exome with a 2-3x genome, allowing for low-cost population-scale sequencing. However, this research-oriented approach lacks the exome depth required for clinical applications. We have developed a clinical Blended Genome Exome (cBGE) to address this gap. In the cBGE process, the amount of exome library present in the blend is increased. This brings the exome to &gt80x mean coverage, in line with the standard for commercially available clinical exome products, while still incorporating whole genome data at 2-3x coverage. This sequencing approach balances cost considerations with clinical utility and can be used for multiple clinical applications (e.g. CDC Tier 1 screening, polygenic risk scores).
We will describe our workflow in detail, including the results of our analytical validation study, which was performed to assess the performance of this clinical-grade assay. A combination of reference samples, and previously tested clinical samples were used to verify that the assay is suitable for its intended use as part of clinical applications. We will also demonstrate the utility of this approach, despite limitations in diagnostic applications, for genetic risk.

Session Title: Omics Technologies Poster Thursday Session

Board 1122T: Comprehensive Study and Implementation of Enhanced Healthcare Data Management Practices in Korea

Authors:

K. Lee[1], J. Choi[2], I. Cho[1], S-J. Kim[1], S. Park[1]; [1]Insilicogen, Inc, Yongin, Korea, Republic of, [2]Bioto, Inc, Daejeon, Korea, Republic of

Abstract:

Healthcare data is vital for advancing medical research, improving patient outcomes, and informing public health policies. Our research aims to enhance data management through the development and implementation of a revised BioResearch Data Standard Registration Form, addressing current gaps in data registration, deposit, management, and utilization in Korea.
We began by surveying and summarizing the current status of health data practices. This foundational work enabled us to identify various data types and structures. Through comparative analysis, we evaluated different methods for data deposit, purification, and distribution, leading to practical operational suggestions.
A key component of our research is preparing an on-premise-cloud interworking plan, ensuring infrastructure support for seamless data integration and accessibility. Based on our findings, we established comprehensive guidelines for the registration, deposit, purification, and

standardization of healthcare data, designed to enhance data quality and interoperability. Furthermore, a specialized clinical epidemiology standardization management system has been established in Korea, and a foundation has been laid to utilize it from data registration to distribution.
Our efforts culminated in constructing a robust database that researchers can readily utilize, facilitating efficient data retrieval and analysis. The revised BioResearch Data Standard Registration Form provides a standardized framework for consistent data collection and management.
This research sets a new benchmark for healthcare data management in Korea. By addressing current challenges and implementing innovative solutions, we aim to significantly improve the quality and usability of healthcare data. This fosters more effective and collaborative research, leading to better health outcomes and more informed public health decisions.

Session Title: Omics Technologies Poster Thursday Session

Board 1123T: Context-Dependent Embedding and Interpretation of Genomic Data Using Large Language Models

Authors:

W. Gong, A. Mahajan; Genentech, Inc, South San Francisco, CA

Abstract:

Background: Recent research has focused on leveraging large-scale single-cell (sc) data to develop specialized foundational models such as GeneFormer and scGPT. Large Language Models (LLMs) have shown promise in interpreting expression data due to the rich semantic context of gene names and their relative rankings, as demonstrated by GenePT-s. However, LLM embeddings have not yet matched the performance of dedicated single-cell foundation models. In this study, we introduce scLLM, a novel method that embeds scRNA-seq data by leveraging existing knowledge of gene functions, pathways, and disease associations.
Results: We created a database that maps biological information from sources like Gene Ontology, Reactome, OpenTargets, and CellMarker2.0 to gene lists. Using a retrieval-augmented generation (RAG) system, we retrieved contextual information according to task-related queries to annotate genes associated with tasks such as cell clustering. The top-ranking genes with contextual annotations were used to generate cell-level embeddings. We evaluated the performance of scLLM, dedicated foundational models, and other LLM-based models (GenePT-s and GenePT-w) in integrating heterogeneous scRNA-seq datasets, without utilizing any batch labels. The adjusted rand index (ARI) was used to quantitatively assess the consistency between

integrated cell clusters and known cell labels. Our results show that scLLM outperforms scGPT in integrating eight scRNA-seq datasets across multiple human/mouse tissues, including failing heart (0.59 vs 0.41), healthy heart (0.42 vs 0.21), pancreas (0.50 vs. 0.097), PBMC (0.54 vs. 0.41), immune atlas (0.21 vs. 0.097), lung atlas (0.32 vs. 0.18), brain (0.31 vs. 0.16) and retina (0.19 vs 0.18) (ARI: scLLM vs. scGPT). Notably, scLLM requires no additional training or fine-tuning and utilizes off-the-shelf language embedding models, enabling easy scalability to atlas-level datasets. Among the language embedding models evaluated, OpenAI's text-embedding-3-small and UAE-Large-V1 yielded the best performance for closed and open-source models, respectively.
Conclusion: We propose that scLLM offers a straightforward and effective method for leveraging LLMs to integrate and interpret scRNA-seq data, providing superior performance without the need for further training or fine-tuning. This approach has the potential to significantly advance the field of single-cell analysis by simplifying the integration process and enhancing the interpretability of gene expression data. Further research could explore the application of scLLM to other types of omics data and investigate its potential in clinical settings.

Session Title: Omics Technologies Poster Thursday Session

Board 1124T: Current genomic deep learning models display decreased performance in cell type specific accessible regions

Authors:

P. Kathail[1], R. W. Shuai[1], R. Chung[1], C. Ye[2], G. Loeb[2], N. Ioannidis[1]; [1]Univ. of California, Berkeley, Berkeley, CA, [2]UCSF, San Francisco, CA

Abstract:

Background: A number of deep learning models have been developed to predict epigenetic features such as chromatin accessibility from DNA sequence. A key application of these models is to probe the functional consequences of genetic variation, particularly disease-associated genetic variation, and current models have shown mixed success in identifying causal GWAS variants and their regulatory mechanisms. Model evaluations commonly report accuracy genome-wide; however, *cis* regulatory elements (CREs), and in particular, cell type specific CREs, make up only a small fraction of the genome yet play critical roles in gene regulation and harbor a large fraction of the heritability of complex diseases.
Results: Here, we evaluate the performance of genomic deep learning models in accessible regions with varying degrees of cell type specificity. We assess two alternate modeling directions in the field: general purpose models trained across thousands of cell types and

epigenetic marks, and models tailored to specific tissues and tasks. We find that the accuracy of genomic deep learning models, including two state-of-the-art general purpose models -- Enformer and Sei -- varies dramatically across the genome and is reduced in cell type specific accessible regions. For example, Enformer makes highly accurate predictions for the least cell type specific regions (median Pearson R 0.76) but performs poorly on the most cell type specific regions (median Pearson R 0.1). Using tissue-specific ATAC-seq data from kidney and immune cells, we identify that increasing a model's capacity to learn cell type specific regulatory syntax -- either through single-task learning or high capacity multi-task models -- is a potential method to improve performance in cell type specific accessible regions. Finally, we use GWAS, eQTL, and allelic imbalance data to extensively evaluate models on their ability to predict variant effects in accessible regions with varying degrees of cell type specificity. We observe a somewhat surprising result, that improvements in reference sequence prediction do not always lead to a direct improvement in variant effect prediction, indicating that novel training strategies are needed to improve variant effect accuracy.
Conclusions: Our results reveal that the accuracy of genomic deep learning models varies dramatically across the genome, and is particularly reduced in cell type specific accessible regions. Our evaluations provide a new perspective on the performance of current genomic deep learning models, and an outlook on strategies to maximize performance in cell type specific accessible regions.

Session Title: Omics Technologies Poster Thursday Session

Board 1125T: De novo mutation identification without parental data: Implications for rare disease gene discovery and heritable effect of ionizing radiation

Authors:

S. Lee, M-h. Choi, J-W. Nam; Hanyang Univ., Seoul, Korea, Republic of

Abstract:

With the advent of whole genome sequencing, it has become possible to directly identify de novo mutations (DNMs), leading to the discovery of many new disease genes and advancing research into mutation mechanisms. However, the necessity for parental information has impeded research in situations where parents are deceased, unavailable or choose not to participate in studies. We have developed a method to call DNMs using sibling data in incomplete trios, where one or both parents' information is missing. This method was evaluated on CEPH and Korean families with one or more siblings, demonstrating high accuracy with a

false discovery rate (FDR) of 26% in approximately half of the genome, referred to as callable rate (CR). After adjusting for CR and FDR, the results were comparable to those obtained using the conventional trio method, which we refer to as complete trios. Moreover, the characteristics of mutations identified in incomplete trios, such as paternal age effect and mutation spectra, closely resembled those observed in complete trios. To assess our method's effectiveness in rare disease gene discovery, we intentionally excluded one parent's information while analyzing 178 Korean rare disease patients. This revealed 8,405 de novo mutations, including 12 loss-of-function variants all absent in parental genomes. Notably, two of these mutations correlated with the patients' phenotypic manifestations. Additionally, we applied this method to analyzing DNMs in offspring of Korean atomic bomb survivors, a population where parental data is often unavailable due to parental mortality, to explore the heritable consequences of ionizing radiation exposure. We called DNMs in 120 offsprings and identified two individuals exhibiting hypermutation. We are currently calling DNMs in additional offspring of parents exposed to ionizing radiation, where parental data is unavailable. Furthermore, we are in the process of comparing the rates of DNMs with those of healthy controls. We anticipate that our method will offer novel insights into the factors affecting DNM occurrence and their implications for disease onset.

Session Title: Omics Technologies Poster Thursday Session

Board 1126T: Deep learning-based multi-omics data integration framework identifies possible drug-repurposing gene targets for sarcopenia

Authors:

M. Reza, K-J. Su, A. Liu, X. Zhang, Y. Gong, Z. Luo, Q. Tian, C. Qiu, H. Shen, H-W. Deng; Tulane Univ., New Orleans, LA

Abstract:

Sarcopenia, marked by the age-related decline in muscle mass, strength, and diminished physical performance, poses a significant public health challenge in the context of a rapidly aging population. Finding effective pharmacological therapies has become more crucial as there are currently no approved drugs that specifically target sarcopenia. Therefore, we developed a new multi-task attention learning algorithm that effectively combines multi-omics data to identify drug targets and discover new treatments for sarcopenia. Our method combines various omics datasets such as transcriptome, methylome, and genome, obtained from a cohort of 1010 individuals aged between 20 and 50, as well as incorporates human protein-protein

interaction networks and drug-target networks. Based on our methods, we identified 639 target genes, and by analyzing PPIs and consulting public databases, we identified seven potential hub genes associated with sarcopenia, namely *ESR1, ATM, CDC42, EP300, PIK3CA, EGF,* and *PTK2B*. The validation of these findings was further supported by incorporating various levels of pathobiological evidence associated with sarcopenia. The analysis of Gene Ontology and KEGG pathways revealed five important functions and signaling pathways closely related to skeletal muscle. Furthermore, the interaction network analysis identified three transcription factors (GATA2, JUN, FOXC1) as key regulators of seven potential genes. Furthermore, the top seven selected genes significantly bond with four drugs- Testosterone, MK-0773, Vorinostat, and Canagliflozin out of 1940 candidates for drug repurposing in sarcopenia. Specifically, Canagliflozin exhibited the binding energy at -9.5 kcal/mol and the inhibition constant of 6.97 µM against the PTK2B target protein. Hence, our method enables multi-omics data integration to identify gene targets associated with lean mass, suggesting that canagliflozin may be a promising therapeutic option to address sarcopenia.
Acknowledgment: This research was made possible with partial support from the National Institutes of Health (R01AR069055, U19AG055373, R01AG061917, P20GM109036).

Session Title: Omics Technologies Poster Thursday Session

Board 1127T: Designing Cell-Type-Specific Promoter Sequences Using Conservative Model-Based Optimization

Authors:

A. Reddy, X. Geng, M. Herschl, S. Kolli, A. Kumar, P. Hsu, S. Levine, N. Ioannidis; Univ. of California, Berkeley, Berkeley, CA

Abstract:

Gene therapies have the potential to treat disease by delivering therapeutic genetic cargo to disease-associated cells. One limitation to their widespread use is the lack of short regulatory sequences, or promoters, that differentially induce the expression of delivered genetic cargo in target cells, minimizing side effects in other cell types. Such cell-type-specific promoters are difficult to discover using existing methods, requiring either manual curation or access to large datasets of promoter-driven expression from both targeted and untargeted cells.
Model-based optimization (MBO) has emerged as an effective method to design biological sequences in an automated manner, and recent promoter design methods have made use of MBO techniques by training machine learning models of promoter-driven expression and using these models to test and optimize promoters in-silico. However, these methods have only been

tested using large training datasets that are expensive to collect, and focus on designing promoters for markedly different cell types, overlooking the complexities associated with designing promoters for closely related cell types that share similar regulatory features. Therefore, we introduce a comprehensive strategy to utilize MBO for designing promoters in a data-efficient manner, with a specific emphasis on discovering promoters for similar cell types. We focus on using conservative objective models (COMs) for MBO since these models are less prone to producing unreliable results when tested using sequences that are very different from the training sequences. Furthermore, we highlight practical considerations such as best practices for improving sequence diversity, getting estimates of model uncertainty, and choosing an optimal set of sequences for experimental validation.
We evaluate our design strategy using three relatively similar blood cancer cell lines (Jurkat, K562, and THP1), and show that our approach is capable of discovering novel cell-type-specific promoters after experimentally validating the designed sequences. **For K562 cells in particular, we discover a promoter that has 75.85% higher cell-type-specificity than the best promoter from the initial dataset used to train our models.** We compare our method to a conventional motif tiling-based approach and to a recent generative models-based approach and show that we outperform these previous approaches. Our approach to sequence design is flexible and can be used to design promoters in many settings to improve the effectiveness of gene therapies.

Session Title: Omics Technologies Poster Thursday Session

Board 1128T: Detecting differential RNA editing sites using single-cell RNA sequencing data

Authors:

W. Wang, R. Yamamoto, J. Li, X. Xiao; Univ. of California Los Angeles, Los Angeles, CA

Abstract:

RNA editing is a post-transcriptional modification that introduces an additional layer of variation in RNA and has been demonstrated to play a crucial role in immune response. Large-scale shifts of RNA editing levels in bulk RNA-Seq data have been reported to be associated with several cancer types. The advent of single-cell RNA-seq (scRNA-Seq) technologies enables investigating RNA editing patterns at the single-cell level. However, there is currently no established method for detecting differential RNA editing sites (DES) using scRNA-Seq data. In this study, we aim to propose a workflow for detecting trustworthy DES using scRNA-Seq data. The proposed workflow involves metacell construction, RNA editing level quantification, DES identification, and false discovery rate (FDR) control. Initially, we utilize the metacell approach to cluster

similar cells within the same condition and perform RNA editing level quantification within each metacell. By examining the editing levels across different conditions within these metacells, we employ three distinct DES identification methods (Wilcoxon rank-sum test, REDIT, and edgeR), each based on different statistical models. Following the DES identification, we apply two FDR control methods (Benjamini-Hochberg procedure and Clipper) to ensure the robustness of our findings.We evaluated the workflow using real data from the ROSMAP snRNA-Seq dataset on Alzheimer's disease, as well as synthetic data. The results indicate that metacell construction significantly enhances RNA editing level quantification. Among the three tested DES methods and two FDR control methods, the combination of REDIT and Clipper demonstrated superior performance, achieving a high true positive rate and a low false positive rate.In conclusion, we proposed a workflow for detecting trustworthy differential RNA editing sites using single-cell RNA-Seq data. Tested on both real and synthetic data, our workflow effectively identifies differential RNA editing sites with a high true positive rate and low false positive rate. This approach has the potential to advance RNA editing research using scRNA-Seq data and improve differential RNA editing sites identification across different cell types and conditions.

Session Title: Omics Technologies Poster Thursday Session

Board 1129T: Detection of repeat expansions with PureTarget

Authors:

M. Eberle[1], G. De Sena Brandine[2], V. Gaysinskaya[2], J. Aiyedun[2], J. Rocha[3], D. Kilburn[2], S. Kingan[4], E. Dolzhenko[2], Z. Kontogeorgiou[5], A. Szabo[6], C. Zarouchlioti[6], R. Thaenert[7], P. Alvarez Jerez[8], K. Billingsley[8], S. Lameiras[9], S. Baulande[9], A. Davidson[10], G. Koutsis[5], G. Karadima[5], S. Tome[11]; [1]PacBio, Oceanside, CA, [2]PacBio, Menlo Park, CA, [3]PacBio, Bel Air, MD, [4]PacBio, San Mateo, CA, [5]Natl. and Kapodistrian Univ. of Athens, Athens, Greece, [6]Univ. Coll. London, London, United Kingdom, [7]Quest Diagnostics, Marlborough, MA, [8]NIH, Bethesda, MD, [9]Inst. Curie, Paris, France, [10]UCL, London, United Kingdom, [11]INSERM, Paris, France

Abstract:

Short tandem repeats (STRs) are DNA sequences composed of repetitions of 1-6bp motifs. Expansions of STRs are the cause of over 60 monogenic diseases, including Huntington's disease, fragile X syndrome, and amyotrophic lateral sclerosis. In addition to their length, the pathogenicity of these STRs is impacted by sequence composition, methylation status and mosaicism. One such example is a repeat in an intron of the *RFC1* gene whose reference sequence consists of a short stretch of AAAAGs while expansions that span hundreds of AAGGGs cause cerebellar ataxia with neuropathy and vestibular areflexia syndrome. Another

example is the *FMR1* repeat whose expansions are typically hypermethylated. Detecting all the characteristics associated with pathogenic repeat expansions traditionally required multiple assays, however long read sequencing of unamplified DNA holds the promise to resolve all of the required features in a single assay.
We describe a robust amplification-free protocol to generate long-read HiFi sequencing libraries containing a panel of loci associated with 20 pathogenic STR expansions. The protocol can be multiplexed to sequence 48 samples at up to 1000x coverage per locus in one sequencing run. To assess the accuracy of this protocol, we sequenced 129 samples with validated pathogenic expansions at 20 loci including *CNBP*, *DMPK*, *RFC1* and *C9orf72*. Combined, we tested 2580 sample-expansion combinations, including technical replicates, for expansions between 66bp and >10kb. Our assay correctly categorized all (129/129) expansions, including the detection of hypermethylation in the *FMR1* expansion and differentiating the pathogenic AAGGG motif in *RFC1*. We identified additional expansions in *FXN*, *RFC1* and *TCF4*, consistent with these loci having carrier frequencies between 1:50 and 1:20. Excluding these three genes, we found no unexpected expansions (0/2064) in any sample-loci combination.
We will also present a detailed characterization of lengths, sequence composition, mosaicism, and methylation of normal and expanded alleles in 150 genomes. Most repeats we profiled exhibit high genetic or epigenetic polymorphism and also mosaicism at the expanded size ranges. Motivated by these results, we describe a novel computational approach that will capture all these modalities to robustly differentiate between normal and abnormal variation at known pathogenic or any other repeats in the human genome.
In summary, we will present a protocol and a set of computational methods for accurately assessing tissue-level molecular landscape of various pathogenic STRs, which can be further adapted to other loci in the human genome.

Session Title: Omics Technologies Poster Thursday Session

Board 1130T: Differential diagnosis of rare genetic diseases aided by large language model

Authors:

H. Li[1], Y. Jian[2], L. Shu[3], H. Duan[2]; [1]Children's Hosp., Zhejiang Univ. Sch. of Med., Hangzhou, China, [2]Zhejiang Univ., Hangzhou, China, [3]Rhode Island Hosp., Providence, RI

Abstract:

Large Language Models (LLMs) like ChatGPT have been around for a while and have transformed many industries, including healthcare and medicine. They are sophisticated neural

network-based models that can understand natural language patterns and produce text that is reasonably trustworthy and human-like. Such conversational AI provides a more adaptive and natural method of human-machine interaction than traditional keyword search techniques such as Orphanet and FindZebia for searching for R&D knowledge. A major challenge is the generation of false information, commonly known as hallucination, which poses a significant risk, particularly in sensitive domains such as healthcare. Furthermore, ensuring the credibility of knowledge generated by LLMs remains a critical concern, especially in specialised domains such as rare diseases (RD) knowledge retrieval and medical diagnosis.In this study, we introduce RDguru, a conversational intelligent agent tailored for RDs, built on the LangChain framework and incorporating GPT-3.5-turbo as an automated reasoning and summarising LLM. RDguru's functionalities are centred around knowledge Q&A and medical advice. We have customised several tools for knowledge-based RD knowledge Q&A. When disease entities are queried and appropriate processing tools are inferred, selected pre-defined tools are invoked to retrieve relevant knowledge fragments from online databases such as Orphanet or pre-loaded Orphadata, ensuring evidence-based answer generation. In addition, RDguru includes a professional tool-assisted medical consultation process to support interpretable and accurate RD diagnosis. This process integrates automated phenotype extraction and context analysis methods for analysing case text reports, PheLR for phenotypic diagnosis interpretability, and RDmaster's phenotype-oriented Q&A strategy for differentiating competing diagnoses.
A demo has been made available at http://rdguru.nbscn.org.

Session Title: Omics Technologies Poster Thursday Session

Board 1131T: Elucidating genetic mechanisms directing in vitro blood cell formation with single cell multiomics

Authors:

R. Qiu[1], A. Alkanaq[2], M. B. Wilken[1], C. Shalaby[1], Q. She[1], G. Pavani[1], C. S. Thom[1,3]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of North Carolina in Chapel Hill Hosp., Chapel Hill, NC, [3]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

An inability to efficiently produce hematopoietic progenitor cells (HPCs) in vitro has hindered derivation of stem cell-derived blood cell therapeutics. We hypothesized that defining in vitro HPC development at a single cell level would elucidate genetic mechanisms that support in vitro hematopoiesis. We used an in vitro model in which human induced pluripotent stem cells (iPSCs) differentiate into hemogenic endothelial cells (HECs) that form HPCs with multilineage

potential (Pavani et al, Blood Adv 2024). This model mirrors in vivo development and produces stromal and vascular endothelial cell populations along with blood cells. We optimized a protocol to isolate nuclei from cells at different culture stages to support multi-modal sequencing analysis (Qiu et al, JoVE 2024). We analyzed 26,080 high quality nuclei from cultured cells during nascent HPC formation by single-cell RNA/ATAC sequencing. Unbiased clustering and annotation revealed 10 major cell types, including stroma, vascular endothelia, HECs, HPCs, and hematopoietic cells at different stages of myeloid, lymphoid, megakaryocytes, or erythroid maturation. There were 242 transcription factor motifs significantly enriched in HPCs compared to HEC precursors ($p<1x10^{-4}$), including JUNB (1.4-fold enriched, $p<9x10^{-72}$) and the Notch-related factor HEY2 (1.4-fold, $p<2x10^{-63}$) known to impact the HEC-to-HPC transition. We then profiled cellular communication among cultured cells. TGFβ and BMP ligands secreted by stroma and vascular endothelia were received by HECs, mirroring interactions that support in vivo hematopoiesis. We also identified IGFBP interactions between vascular endothelia and HECs, indicating metabolic support. Finally, we found robust expression of collagen and extracellular matrix components in stroma and endothelia; collagen-mediated interactions support HEC specification and development in vivo. Leveraging multimodal single cell analyses in a clinically relevant in vitro model, our findings elucidate known and novel metabolic and extracellular matrix components that directly impact HEC biology and HPC formation. These findings reveal opportunities to optimize culture conditions and highlight the relevance of hematopoietic niche support in the in vitro system.

Session Title: Omics Technologies Poster Thursday Session

Board 1132T: Enabling low cell input and high throughput processing of cell and nuclei fixation for single-cell RNA profiling using combinatorial barcoding

Authors:

S. Schroeder, O. Kaplan, V. Tran, S. Marrujo, A. Salvino, A. Sapre, I. Dowsett, Z. Sayar, A. Koriath, A. Sova, J. Siputro, G. Geiss, C. Roco, A. Rosenberg; Parse BioSci.s, Seattle, WA

Abstract:

Single cell analysis is a critical approach to tease out nuanced information in any biological system. Combinatorial barcoding is an extremely efficient method for scRNA-seq at larger scale, achieving both sensitivity in gene detection and flexibility for sample input. However, this approach is challenging when starting with small sample inputs of less than 100,000 cells. To address this, we have developed a modified combinatorial barcoding approach that allows

researchers working with low sample inputs to fix, store and barcode their cells efficiently without the need for costly or specialized instruments. This approach can be used to fix and profile up to 96 samples in parallel, results in a shorter workflow, and enables consistent handling which minimizes both experimental error and burden on scaling. As a proof-of-concept, we have generated a screen of 96-cell culture samples with input amounts starting at 10,000-20,000 cells. We present a method for fixing and isolating cells and nuclei at high capture efficiency (>70%). The capture efficiency remains high even at the lowest input values. This method maintains the stability of the fixed sample for extended storage at -80℃ and is compatible with the combinatorial barcoding process.
The design of this workflow maximizes sample recovery while eliminating batch effects to provide accurate and comprehensive results on the transcriptional effect of experimental perturbations. This low-input fixation and capture method will be critical for the research community to expand their investigative scope to include precious samples at a larger scale, and to unlock new insights in biology for years to come.

Session Title: Omics Technologies Poster Thursday Session

Board 1133T: Enhanced Library Preparation Methods for Targeted Oxford Nanopore Technologies (ONT) Long-Read Sequencing

Authors:

Y-C. Ho, P-L. Chen; Natl. Taiwan Univ., Taipei, Taiwan

Abstract:

Advancements in next-generation sequencing (NGS) technologies have improved, yet many regions within the human genome, such as large structural variations, highly repetitive areas, and pseudogenes, remain challenging with short-read sequencing. Third-generation long-read sequencing technologies, like Oxford Nanopore Technologies (ONT), offer direct sequencing of target genes, addressing complex genomic structures like inversions, duplications, or translocations. However, targeted long-read sequencing (T-LRS) methods are still underdeveloped. This study aims to refine T-LRS by developing an effective method to target specific Regions of Interest (ROI) associated with diseases. We explored specialized library preparation methods to enhance T-LRS, focusing on the differences between circularized and linear DNA, and the effects of varying probe and primer characteristics on the technique.
We evaluated a specialized library preparation approach to increase adapter ligation efficiency and specificity, reduce adapter dimer formation, and employed target-specific primers for Rolling Circle Amplification (RCA) and Multiple Displacement Amplification (MDA) on both

circular and linear DNA. This methodology aimed to assess the fragment sizes, efficiency, and uniformity of the amplification. Using 120 bp probes at varied intervals, we examined the effectiveness of targeted enrichment.
Observations indicated that NEB Blunt/TA DNA ligase partially enhanced library preparation efficiency and specificity. This improved method could be applied to DNA circularization, comparing the effects between circular and conventional linear DNA. In experiments using biotinylated-probe capture, targeted enrichment was achieved with random hexamer MDA, though uniformity still requires improvement.
Our goal is to enhance the enrichment fold for targeted long-read sequencing through novel library preparation and DNA circularization techniques. We are also considering the integration of adaptive sampling from ONT to further increase enrichment efficiency. This approach is intended for complex genomic regions associated with pharmacogenomics, hearing impairment, or repeat expansion disorders.

Session Title: Omics Technologies Poster Thursday Session

Board 1134T: Estimating gene regulatory noise by measuring gene expression dispersion to assess robustness across cell types using single-cell gene expression data

Authors:

B. Jamison, Y. Gilad; Univ. of Chicago, Chicago, IL

Abstract:

Individual cells must be robust to genetic and environmental perturbations to maintain cellular identity and tissue function. A lack of robustness, manifested as an increase in gene expression noise, has been implicated in cancer, schizophrenia, and other complex traits. However, such observations have been made with bulk RNA-seq data which cannot differentiate between increased noise within cell types and changes in cell type composition of tissues. Bulk data is fundamentally limited for estimating noise because it can only capture averages of expression states. scRNA-seq, since it captures expression within individual cells, is uniquely equipped to measure moments of the expression distribution other than the expected value, including variance. Further, studies using scRNA-seq have already shown that even similar cells can have significant differences in gene regulation, meaning that cell states can often be robust with respect to regulatory variation. Studying how gene regulatory noise is controlled across different cell types and individuals can improve our understanding of cellular behavior and complex diseases. However, most current scRNA-seq datasets are underpowered with respect

to the task of estimating regulatory noise, with too few cells and insufficient sequencing depth, making it difficult to measure gene expression dispersion accurately. To address this, we collected a new, much larger scRNA-seq dataset from heterogenous differentiating cultures (HDCs) of iPSC-derived cardiac cell types. This dataset includes 75,000 cells from each of 3 lines, with 400,000 reads per cell—an order of magnitude more data per cell than typical scRNA-seq experiments. Because current datasets are underpowered for measuring gene regulatory dispersion, it is unclear what the best analytical approach is for this task. This dataset allows us to effectively compare methods for measuring gene regulatory dispersion in single-cell data, using both parametric (e.g., empirical Bayesian inference, linear regression) and non-parametric (e.g., gene homeostasis Z-index) approaches. Our findings will inform future studies on the genetic control of regulatory noise, aiding in the mapping of cell-type-specific robustness QTLs and enhancing our understanding of cellular resilience in health and disease.

Session Title: Omics Technologies Poster Thursday Session

Board 1135T: Evaluating transcriptomic signatures in post-mortem brain tissue of ALS and related disorders using innovative long-read sequencing technologies

Authors:

A. Jain[1], E. Udine[1], I. Cordts[1], M. DeJesusHernandez[1], N. Finch[1], R. Crook[1], M. Baker[1], N. Graff-Radford[1], B. Boeve[2], R. Peterson[2], K. Josephs[2], J. Shah[1], N. Staff[2], B. Oskarsson[1], S. Tian[3], L. Petrucelli[1], D. Dickson[1], R. Rademakers[1,4], M. van Blitterswijk[1]; [1]Mayo Clinic, Jacksonville, FL, [2]Mayo Clinic, Jacksonville, MN, [3]Mayo Clinic, Rochester, MN, [4]VIB Ctr. for Molecular Neurology, Antwerp, Belgium

Abstract:

Emerging evidence has revealed widespread splicing abnormalities in amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD). A pathological hallmark of these debilitating neurodegenerative diseases is the mislocalization and aggregation of TDP-43, an RNA-binding protein with a crucial role in RNA metabolism and splicing. However, common denominators leading to TDP-43 accumulation are not fully understood. While short-read RNA sequencing has been crucial in determining differences in gene expression and uncovering important molecular pathways, novel long-read RNA sequencing technologies can improve upon those efforts. Long-read sequencing can capture the transcriptional diversity of ALS and related disorders, accurately identifying and quantifying differentially expressed genes as well as transcript variants. As such, in our project, we leverage this technology to

examine a primary affected region, the motor cortex, from fully-characterized patients with and without well-known pathogenic *C9orf72* repeat expansions (n~60), as well as control subjects (n~20). Long-read sequencing is performed on the Revio platform (Pacific Biosciences), using the innovative Kinnex method. Full-length transcriptomic data is analyzed with a custom workflow that incorporates various tools (pbmm2, isoseq, pigeon, FLAIR, etc.). Additionally, complementary short-read RNA sequencing data from the same individuals is integrated to boost the quantitative power of our analysis. Multiple transcripts obtained are classified into major structural categories, including full splice matches, incomplete splice matches, and novel transcripts. Genes and individual transcripts are then quantified, enabling the identification of isoform-specific differences in expression and detection of alternative splicing events, thereby nominating disease-relevant genes, transcripts, and/or pathways. Our integrative approach, combining state-of-the-art long- and short-read sequencing technologies, assists in mapping the transcriptomic landscape of ALS, possibly contributing to the observed clinical and pathological heterogeneity. Ultimately, knowledge gained from this project may aid our ability to diagnose, monitor, and/or treat ALS and related disorders.

Session Title: Omics Technologies Poster Thursday Session

Board 1136T: Exploring Pediatric Disease Genomics with the Upgraded Variant WorkBench in the Kids First Data Resource Portal

Authors:

Y. Guo[1], J. Rozowsky[2], J-P. Thibert[3], J. Costanza[3], Q. Li[1], M. Mattioni[2], E. Wenger[1], D. Higgins[1], Y. Zhu[1], A. Heath[1], V. Ferretti[4], A. Resnick[1], Kids First Data Resource Center; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Velsera, Cambridge, MA, [3]Ctr. Hosp.ier Univ.ire Sainte-Justine, Montréal, QC, Canada, [4]Ctr. Hosp.ier Univ.ire Sainte-Justine, Montreal, QC, Canada

Abstract:

Aiming at facilitating researchers to uncover new insights into the biology of childhood cancers and structural birth defects, the Gabriella Miller Kids First Pediatric Research Program (Kids First) was initiated and Kids First Data Resource Center (KFDRC) developed the Kids First Data Resource Portal (KFDRP; https://portal.kidsfirstdrc.org/), which is a centralized data platform for both Kids First and collaborative cohorts. We as KFDRC present an upgraded version of the Variant WorkBench as part of the KFDRP, with more data incorporated, on a more efficient platform, and in a more streamlined data flow design. First, the current collection of Kids First data include reharmonized genomics data of over 174,000 files in 28,700 participants from 35 studies, all in the annotation VEP version 105. We also provide updated variant/gene annotation

databases from more than 50 public resources (e.g. gnomAD, ClinVar, HPO etc.). Second, the Variant WorkBench is running on Velsera's Cavatica Data Studio platform with a new Spark version 3.5.1 plus Python 3.11, achieving a ~10 fold acceleration in terms of executing pyspark codes when compared to the previous version. Third, we redesigned the data flow from KFDRP to Variant WorkBench, where portal users can now import Kids First data with which they have dbGaP approval directly to a Cavatica project and start analyzing in Variant WorkBench. As an example, we use Variant WorkBench to replicate findings of gene AMOTL1 from a previous publication in which Kids First data served as a data source, and to expand the AMOTL1 variant search in all available Kids First studies, plotting phenotypic distribution of participants with such variants. In conclusion, the upgraded Variant WorkBench enables accelerated exploration of pediatric disease genomics under the Kids First program.

Session Title: Omics Technologies Poster Thursday Session

Board 1137T: Feasibility of a Complete PCR-free Whole Exome Sequencing Using a New Integrated Library Preparation and Sequencing Workflow

Authors:

X. Qi, A. Mah, M. Ojeda, K. Wiseman, J. Moreno, B. Lajoie, S. Kruglyak, J. Zhao, S. Levy; Element BioSci.s, San Diego, CA

Abstract:

Over the last 15 years, whole exome sequencing (WES) has been demonstrated as a robust and cost-effective approach for research that can potentially help detect genetic variants across a broad spectrum of inherited human diseases and disorders. Despite ongoing efforts to shorten hybridization time and develop automation solutions for target capture workflows, the complex protocols, lengthy post-enrichment wash processes, and stringent temperature requirements continue to contribute to assay complexity and variability. Additionally, library preparation and PCR amplification steps limit the sensitivity of the assay due to the introduction of enzyme-based errors. Low concordance of INDELs is often seen when comparing WES and PCR-free whole genome sequencing (WGS).

We present a unique exome workflow that addresses these challenges by eliminating the need for PCR and seamlessly integrating post-hybridization steps directly into the sequencing process, removing several steps and improving the resolution of the workflow. After the enrichment step, the hybridization reaction is directly loaded onto the sequencing instrument. Applying this technology to a 24-plex exome enrichment with genomic reference DNA, high on-target rates

(>90%), high coverage uniformity (fold-80 penalty 1.3-1.4), and a mean sample coverage depth of ~70x were observed. Eliminating post-enrichment amplification results in a 3- to 4-fold lower duplication rate relative to the traditional WES workflow.
Eliminating PCR in the workflow has several positive effects. An improved indel F1 score that partially bridges the gap between traditional exome sequencing and whole genome PCR-free sequencing was observed. Initial results from an in-silico mixing of HG001 and HG002 show a ~13-26% reduction in false negatives (FN) and a ~65-72% reduction in false positives (FP) when comparing the PCR-free preparation to the traditional in-solution PCR-plus, demonstrating that eliminating PCR from target selection workflows has important opportunities for improving the resolution of low-frequency variants.

Session Title: Omics Technologies Poster Thursday Session

Board 1138T: From pixels to genes: leveraging spatial transcriptomics for single-cell resolution gene expression prediction in histology images

Authors:

S. Yao[1], D. Zhang[1], M. Li[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Spatial transcriptomics (ST) technologies have revolutionized our ability to profile gene expression across tissues while retaining crucial spatial context. However, the broad adoption of commercial ST platforms is hampered by their prohibitive costs and long turnaround time. In contrast, generating hematoxylin and eosin (H&E)-stained histology images is substantially more cost-effective. Previous studies have indicated correlations between gene expression patterns and histology image features, suggesting the potential for gene expression prediction from such images. Here we propose an AI model designed to predict single-cell resolution gene expression from histology images by leveraging information from diverse ST data types. Our model aims to incorporate ST data generated from popular ST platforms, including Visium, Visium HD, Xenium, MERSCOPE, and CosMx, into its training. We use a hierarchical vision transformer model to extract image features that capture both local and global tissue structures from histology images. A neural network then learns the relationship between these image features and gene expression. When training with ST data from both low-resolution platforms, such as Visium, and high-resolution platforms, such as Xenium, our model can use the neural network trained on Xenium data to initialize the neural network for Visium data, thereby enhancing prediction accuracy. By integrating ST training data of different spatial resolutions and gene coverages, we

can predict whole-transcriptome gene expression at single-cell resolution. Applying our model to mouse brain and human breast cancer samples, we demonstrate its capability to provide accurate single-cell resolution gene expression predictions across various samples and subjects. The model also identifies spatial domains and uncovers spatial patterns that are not easily discernible and would require careful pathological examination. The accuracy and robustness of our predictions were confirmed through extensive evaluations, including numerical assessment of prediction accuracy and reliability using large simulation sets and manual annotations by pathologists. Overall, our method is versatile and capable of integrating histology images with various ST data types. By virtually predicting single-cell resolution spatial gene expression across different samples and subjects, our model can significantly save time and money for researchers, thereby accelerating discoveries in disease research and precision medicine.

Session Title: Omics Technologies Poster Thursday Session

Board 1139T: From Validation to Diagnosis: Metadata Analysis of Long-Read Sequencing for Clinical Insights

Authors:

S. Drury, A. M. Tabish, E. Mor, H. Mor Shaked; Geneyx Genomics, Herzliya, Israel

Abstract:

Short read sequencing (SRS) is established in clinical practice for inherited disorders and whilst this has transformed the diagnostic odyssey for countless patients, there remains a need to solve the remaining undiagnosed. Long read sequencing (LRS) is a solution for uncovering variants not identified via SRS, however, has not yet been widely adopted. Here we systematically investigated the validation and diagnostic yield of LRS, to identify the added value of LRS and where in the clinical pathway LRS could be implemented. We performed a literature search for studies using LRS technology and split them into three groups 1: validation group (LRS re-confirmed variants, 232 cases from 14 studies), 2: case reports (37 studies), 3: diagnostic group (LRS identified variants, 366 cases from 19 studies). In the validation group, 231 cases were validated by LRS with a validation yield of 99% (231/230), and 50% of the validated cases had repeat expansion in their molecular etiology. One missed variant was in the *PKD1* gene (c.6223_6224insTT, p.Arg2075LeufsTer42). In the diagnostic group, 366 cases were solved by LRS (22% by PacBio and 14% by ONT) with a diagnostic yield of 34.4% (366/697). *DMD* (108), *SLC12A3* (42), *OPN1LW/OPN1MW* (38), *TP53* (12) and *ABHD12*(10) were the most commonly reported genes in the diagnostic group. In most studies, identified variants were not LRS exclusive, as SRS would have identified these variants by employing different

variant callers or manual inspection of reads. The benefits of LRS were seen in solving particular variant types (eg. fusions, complex SV, repeats expansions) and due to phasing analysis. Further, LRS resolved repeat expansion variants in 45 genes, with *ATXN10* being the most commonly observed gene followed by *HTT*, *FMR1*, *DMPK*, *RFC1* and others, highlighting the clear utility of LRS in repeat expansion disorders. This analysis showed ~100% validation yield of LRS, while testing on diverse variant effect where any other single technique would fall short, highlighting utility of LRS as one technique for all type of genetic variants in validation studies. LRS also showed superior diagnostic yield in characterizing complex structural variant, repeat expansion and SNVs. Trends are emerging in clinical applicability of LRS. Major benefit will be driven by enabling use of one technology for all types of variants- a paradigm shift in molecular diagnostics. LRS as 1st-tier testing in NICU or paediatric settings and complex oncological cases is already in practice. Large-scale applicability requires further implementation studies and clear messaging to the clinical community about diagnostic yield and utility.

Session Title: Omics Technologies Poster Thursday Session

Board 1140T: Haplotype calculation with SNV and Structural Variations for long-read sequencing data

Authors:

L. Luo, Y. Wu, C. Hancock, C. Hall, K. Levan, E. Bouton, C. Liu; SoftGenetics, State College, PA

Abstract:

Haplotype calculation is crucial in pedigree analysis for inherited disease with single nucleotide variants (SNVs) and structural variations (SVs). Phased SNVs and SVs from haplotype calculation are also essential for star allele detection in pharmacogenomics. For example, a long deletion that causes the loss of an exon and/or a single nucleotide variant in the HBB gene results in Beta-thalassemia. Repeat expansion of C9ORF72 gene is the cause of frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS). Pharmaco-genes, such as CYP2D6, pose challenges due to various SVs, including deletion, duplication, inversion as well as SNVs.
Compared with short-read technology, the reads of long-read sequencing data can capture most of a whole gene region, allowing the alignment of reads with both SNVs and SVs. However, long-read sequencing data, especially Nanopore data have higher sequencing error rates than short-read data. Therefore, we introduced a novel haplotype calculation method that phases SNVs and SVs simultaneously and sorts reads into different haplotypes using long-read data. A method to calculate the effective SNV was developed, based on neighborhood perfect matches, which can distinguish real variants from sequencing errors, providing more accurate haplotype

assignment. To phase both SNVs and SVs with high accuracy, SVs are given a weighted factor when sorting the reads for different haplotypes. SVs, such as long deletion/insertions typically cause coverage imbalance during the PCR process for amplicon data, resulting in low coverage of variants, especially small INDELs in one allele. A consensus calculation is also applied to the phased alleles to detect such variants with lower coverage.
The method was demonstrated on NCBI public long-read data of HBB gene, where an over 800 base pair insertion was detected and separated into different alleles. Additionally, both deletion and repeat expansion were detected and phased into diplotypes for Pacbio amplicon data of the C9ORF72 gene. The method also provided accurate phased variants for star allele detection in whole genome long-read data with pharmacogenetics for Coriell samples.

Session Title: Omics Technologies Poster Thursday Session

Board 1141T: Human pancreatic islet single cell transcriptomes reveal convergent T2D genetic and environmental β-cell gene expression effects.

Authors:

K. Bandesh[1], E. Motakis[1], S. Nargund[1], R. Kursawe[1], V. Selvam[1], R. M. Bhuiyan[1,2], G. N. Eryilmaz[1], S. N. Krishnan[1,2], D. Ucar[1,2,3], M. L. Stitzel[1,2,3]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]Dept. of Genetics and Genome Sci., UConn Hlth., Farmington, CT, [3]Inst. for Systems Genomics, UConn, Farmington, CT

Abstract:

Pancreatic islets are central to type 2 diabetes (T2D) genetics and etiology, harboring many T2D-associated genetic variants linked to crucial insulin regulatory genes like *ADCY5* and *KCNJ11*. Their mixed cell type composition impedes precise understanding of T2D genetic or pathophysiologic effects on islet function. Here, we completed human pancreatic islet single cell RNA-seq (scRNA-seq) in a 48-donor cohort (17 individuals with T2D, 14 prediabetes, and 17 without diabetes, ND), to decipher such cell-type-specific changes. Unsupervised clustering of 245,878 scRNA-seq profiles listed 14 cell types and revealed 15% reduction in islet β-cell proportion for T2D vs. ND islets ($P = 0.009$). β-cell reduction correlated with higher HbA1c across all donors ($r = -0.38$; $P = 0.008$). Pseudo-bulk analysis identified 511 differentially expressed genes in T2D vs. ND β-cells (FDR < 5%, |fold-change, FC| ≥ 1.5). Twenty percent of T2D upregulated genes were implicated in neuroactive-ligand receptor interaction with multiple genes encoding receptors for neurotransmitters - glutamate (*GRIN1*, *GRIN2A*), acetylcholine (*CHRNA3*, *CHRNA5*), and catecholamine (*ADRB1*). Apoptotic genes (e.g., *FAIM2*, *GAS6*, and *HGF*)

were also upregulated in T2D β-cells. Conversely, T2D downregulated genes were enriched in lipid metabolic processes and vitamin A/carotenoid metabolism. Additionally, T2D donor β-cells exhibited significant downregulation of genes mutated in maturity onset diabetes of the young (MODY, e.g., *HNF1A*, *HNF4A*, and *SLC2A2*). Given the dominant inheritance and highly penetrant genetic effects of all MODY variants, downregulation of these genes prompted further investigation into genes with expression patterns mirroring T2D genetic risk. By examining the eQTL associations in published islet datasets and our own, we currently determined 17 genes with concordant T2D genetic and environmental influences, including *STARD10, RNF6, SNX11, DGKB, GAD1,* and *ST6GAL1*. For instance, *STARD10*, crucial for intracellular lipid transfer, houses a T2D GWAS variant rs7109575 (G/A; G allele ↑ T2D risk, $P = 6.6 \times 10^{-64}$; OR = 1.10). Our analysis in whole islets showed that T2D risk allele G correlates with ↓ *STARD10* expression ($P = 1.7 \times 10^{-35}$, Effect = -0.56), consistent with reduced *STARD10* expression in T2D β-cells ($\log_2 FC = -0.74$; P = 0.001). These data and analyses nominate genes that merit high prioritization for therapeutic intervention based upon their modulation by both T2D genetic and environmental factors. Our study presents cell-type specific transcriptomic signatures as a valuable and publicly accessible resource for research in human islet biology, especially in T2D context.

Session Title: Omics Technologies Poster Thursday Session

Board 1142T: In vivo-in vitro genomic comparison reveals TNFa as a missing factor that supports in vitro hematopoiesis

Authors:

C. Thom, C. Shalaby, R. Qiu; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Inefficient blood cell production in vitro is a major hurdle for blood cell-based cell therapeutics development. In vitro systems can differentiate induced pluripotent stem cells (iPSCs) into hemogenic endothelial cells (HECs), which are direct precursors to hematopoietic stem and progenitor cells (HSPCs) that can form multiple blood lineages. Stromal and vascular endothelial cells are also generated in this culture system, but may not fully recapitulate cell types and niche factors that support HEC and HSPC formation in vivo. This study aimed to use single cell genomics approaches to identify factors that support in vivo HSPC formation and use related insights to augment in vitro HSPC yield.
To identify factors that support in vivo HSPC yield, we used CellChat to define 34 significant cell-cell interactions pathways among public single cell RNAseq data from human embryonic cells during HEC and HSPC formation (Calvanese et al, Nature 2022, padj<0.05). Among these

pathways were TGFb, VEGF, CXCL, and Wnt signaling from stroma and mesenchymal cells to HECs: These all support HEC formation and function. We also identified inflammatory (TNFa) and metabolic (Visfatin, Osteopontin) signals from circulating 'primitive' hematopoietic cells on HECs.
We hypothesized that key signaling modalities might be deficient in vitro, limiting HSPC formation. Single cell RNAseq of a clinically relevant in vitro hematopoietic system revealed that, while the culture system retained many key cell interaction activities, TNFa signaling was absent during nascent HSPC formation. TNFa signaling deficiency was confirmed by adding exogenous recombinant TNFa to cultures during HEC and HSPC formation, which doubled functional HSPC yield ($p<0.05$).
Our results reveal systemic and niche signaling factors that support hematopoiesis at a single cell level. Timely TNFa addition can bring in vitro HSPC formation closer to clinical efficiency. Further augmenting HEC and HSPC production via targeted genetic or environmental manipulation will enhance in vitro derivation of blood cell-based therapeutics. A similar in vivo-in vitro comparison approach at the single cell level may facilitate organoid development for genetic disease modeling in other systems.

Session Title: Omics Technologies Poster Thursday Session

Board 1143T: Infrared-laser based single cell sorting for advancing spatial omics

Authors:

H. Kim[1], A. Lee[1], S. Lee[2]; [1]Meteor Biotech, Seoul, Korea, Republic of, [2]Meteor Biotech, SEOUL, Korea, Republic of

Abstract:

Contextualizing the spatial arrangement of genetic elements in individual cells and the greater clusters and structures organizing them brings insight into cellular heterogeneity, microenvironment interactions, and dynamics of gene expression. However, accessing the intersection between single-cell technologies and spatial omics has several hurdles. First, the most frequent method of single-cell isolation, flow cytometry, requires a sample homogenization step that makes original spatial information unrecoverable. Although advanced imaging-based single-cell methods like Xenium and Visium from 10X Genomics are highly effective, their high cost impose considerable barriers. We present Spatially-resolved Laser Activated Cell Sorting (SLACS) as an affordable, powerfully precise, and high-throughput technique for intersecting genetic and spatial data at single-cell resolution. It employs an

infrared laser that is gentler than the harsh ultraviolet laser employed in standard laser capture microdissection techniques—in other words, causes less stress to the target cell or region, allowing the preservation and extraction of genomic data more closely reflecting its natural, in situ state. Notably, SLACS manages to facilitate this collection of high-quality omics data at an isolation rate of one target per second. A sequencing experiment of a mixed sample of 30,000 HL60 human leukemia cell line control cells and 10,000 SLACS-isolated MCF7 breast adenocarcinoma cell line experimental cells produced distinct clustering effects upon data analysis. A clearly differentiated cluster was affirmed to represent MCF7 cells by its high expression of the estrogen receptor alpha (ESR1) gene and minimal expression of estrogen receptor beta (ESR2) gene. This validates the effectiveness of SLACS in capturing genetic expression data at the single-cell level. Additionally, subsequent sequencing of a mixed-species sample comprising 30,000 HL60 cells and 10,000 SLACS-isolated cells from a mouse spleen sample provided further evidence of SLACS's compatibility with both cell lines and tissue samples. Continuing research expands into the capturing and sequencing of single-cell samples from a wide range of organ, organoid, and tumor samples. Greater integration of SLACS can enrich and expedite research aiming to pinpoint the interplaying ways in which spatial arrangement of genes and their expression underpin human biology, disease, and development.

Session Title: Omics Technologies Poster Thursday Session

Board 1144T: Integrating single-cell transcriptome and epigenome information with spatial analysis at scale with massively parallelized barcoding

Authors:

M. Nakamoto[1], **M. Masuda**[1], A. Holmes[1], B. Biddy[1], D. Skinner[1], S. Khare[1], H. Wang[2], R. Chen[2], M. Ray[2], F. Schlesinger[1]; [1]Scale BioSci.s, San Diego, CA, [2]Vizgen, Cambridge, MA

Abstract:

The advent of high-throughput sequencing technologies has revolutionized our understanding of cellular heterogeneity, however while transcriptome and epigenome information unlock understanding of heterogeneity within samples, it lacks spatial information important for understanding the organization and interactions within tissues. Conversely, spatial transcriptomics techniques provide spatial context but can be limited in the number of genes and modalities profiled. Integrating spatial and single-cell sequencing datasets can bridge this gap, offering insights into the spatial organization of gene expression dynamics transcriptome-wide within complex tissues.

In this study, we performed integration of single-cell RNA and single cell methylation data using the ScaleBio scRNA kit and the Vizgen Merfish platform. Specifically, adjacent brain slices 10 um or 300 um thick were collected from 3 separate brain regions of a mouse. The 10 um samples were profiled using the Vizgen's fresh frozen protocol and reagents on the MERSCOPE using a 500 gene pan-neuron panel. The 300um slice was used to perform transcriptome-wide RNA profiling as well as genome-wide single-cell methylation analysis using the ScaleBio scRNA and scMethylation kits. Pipeline outputs from both platforms were then integrated and analyzed. Our analysis revealed clean and high-quality data obtained from the ScaleBio scRNAseq, Scalebio scMethylation, and spatial transcriptomics datasets. Using overlapping genes found on both platforms the spatial and single-cell transcriptome datasets were integrated. Notably, mapping of the unbiased clusters identified in the scRNA dataset based on this integration revealed spatially distinct localization of these clusters. Next, genes present in both panels were used to examine the accuracy of our integration approach. Importantly, we observed high similarity between the expression of genes observed in the Vizgen data with the imputed location of the same gene in the scRNA data. Next we examined the single-cell methylation data, using differentially methylation regions to identify different cell types in the data. Finally the genome-wide single-cell methylation profiles were mapped onto spatial transcriptome data showing the underlying epigenomic states of these populations.Our study, one of the first to integrate these modalities, demonstrates the ability to integrate spatial, single-cell transcriptomic, and epigenetic data for comprehensive spatial analysis, providing valuable insights into the spatial organization of gene expression as well as the spatial regulatory landscape within complex tissues such as the mouse brain.

Session Title: Omics Technologies Poster Thursday Session

Board 1145T: Interplay of genetic and environmental perturbations on the human plasma proteome

Authors:

K. Tsuo[1,2], A. M. Argentieri[2,1], D. Gadd[3,4], M. Kurki[1,2,5], Z. Zheng[1,2,5], D. Baird[6], R. Marioni[3,4], C. Foley[3,4], H. Huang[2,1], B. Sun[6,7], C-Y. Chen[8], M. Daly[2,1,5], A. R. Martin[2,1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Optima Partners, Edinburgh, United Kingdom, [4]Univ. of Edinburgh, Edinburgh, United Kingdom, [5]Univ. of Helsinki, Helsinki, Finland, [6]Biogen Inc., Cambridge, MA, [7]Univ. of Cambridge, Cambridge, United Kingdom, [8]Biogen, Inc., Cambridge, MA

Abstract:

Thousands of associations between plasma protein levels and incident disease have been discovered, and prediction models that include proteomic associations with disease tend to outperform clinical models. To understand underlying mechanisms that may be driving these protein-disease associations, we aimed to determine whether the effects of these proteins on disease may be explained by either 1) causal associations or 2) environmental effects. Previously, ~3k plasma proteins were measured in ~47k individuals in the UK Biobank (UKB), and >3k associations were detected between protein levels and time-to-onset for 19 diseases. In this study, we first employed Mendelian Randomization (MR) using cis-pQTLs from UKB for these disease-associated proteins and publicly available GWAS of the disease outcomes. We identified 24 significant protein-disease pairs, indicating that these GWAS risk loci impact disease risk through protein level changes. MR support was significantly enriched in autoimmune diseases.
Next, we assessed whether proteins associated with disease but lacking putative causal support may be byproducts of the effects of environmental exposures. We focused on smoking due to its known large effects on disease. We found that the plasma proteome is highly perturbed by smoking: 50% of the ~3k plasma proteins measured in UKB participants are significantly associated with smoking status (Bonferroni $p < 1.71e-5$). Of the disease-associated proteins that had pQTL instruments for MR (N=505), the majority (N=302, 60%) are significantly associated with smoking and have no evidence of causal association with disease ($p > 0.005$ in MR). To quantify the value of smoking-associated proteins for incident disease prediction, we performed feature selection on this subset of proteins using LASSO. This protein score predicts smoking status with high accuracy in a held-out set in UKB (AUC=0.95) and replicates in another biobank, FinnGen (AUC=0.87). Further, it captures time since smoking in previous smokers and pack-years in current smokers, and predicts incident lung cancer (C-index [SE] = 0.861 [0.009]) and emphysema (0.853 [0.010]) better than self-reported smoking status (0.843 [0.008] and 0.832 [0.011], respectively). Altogether, our study disentangles disease-associated proteins into putatively causal determinants of disease vs. environmental biomarkers. We illustrate that the utility of the plasma proteome in predicting disease is most likely due to its ability to capture important effects of environmental risk factors, providing a critically needed biomarker for elucidating molecular mechanisms of differences in disease risk among populations.

Session Title: Omics Technologies Poster Thursday Session

Board 1146T: JARVIS 2.0: A non-coding variant effect predictor based on 500,000 UK Biobank WGS samples that surpasses existing benchmarks for prioritisation of common variants ⭐

Authors:

A. Tsakiroglou[1], M. Karpinski[1], P. Martin-Gonzales[1], M. Monti[1], Q. Wang[2], S. Petrovski[1,3], D. Vitsios[1]; [1]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, [2]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Waltham, MA, [3]Dept. of Med., Univ. of Melbourne, Austin Hlth., Melbourne, Victoria, Australia

Abstract:

It is well established that non-coding DNA regions play an important role in gene regulation and disease development. Existing computational variant effect prioritisation (VEP) tools have shown acceptable performance in rare and germline variants in non-coding regions. However, establishing VEP tools able to prioritize both rare and common variants in terms of pathogenicity likelihood has historically proven more challenging. To address this, we propose JARVIS 2.0, a Random Forest-based model for non-coding variant prioritisation, which extends our previously developed JARVIS score (Vitsios et al., 2021): i) We now incorporate an expanded set of high quality functional annotations, chromatin states and methylation assay data produced by the ENCODE and Roadmap Epigenomics consortia, and ii) we introduce an updated genome-wide variation intolerance score (gwRVIS) calculated from ~500,000 individuals from the UK Biobank. We have also further optimized and studied the impact of including individuals from multiple ancestries in the construction of the new gwRVIS score, showing that balanced ancestry representation and incorporating context from longer DNA sequence windows results in more meaningful and discriminative intolerance scores.
We evaluated the performance of JARVIS 2.0 in a diverse set of eight external benchmarks including rare germline variants, variants associated with dominant and recessive Mendelian traits, rare somatic variants, common regulatory variants from curated expression quantitative trait locus (eQTL) data and common disease associated variants from genome-wide association studies (GWAS). Compared to well-established VEP tools (DEEPSEA, FINSURF, LINSIGHT, ncER and CADD), our proposed model showed superior performance in all test sets consisting of common variants, achieving a ROC AUC=0.909 for common regulatory eQTL variants and AUC=0.725 in GWAS hit SNVs from GWAS catalog. For rare somatic variants from COSMIC, Jarvis 2.0, LINSIGHT and ncER achieved similar top performance, surpassing all other scores. Finally, JARVIS 2.0 achieved consistently high performance across all test sets consisting of rare germline or Mendelian variants (AUC > 0.921).

We anticipate that JARVIS 2.0 will help us enhance our understanding of the non-coding genome and also aim to incorporate information from additional functional annotations as well as DNA sequence context to improve prediction performance.

Session Title: Omics Technologies Poster Thursday Session

Board 1147T: Leveraging deep neural networks to identify single-cell gene expression profiles that characterize Parkinson's disease

Authors:

M. Fiorini, J. Li, E. Fon, S. Farhan, R. Thomas; McGill Univ., Montreal, QC, Canada

Abstract:

Single-cell RNA (scRNAseq) is a promising technology for characterizing pathological interactions in disease by facilitating global gene expression profiling at the level of a single cell. However, scRNAseq-based differential gene expression analysis is impeded by sparse gene counts, technical noise, limited biological samples, and the burden of multiple hypothesis testing. Deep neural networks (DNNs) are machine learning models that can detect complex patterns in the data to predict classifications; thus, DNNs could leverage single-cell transcriptional profiles as input to predict the disease status of individual cells. Local surrogate models can then be used to decode the DNNs to determine the most important gene expression profiles for classifying a particular cell as originating from a diseased or healthy individual. In this work, we introduce a novel framework that utilizes scRNAseq and DNNs to identify disease-associated, cell-type specific gene expression profiles that characterize Parkinson's disease (PD). PD is the second most common neurodegenerative disease, characterized by the progressive loss of dopaminergic neurons in the substantia nigra pars compacta, located in the midbrain. Application of our framework to a public data set of post-mortem midbrain tissues from six individuals with PD and eight healthy controls sequenced at a single-nuclei resolution revealed remarkably high disease classification accuracy across seven distinct cell types comprising the tissue (mean balanced accuracy = 0.987), including astrocytes, dopaminergic neurons, endothelial cells, microglia, non-dopaminergic neurons, oligodendrocytes, and oligodendrocyte precursor cells. Applying Local Interpretable Model-Agnostic Explainer (LIME) to decode the DNNs identified the genes whose expression patterns were most important for classifying cells by disease status. Re-training the DNNs with the top LIME-identified genes revealed high disease status classification accuracy (mean balanced accuracy = 0.941), nominating cell-type specific gene sets that characterize PD. Future work will interrogate the LIME-identified genes sets for their enrichment in known biological processes, their variant burden in PD, and their

ability to stratify cells according to disease status in different datasets characterizing individuals with PD and healthy controls. Following our proof-of-concept analysis in PD, we will assemble an open-source computational pipeline that facilitates the use of DNNs to identify gene expression profiles across diseases.

Session Title: Omics Technologies Poster Thursday Session

Board 1148T: Lifespan cellular dynamics and disease associations in the human prefrontal cortex ⭐

Authors:

H. Yang[1], T. Clarence[1], M. Scott[2], D. Lee[1], J. Bendl[1], F. Prashant[1], C. A. McClung[2], J. Fullard[1], G. Hoffman[1], K. Girdhar[1], P. Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Understanding the cellular dynamics in the human brain across the lifespan is crucial for elucidating the development and progression of various diseases. To achieve this, we conducted single-nucleus RNA-seq (snRNA-seq) profiling of 284 postmortem prefrontal cortex (PFC) samples from individuals spanning neonatal to late adulthood stages. This data was integrated with published snRNA-seq data, including fetal donors, resulting in a comprehensive dataset of 1,452,103 nuclei from 311 individuals. Utilizing the Uniform Manifold Approximation and Projection of Maturation (UMAT), we visualized the developmental trajectories of diverse brain cell types. Our analysis of astrocytes revealed distinct developmental patterns for fibrous and protoplasmic astrocytes, with fibrous astrocytes enriched in white matter and protoplasmic astrocytes enriched in gray matter; the latter matured later and was more significantly associated with migraines. Microglia development showed stable enrichment in neurological and immunological traits. Oligodendrocyte development exhibited a transition from psychiatric trait enrichment in oligodendrocyte precursor cells (OPCs) to Alzheimer's disease in mature oligodendrocytes. Both excitatory and inhibitory neurons showed significant associations with psychiatric diseases from early developmental stages, with unique enrichment patterns for Tourette's syndrome and obesity, which become relevant after early development. Furthermore, dynamically expressed genes module analysis showed that developmental genes in all brain cell types were linked to psychiatric traits, while senescent genes were associated with immunological traits in glial cells (astrocytes, microglia, oligodendrocytes) and psychiatric traits in neurons (excitatory and inhibitory). These findings underscore the importance of

understanding cellular and molecular mechanisms underlying brain development and aging, providing insights into disease susceptibility across the human lifespan.

Session Title: Omics Technologies Poster Thursday Session

Board 1149T: Long-read sequencing enables structural variant discovery and identification in haptoglobin.

Authors:

Y. Yang[1], A. Griswold[2], T. Starks[3], G. Byrd[4], L. Adams[2], E. Martin[5], M. Pericak-Vance[6], J. Haines[7], W. Bush[7]; [1]Case Western Reserve Univ., Cleveland Heights, OH, [2]Univ. of Miami, Miami, FL, [3]Wake Forest Sch. of Med., Winston-Salem, NC, [4]NC A&T State Univ, Greensboro, NC, [5]Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Univ. of Miami Miller Sch. of Med, Miami, FL, [7]Case Western Reserve Univ, Cleveland, OH

Abstract:

Complex structural variants (SV) in genes such as **haptoglobin (HP)**, FMR1, and CYP2D6 have important consequences for disease risk and drug metabolism. Accurate typing of these highly polymorphic genes in a large sample depends on using cumbersome PCR-based methods to build imputation reference panels. *Long read (LR) sequencing has the potential to provide more rapid and accurate SV imputation panels, and for novel SV discovery. We demonstrate this potential with a focused study of the HP gene, where we used LR and short read whole genome sequencing (WGS) to build improved imputation panels for haptoglobin in African-descent populations. We also discover a novel HP1 subtype previously undetected by traditional means.*
HP SVs have been associated with many diseases including cardiovascular and neurological disorders. HP contains a tandem two-exon repeat, where exons 3 and 4 are sequence-similar to exons 5 and 6, leading to alleles containing 1 (HP1) or 2 (HP2) copies of this exon pair. Evidence suggests HP2 is the ancient structural form, with recurrent deletions throughout human evolutionary history leading to subtypes HP1F (exons 3 & 4) and HP1S (exons 5 & 6) distinguished through gel migration speed. However, studies have so far predominantly focused on HP in European populations imputed from array-based genotyping.
We analyzed long-read and WGS data from 32 participants from the African American cohort of the Alzheimer's Disease Sequencing Project. Haptoglobin genotypes were called by manual inspection of aligned long-reads. We used the TOPMed server to phase WGS variants and Minimac4 to build a reference imputation panel for calling HP1/HP2. Imputation accuracy was evaluated via leave-one-out cross-validation. We were able to impute HP status from multi-SNP haplotypes in the surrounding region with high accuracy (r2 > 0.95). We achieved higher

accuracy with WGS compared to array-based SNP sets, but the array-based panel maintained accuracy comparable to existing panels.
We also created WGS-based imputation panels for HP1 subtypes with r2 > 0.90. Long-read data resolves a third HP1 subtype from a deletion of exons 4 & 5 within the ancestral HP2 allele, resulting in a combination of exons 3 & 6. This structural variant subtype has not been previously reported and may be unique to African-descent populations. Since exons 3/5 encodes for a multimerization domain and exons 4/6 encodes for the α-chain, HP1 subtypes may have different functional activity.
Long read sequencing successfully reveals previously unrecognized complex structural variants. We demonstrate that LR can complement other sequencing methods in building imputation panels for variant detection.

Session Title: Omics Technologies Poster Thursday Session

Board 1150T: MAJIQ-CLIN: A novel tool for the identification of Mendelian disease-causing variants from RNA-seq data ⭐

Authors:

D. Issakova[1], J. K. Aicher[1], B. Slaff[1], S. Jewell[1], G. R. Grant[1], N. F. Lahens[1], E. Bhoj[2], Y. Barash[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Exome sequencing (ES) is the current standard of care for patients with suspected Mendelian genetic disorders; however, the diagnostic rate is only 31%. Traditional ES analysis methods have difficulty incorporating differences in RNA splicing - a highly regulated process in which different segments of pre-mRNA are joined together resulting in different mature mRNA transcripts. It is estimated that 15-50% of human pathogenic variants alter splicing, with numerous cases of splice-altering variants being causal for known Mendelian disorders. Several recent works pointed to the promise of aiding clinical diagnostics by detecting aberrant splicing from patients' RNA sequencing, thus raising the need for a reliable tool for this task.
For a tool to be considered reliable for detecting splicing aberrations from RNA-Seq, it must accurately detect splicing changes in a patient sample while accounting for factors such as uniqueness to the patient, aberration type, change magnitude, and confounding factors. It should reliably agree with solved cases ranking the causal variation as high as possible and use resources (memory and runtime) realistic for a clinical setting. In recent years a few tools were developed to address this need, specifically LeafCutterMD and FRASER. While both served as

good proof of concepts for clinical diagnostics using RNA-Seq, our analysis indicates several challenges remain in terms of accuracy, efficiency, and usability.
Here we present MAJIQ-CLIN, a pipeline for detecting splicing aberrations in a given patient's RNA-Seq sample compared to a large cohort of controls. We assess the performance of MAJIQ-CLIN and other tools using synthetic data with spiked in splice variations and demonstrate that MAJIQ-CLIN improves significantly upon the performance of FRASER and LeafCutterMD across different types of splicing aberrations and levels of variant inclusion. MAJIQ-CLIN can also reuse quantifications of controls and models of unobserved confounders, resulting in significantly faster performance. To demonstrate MAJIQ-CLIN's performance on real data we used solved cases from Cummings et al., all of which were solved by MAJIQ-CLIN. We are currently in the process of testing MAJIQ-CLIN on additional datasets from our collaborators at the University of Southampton, the University of Washington, and the Undiagnosed Diseases Network. Given the significant improvement in performance offered by MAJIQ-CLIN we hope that clinical geneticists will now be able to incorporate RNA-Seq into their analysis pipeline and improve molecular diagnosis rates for the patients and their families.

Session Title: Omics Technologies Poster Thursday Session

Board 1151T: Maternal and Fetal Genetic Associations between the Circulating Proteome and Preeclampsia ⭐

Authors:

J. Hu[1], C. CAO[2], B. Ferolito[3], K. Gravel-Pucillo[4], M. Maher[5], J. Valliere[5], A. Pereira[6], J-P. Casas[7], R. Saxena[8], K. Gray[9]; [1]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [2]MGH, Boston, MA, [3]VA Boston, Jamaica Plain, MA, [4]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Harvard Med. Sch., Boston, MA, [7]Novartis BioMed. Res., Cambridge, MA, [8]Mass Gen. Hosp., Harvard Med. Sch., Boston, MA, [9]Univ. of Washington, Seattle, WA

Abstract:

Preeclampsia (PE), defined as new-onset hypertension after 20 weeks gestation with proteinuria or other end organ damage, is the most severe type of hypertensive disorders of pregnancy, which is a group of serious pregnancy complications and a leading cause of maternal and neonatal morbidity and mortality. PE is heritable, with contributions from both the maternal and fetal genomes. However, the etiology of PE is not fully understood, leading to the lack of effective preventive and therapeutic strategies. To better understand the pathophysiology of PE, we examined the genetic associations between the circulating proteome and PE based on

published maternal and fetal GWAS of PE using Mendelian randomization analyses. Instrument variables for proteins were selected based on *cis*-pQTLs identified by published GWAS of plasma proteomics in the deCODE (discovery; n=36,000; 6,954 instruments across 1,888 genes for 1,914 unique proteins), the Fenland study (replication; n=10,708; 2,881 instruments across 1,510 genes for 1,510 unique proteins), and the Atherosclerosis Risk in Communities (ARIC) study (replication; n=7,213; 1,612 instruments across 1,594 genes for 1,594 unique proteins). Proteomic profiling in all three studies was performed using the SOMAscan platform. Further replication was conducted based on genetic instruments of circulating proteins profiled using the Olink platform in the UK Biobank (n=54,219; 1,955 instruments across 1,954 genes for 1,954 unique proteins). In the discovery analysis using maternal GWAS data, we identified 10 proteins with significant causal associations with PE (FDR<0.05). Six of these proteins were replicated in both Fenland and ARIC (p<0.05 with the same direction of association) and had the same direction of association in UK Biobank. Then, in the analysis using fetal GWAS data, we applied the same analytical approach, and identified and replicated three proteins that had significant causal associations with PE. Interestingly, the top proteins identified from the maternal (PZP, pregnancy zone protein) and fetal (PSG4, pregnancy specific beta-1-glycoprotein 4) analyses have a high level of production in pregnancy and are involved in immune regulation. Notably, F7 (coagulation factor VII) was the only protein identified and replicated in both maternal and fetal analyses. Our findings reveal proteins that may be causally involved in the development of PE and should be investigated for future therapeutic targeting.

Session Title: Omics Technologies Poster Thursday Session

Board 1152T: Modeling nascent transcription from chromatin landscape and structure

Authors:

M. Pielies Avelli[1], A. Sigurdsson[2], T. Narita[3], C. Choudhary[4], S. Rasmussen[1]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Copenhagen Univ., Copenhagen, Denmark, [2]NNFCPR, Copenhagen Univ., Copenhagen, Denmark, [3]Novo Nordisk Fndn. Ctr. for Protein Res., Copenhagen Univ., Copenhagen, Denmark, [4]Novo Nordisk Fndn. Ctr. for Protein Res., Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

Different cell types and their associated functionalities can emerge from a single genomic sequence when certain regions are expressed while others remain silenced. The study of gene regulation and its potential malfunctioning in different cellular contexts is hence pivotal to

understand both development and disease. We present the Chromatin Landscape and Structure to Expression Regressor (CLASTER), an epigenetic-based deep neural network that can integrate different data modalities describing the chromatin landscape and its 3D structure in their raw format. CLASTER effectively translates them into nascent transcription levels measured by EU-seq at a kilobasepair resolution. Our predictions reached a Pearson correlation with targets above r=0.86 at both bin and gene levels, without relying on DNA sequence nor explicitly extracted chromatin features. The model mostly used the information found within 10 kbp of the predicted locus to perform the predictions, even when a wide genomic region of 1 Mbp was available. Explicit modeling of long-range interactions using multi-headed attention and high-resolution chromatin contact maps had little impact on model performance, despite the model correctly identifying elements in these inputs influencing nascent transcription. The trained model then served as a platform to predict the transcriptional impact of simulated epigenetic silencing perturbations. Our results point towards a rather local, integrative and combinatorial paradigm of gene regulation, where changes in the chromatin environment surrounding a gene shape its context-specific transcription. We conclude that the predominant locality and limitations of current machine learning approaches might emerge as a genuine signature of genomic organization, having broad implications for future modeling approaches.

Session Title: Omics Technologies Poster Thursday Session

Board 1153T: Mosaic Variant Detection from Whole Genome Sequencing

Authors:

B. Lajoie, A. Altomare, M. Ambroso, V. Dien, M. Kellinger, K. Wiseman, S. Billings, S. Kruglyak; Element BioSci.s, San Diego, CA

Abstract:

The detection of mosaic variants has applications in the study of cancer, inherited disease, and aging. Whole genome sequencing (WGS) from blood offers a promising approach for mosaicism detection, but it comes with two requirements beyond what is needed in WGS for standard germline variant calling. First, higher coverage is required to achieve good sensitivity given the low allele frequency (AF) of most mosaic variants. Second, higher accuracy is required to avoid a large number of false positive low frequency calls that stem from library preparation and sequencing errors.
The mosaic truth set for sample HG002 recently released by the National Institute of Standards and the Medical Device Innovation Consortium enables benchmarking of mosaic variant calling.

The truth set contains 85 SNVs determined to be > 5% AF but < 30% AF and not contained in the 4.2.1 variant calling truth set. The release covers nearly 90% of the genome (non-gap bases) and provides regions determined not to have mosaic variants above 2% so that both sensitivity and precision can be calculated.
In this study, we combined PCR-free library preparation with a high-accuracy sequencing chemistry that is capable of generating Q50 accuracy for the majority of bases to sequence HG002 at 70X coverage. We call all 85 mosaic variants with DeepSomatic, and all but 4 pass filters. We also make only 341 additional passing filter calls across the confident regions provided with the truth set.
This work details our methodology, examines sensitivity and precision as a function of sequencing coverage, and investigates the nature of the additional calls. Our findings underscore the potential of high-accuracy WGS for reliable mosaic variant detection.

Session Title: Omics Technologies Poster Thursday Session

Board 1154T: Multi-year follow-up analysis of infants and children post-rapid whole genome sequencing

Authors:

L. Tobin[1], E. Sanford Kobayashi[1], E. F. Juarez[1], A. D. Besterman[1], R. Reimers[1], O. Vargas-Shiraishi[2], A. J. Schwarz[2], J. Knight[2], M. N. Bainbridge[1]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Children's Hosp. of Orange County (CHOC), Orange, CA

Abstract:

Much data exists on the short-term clinical utility and outcomes for critically ill infants and children who have received rapid whole genome sequencing (rWGS), but long-term follow-up is lacking. In this study, a physician manually reviewed the electronic health records (EHR) of 67 critically ill infants and children across two institutions who received rWGS 6-8 years ago. The diagnostic rate of this group was 46%, and the average length of follow-up was 5.4 years. Patients with a genetic diagnosis had a longer average follow-up time compared to those without a diagnosis (5.9 years vs 4.8 years, p=0.026). Undiagnosed patients averaged 1.7 readmissions per follow-up year compared to 2.5 for diagnosed patients; this difference was not statistically significant. Children with a genetic diagnosis had more subspeciality appointments per follow up year than those without a diagnosis (9.4 vs 6.9, p=0.036). Notably, 27% of patients had additional genetic testing ordered by a clinician post-genomic sequencing, including exomes, chromosomal microarrays and gene panels; no new diagnoses emerged from this additional testing. Rates of additional genomic testing orders post-rWGS were consistent across

the diagnosed and undiagnosed groups. Two thirds (66%) of the cohort had a documented new clinical phenotype in the EHR during their follow-up period. There were 7 patients with a formal clinician-requested reanalysis of their rWGS. One of these cases yielded a new genetic diagnosis: a pathogenic *de novo* variant in *IGF2* (Russell-Silver syndrome). Reanalysis of the patients using the new phenotypic data yielded four new candidate diagnoses: *ANK2* (cardiac arrhythmia), *CSNK2B* (Poirier-Bienvenu neurodevelopmental syndrome), *TRPM3* (Neurodevelopmental disorder with hypotonia, dysmorphic facies, and skeletal anomalies), and *SYNE2* (Emery-Dreifuss muscular dystrophy). The mortality during the follow up period was 9% and 10 patients were listed for solid organ transplant. Both diagnosed and undiagnosed patients continued to be high healthcare users in the years post-rWGS, though diagnosed patients received significantly more specialist care. In the future we plan to scale this study to include all sequenced patients. To feasibly identify and extract new post-rWGS EHR information, we plan to use natural language processing (NLP) tools. This data will be utilized by a machine learning algorithm trained to identify patients who are likely to have a new genetic diagnosis. This initial study found individuals who can be used in the training of the algorithm and validation of the NLP results.

Session Title: Omics Technologies Poster Thursday Session

Board 1155T: MuVEH and mitoMuVEH improve discovery of genetic variation from single cells

Authors:

M. Ransom[1,2], K. L. Engel[2], B. M. Stevens[1,2], C. T. Jordan[1], A. Gillen[1,3]; [1]Div. of Hematology, Univ. of Colorado Sch. of Med., Aurora, CO, [2]RefinedSci., Aurora, CO, [3]Rocky Mountain Regional VA Med. Ctr., Aurora, CO

Abstract:

Accurately capturing the genetic underpinnings and clonal structure of heterogeneous malignancies at single-cell resolution is critical for understanding the mechanisms of drug response and disease evolution. Here, we introduce an accessible, multiplexable, targeted mutation enrichment approach and analysis pipeline called MuVEH (Multiplexed Variant Enrichment by Hybridization) that increases the resolution of variant detection in scRNA-seq analysis. This technique allows the targeting of more than 100 genes of interest in up to 12 single-cell sequencing libraries simultaneously and is completely agnostic to the nature of these libraries. When applied specifically to the mitochondrial chromosome ("mitoMuVEH"), this technique can also be used to infer clonal relationships between individual cells. Final libraries

from any single-cell sequencing technique can be used with MuVEH, and the only requirement for mitoMuVEH is capture of mitochondrial RNA or DNA. We applied these approaches to primary bone marrow specimens from acute myeloid leukemia (AML) patients collected at diagnosis and after relapse following Venetoclax+Azacitidine (Ven/Aza) therapy. We found that MuVEH efficiently drove targeted sequences to near-saturation, extracting virtually all information present in the libraries for these targets, even at low sequencing depths. Notably, MuVEH accomplishes this without special handling or primer spike-in during cell capture, allowing for the enrichment of any existing single-cell library. We then used these data to demonstrate clonal evolution and changing mutational burden in response to treatment at single-cell resolution in these patients. In addition to multiplexed enrichment of indexed sequencing libraries, this technique can also be applied to individual full-length single-cell cDNA libraries and paired with long-read sequencing. This allows extraction of information about complex mutations such as rearrangements, translocations, and duplications, as well as alternative splicing. MuVEH probe sets can also be repurposed to enrich bulk sequencing libraries for the targets of interest, further increasing their utility and value. Ultimately, these approaches have the potential to extract additional biological insights from precious patient samples and provide insight into the contributions clonality and genotype have during disease progression at single-cell resolution.

Session Title: Omics Technologies Poster Thursday Session

Board 1156T: needLR: A structural variant filtering and prioritization tool for long-read sequencing data

Authors:

J. Gustafson, S. Gibson, N. Damaraju, M. P. Zalusky, J. Goffena, S. H. Storz, K. Hoekzema, The 1000 Genomes ONT Sequencing Consortium, E. Eichler, D. Miller; Univ. of Washington, Seattle, WA

Abstract:

More than half of individuals with a suspected Mendelian condition remain undiagnosed after a comprehensive clinical evaluation, limiting their ability to benefit from precision therapies, N-of-1 treatments, and to receive accurate counseling regarding prognosis or recurrence risk. While there are several reasons for this low solve rate, one salient factor is that it is difficult to reliably identify and resolve structural variants (SVs; genomic variants ≥50 bp) using traditional short-read sequencing-based approaches. Advancements in long read sequencing (LRS) bridge that gap by allowing for accurate and comprehensive identification of SVs, resulting in an increase of

the average number of SV calls per genome from ~10,000 to ~25,000. The identification of more SVs per genome has introduced the need for a comprehensive, LRS-based database of population-scale SVs as well as tools for filtering and prioritizing candidate SVs in unsolved individuals. Here, we present needLR, a tool which uses SV calls from samples from the 1000 Genomes Project sequenced to high coverage to assign population allele frequencies to SVs detected by LRS and performs downstream SV filtering and prioritization. needLR incorporates an SV merging tool, JASMINE (Jointly Accurate SV Merging with Intersample Network Edges) to identify analogous SVs between samples, both within the 1KGP cohort and between the 1KGP cohort and the query sample. needLR then ranks each SV in the query sample by how often it is seen in the 1KGP cohort, creating a frequency-based prioritization framework. Finally, needLR annotates each SV in the query sample with medically and genomically-informative information such as if the SV intersects with a protein coding gene/exon and if that gene is associated with an OMIM phenotype, as well as if the SV is proximal to repetitive DNA sequence, segmental duplications, centromeres, and telomeres. In practice, needLR has successfully identified pathogenic SVs that were not discovered by short-read sequencing of unsolved individuals. needLR is available as a downloadable, open-source command line tool and as an interactive web application.

Session Title: Omics Technologies Poster Thursday Session

Board 1157T: Neural transcription factors identified driving brain organoid maturation

Authors:

S. Gonzales[1], X. Liu[2], S. Xie[3], J. Gu[1], M-Y. Lee[4], X. Wang[1]; [1]Florida Intl. Univ., Miami, FL, [2]Natl. Day Sch., Beijing, China, [3]Univ. of Michigan, Ann Arbor, MI, [4]Univ. of North Texas, Denton, TX

Abstract:

Human brain organoids (HBOs) are simplified versions of human brain tissues developed from pluripotent stem cells. HBOs can be used to model various neurodevelopmental and neurodegenerative diseases, study disease pathology in a controlled environment and observe how these diseases progress at the cellular and molecular levels. HBOs provide a more accurate representation of human brain development compared to animal models. Nonetheless, insufficient maturity of HBOs is one of challenges in using HBOs for brain disease modeling including neurodegenerative diseases. To identify genes that regulate the maturity of HBOs, we performed differential gene expression (DEG) analysis. This was followed by the prediction and annotation of enriched transcription factors (TFs) between fetal neural tissue and healthy HBOs.

Our goal was to assess transcriptional differences and identify TFs that are crucial for neural differentiation but are under-expressed in brain organoids.
753 significant DEGs (Bonferroni corrected alpha=$1.9x10^{-6}$) were identified with the Wilcoxon Rank Sum (WRS) test on 26,624 genes between 8 normal HBOs integrated from two GEO datasets (GSE195505 and GSE201897) and 57 fetal brain tissue from the BrainSpan Developmental Transcriptome. To identify TFs driving these 753 DEGs, we employed MAGIC, a python tool that uses ChIP-seq data to identify significant TFs in a given gene list (FDR-adjusted $p < 0.05$) that are driving changes in biological states. Furthermore, significantly enriched TFs were annotated with Gene Ontology (GO) knowledge base.
Two TFs (*EZH2* and *SUZ12*) identified could potentially serve as targets for improving organoid maturation. *EZH2* and *SUZ12* are involved in cell differentiation and proliferation, as well as neurodevelopment: *EZH2* (MAGIC Enrichment Score (MES)=3.99, FDR-adjusted $p = 1.5x10^{-5}$) is involved in the regulation of gliogenesis (GO:0014013) and dendrite development (GO:1900006), cerebellar cortex (GO:0021695) and hippocampus development (GO:0021766), stem cell differentiation (GO:0048863), and neurogenesis regulation (GO:0050767). *SUZ12* (MES=4.39, FDR-adjusted $p=1.2x10^{-5}$) is involved in regulation of cell differentiation and proliferation, particularly for oligodendrocytes. *OLIG3* (WRS $p = 8.3x10^{-9}$) was significantly differentially expressed between healthy HBOs and fetal tissue. It is involved in neuron differentiation and is a transcriptional target for both *SUZ12* and *EZH2*. A verification study using 6 tissues (GSE160690, GSE146878, and GSE78566 combined) and 10 HBOs (GSE222365, GSE155382, and GSE146878 combined) confirmed our findings.

Session Title: Omics Technologies Poster Thursday Session

Board 1158T: Novel Drivers of α1A Adrenergic Receptor-mediated Heart Failure Identified through Correction for Cell-Type Shifts in Bulk Cardiac Transcriptomes

Authors:

B. Gural[1], L. Kirkland[1], S. Peyton[1], M. Rosa-Garrido[2], S. Swift[3], M. Patterson[3], B. Jensen[1], C. Rau[4]; [1]UNC - Chapel Hill, Chapel Hill, NC, [2]Univ. of Alabama at Birmingham, Birmingham, AL, [3]Med. Coll. of Wisconsin, Milwaukee, WI, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Background: Advances in single cell sequencing have highlighted the importance of cell-type composition in understanding cardiac phenotypes and disease progression. Heart failure and

other cardiac conditions often involve changes in cell-type composition, which can obscure true gene expression patterns in bulk RNAseq data. Our simulations showed that even minor changes in cell type proportions, such as a 10% reduction in cardiomyocytes—a scenario less severe than those occurring during myocardial infarction—can cause 25% of genes to be found as differentially expressed.
Methods: We developed and tested a novel computational method to estimate cellular proportions from bulk RNA sequencing of cardiac tissue and use them to inform transcriptomic analysis. Our study utilized a cardiomyocyte-specific α1A-adrenergic receptor (α1A-AR) knockout mouse model to investigate cardiac responses to myocardial infarction (MI). We compared gene expression in the left ventricles of wild-type (WT) and α1A-AR knockout mice, post-MI or after sham surgery. We measured gene expression in infarct border zones using bulk RNA sequencing, estimated the underlying cellular contributions to total expression, then validated our findings with RNAscope and IHC.
Results:We found that α1A-AR KO mice experience significant loss of cardiomyocytes and increase in fibroblasts following myocardial infarction. Initial analyses revealed many changes in gene expression attributed to MI or genotype, further analysis adjusted for cell type proportions demonstrated that these changes are mediated by shifts in cellular composition. Post-adjustment, we found considerable loss of association between MI and hallmark processes of the altered cell types, with concurrent increases in association of α1A-AR KO x MI with terms related to coordinated heart function and cardiomyocyte sub-states.
Conclusions: Our approach distinguishes between changes in gene expression driven by alterations in cell state versus those due to shifts in cell abundance, offering a significant advancement in the analysis of bulk RNAseq data. This study improves our understanding of the role of cardiac cell types in cardiovascular disease and lays the foundation for more nuanced transcriptomic studies in cellularly heterogeneous tissue.

Session Title: Omics Technologies Poster Thursday Session

Board 1159T: Optimization and quality control of library preparation in bulk and single-cell ATAC-sequencing

Authors:

B. Strauch[1], W-J. Kim[2], **A. Dorn**[1]; [1]Agilent Technologies, Waldbronn, Germany, [2]Sch. of Dentistry, Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

The assay for transposase-accessible chromatin with sequencing (ATAC-Seq) is currently the most widely used technique for determining chromatin accessibility and relies on the activity of the hyperactive DNA transposase Tn5. ATAC-Seq libraries aim to have a nucleosomal fragmentation pattern that is crucial to achieving reliable outcomes. Here, we investigated and optimized various conditions for both bulk ATAC-Seq and single-cell ATAC-Seq for improving ATAC-Seq library preparations. Bulk ATAC-Seq library preparation was performed according to the original publication[1]. Single-cell ATAC-Seq library preparation was performed using the Chromium Next GEM Single Cell Multiome ATAC kit (10x Genomics). Library quality control (QC) was performed using the Agilent 4150 TapeStation System with the High Sensitivity D5000 ScreenTape. For bulk ATAC-Seq, a custom protocol was established by optimizing total cell number, lysis conditions and Tn5 concentration that shows equal to superior nucleosomal patterning of the library than comparable commercially available kits. For single-cell ATAC-Seq, methods were tested to assess optimal lysis conditions for efficient isolation of individual nuclei. Determining the amount of nuclei RNA turned out to be a reliable indicator to confirm optimal lysis conditions and high quality libraries. The findings from this study offer insights into improving and how to assess the efficiency and quality of ATAC-Seq library preparations for both bulk ATAC-Seq and single-cell ATAC-Seq.

Session Title: Omics Technologies Poster Thursday Session

Board 1160T: Optimized workflows for custom targeted enrichment, from oligo probe design to NGS data analysis

Authors:

S. Angelow, J. Nery, Y. Shin, S. Chen, Y. Han, N. Shah, P. Prashar, S. Kuersten, B. Frey; Illumina, San Diego, CA

Abstract:

Targeted enrichment allows sequencing of a subset of genes or genomic regions of interest by hybridization to sequence-specific biotinylated DNA probes and isolation with streptavidin linked magnetic beads. This method allows researchers the flexibility to sequence a large range of targets, from whole exomes to just several genes or variant hotspots. Successful target enrichment relies on a high-quality probe pool and library prep workflow to provide high target coverage uniformity and to minimize off-target reads and dropouts. Here we describe the end-to-end workflow from panel design in DesignStudio, to secondary analysis in BaseSpace Sequencing Hub for custom DNA and RNA targeted enrichment panels. DNA panels targeting a range of panel sizes were designed using Illumina DesignStudio, an online tool for custom assay

design. DesignStudio uses proprietary design algorithms to design and select 120bp, double-stranded probes that provide high on-target enrichment and uniformity. DNA Panels were designed with different probe densities to determine impact on performance. In addition, DNA panels were designed as either standalone panels or as a spike-in to the Twist Bioscience for Illumina Exome 2.5 Panel. Spike-in panels can be used to enhance performance of existing regions or supplement additional content into the panel. RNA panels were designed by the Illumina Concierge team, which used an exon informed design strategy that allows for detection of novel isoforms or fusion transcripts. Libraries were prepared and enriched using Illumina DNA Prep with Enrichment and Illumina RNA Prep with Enrichment, for DNA and RNA panels respectively. Pooled and enriched libraries were sequenced on NovaSeq 6000 System, using S4 flow cells and 2 x 101 bp read lengths. Analysis was performed in BaseSpace Sequence Hub using the DRAGEN suite of apps. DRAGEN FASTQ Toolkit was used to subsample coverage of each panel to roughly 350x. For DNA panels, DRAGEN Enrichment was used for secondary analysis to determine enrichment metrics and for germline variant calling. For RNA panels, DRAGEN RNA was used for gene expression analysis and fusion detection. Data presented here demonstrates the high performance of targeted enrichment panels with accurate variant calling and sensitive detection of RNA transcripts.

Session Title: Omics Technologies Poster Thursday Session

Board 1161T: Orthogonal validation of sequencing in the clinical setting: Lessons from assessing concordance between whole genome sequencing and genotyping arrays with 133,000 samples

Authors:

M. Gatzen[1,2], S. E. Hofherr[1,2], C. Kachulis[1,2], Y. Gao[1,2], N. J. Lennon[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Broad Clinical Labs, Burlington, MA

Abstract:

Recent innovations in genomics have led to remarkable improvements in efficiency, throughput, and accuracy of sequencing, thereby enabling unprecedented opportunities for research and clinical applications. Whole genome sequencing (WGS) is considered the gold standard in clinical testing, but despite the technological progress it remains imperative to ascertain scientific accuracy, especially in the context of clinical reporting.
Regulatory bodies, most notably the New York State Clinical Laboratory Evaluation Program (CLEP), have implemented important criteria for evaluating the validity of sequencing tests requiring confirmation of results using an orthogonal method of measurement. While this is a

sensible approach for single gene and targeted panels with limited reportable ranges, orthogonal testing can provide a challenge to clinical laboratories due to limitations in throughput or accuracy of established validated assays.
We evaluate the accuracy of WGS as a clinical test by analyzing the concordance of WGS with genotyping array data for an extensive set of samples across various ancestries. By leveraging the latest release of the All of Us Research Program, we analyzed 133,023 unrelated samples, which were sequenced at the Broad Institute following our clinical pipelines, and compared the WGS variant calls with Global Diversity Array (GDA) genotyping data, which was also generated as part of All of Us, functioning as an orthogonal method of measurement. This affords us the ability to evaluate over a billion variant calls, capturing a wide range of genetic diversity and covering an extensive set of clinically reportable regions in the genome. After employing established filtering criteria we find an overall positive predictive value (PPV) of 99.5% across the exome, covering variants in 16,790 genes. We identify specific regions where WGS shows decreased precision while taking into account their clinical relevance. We also recognize potential biases associated with genotyping arrays and consider potential limitations when treating GDA data as truth, and augment our analysis with high-quality variant calls provided by the Genome in a Bottle consortium.
By utilizing the vast amount of data in the All of Us Research Program we establish an understanding of the validity of WGS as a clinical test which can help guide decisions about the importance and suitability of orthogonal validations in genetics. Understanding the strengths and weaknesses of sequencing in a clinical setting is instrumental for defining regulations for safe and effective tests and will ensure increased access for patients to the benefits of genetic testing as a standard of care.

Session Title: Omics Technologies Poster Thursday Session

Board 1162T: Piloting the 48-hour sequencing workflow for BadgerSeq: a decentralized model for ultra-rapid long-read diagnostic genome sequencing

Authors:

S. Meyn[1], S. Z. Davis[1], D. Pavelec[1], B. Ross[1], J. Heilmann[1], L. Frater-Rubsam[1], X. Shao[1], V. Horner[2], A. Hall[3], B. Webb[4]; [1]Univ. of Wisconsin Sch. of Med., Madison, WI, [2]Univ Wisconsin Madison, Madison, WI, [3]Univ of WI - Madison, Madison, WI, [4]Univ. of Wisconsin- Madison, Madison, WI

Abstract:

&gt100,000 infants are born annually in the United States with a rare genetic disorder/birth defect, representing &gt20% of level IV NICU patients. As prompt diagnosis is critical, rapid genome sequencing is becoming the genetic test of choice in the NICU. Yet, lack of genetic services limits access for infants in regional hospitals, NGS cannot detect many clinically important variants, and results are frequently returned more than 10 days after presentation due to delays in initiating genetics consults, shipping samples, and long sequencing times. *BadgerSeq* addresses these challenges by using a genome sequencing first approach that evaluates all NICU patients daily by text mining the patient's EMR for HPO terms, then input into machine learning-based software which ranks patients for sequencing. Infants whose scores exceed a threshold are offered trio genome sequencing. BadgerSeq's second innovation is the use of local long-read Nanopore sequencing coupled with AI-assisted tertiary variant analysis. Nanopore sequencing enables detection of more SNVs and structural variants than NGS and, unlike NGS, can be used to analyze repeat expansions, resolve pseudogenes, and identify epigenomic changes. We are piloting our sequencing workflow using proband + parent trios from our Undiagnosed Disease Program. We now report the results for our first 10 trios: 3 previously solved cases, as well as 7 cases that remained undiagnosed after NGS genome sequencing. For these trios, library preparation averaged 3-4 hours. Each library was loaded onto 2 flow cells and sequenced for 36 hours, which yielded an average output of 112 Gb/genome, with an average N50 of 7.3 kb. Base calling and alignment was performed on the fly by the PromethION P-24 sequencer, with assembly/variant calling carried out in 4 hours on a local HPC, following a modified Nanopore workflow. Resultant VCF files were uploaded for tertiary cloud-based analysis by Fabric Genomics GEM software, which ranked candidate causal variants after 2 hours of analysis. The combined times for these first 10 cases averaged 46-48 hours from DNA to completion of initial tertiary analysis, including 60-90 minutes for data transfer. This ultrarapid workflow correctly identified the pathogenic SNVs and structural variants in the 3 positive control samples and suggested candidate variants for 3 of the 7 unknowns. Our initial results demonstrate the feasibility of *BadgerSeq's* innovative ultra-rapid sequencing workflow, a workflow that will offer academic and regional hospitals a faster and more comprehensive long-read alternative to the current hub and spoke model of centralized high volume rapid short-read genome sequencing.

Session Title: Omics Technologies Poster Thursday Session

Board 1163T: PRISM: Refraction of Causal Analyses upon Proteomics Data Into Bands of Efficacious Drug Targets

Authors:

J. Kintzle[1], S. Kulm[1], P. Di Domenico[2], **G. Busby**[1], G. Bottà[3]; [1]Allelica, New York City, NY, [2]Allelica, Rome, Italy, [3]Allelica, New York, NY

Abstract:

Drug development is a notoriously slow, expensive and risky endeavor. Although, drugs targeting proteins with strong genetic links to the disease, especially those with high causal likelihood, are more likely to succeed in clinical trials. The convergence of large proteomics datasets and powerful causal inference methods enables the identification of proteins with strong causal evidence for disease that are relatively likely to make successful drug targets. Here, we achieve this convergence of novel, high-powered, efficacious drug target identification in a process we call PRISM. We begin by applying association, mediation and mendelian randomization results to data from the UK Biobank Pharma Project. The mendelian randomization analysis generated 4.23x10-10 total genotype-protein associations, which were then filtered into instrumental variables for the use in 102 total analyses, following MR-STROBE principles. The multiple signals of disease causality per target were condensed into a single prediction of the target's drug efficacy by a cross validation modeling framework. Targets were labeled according to the efficacy of the available drug and indication as provided by the OpenTargets database. We found that the PRISM-based model predictions had an area under the receiver operator curve of 0.854 (95% CI 0.827 - 0.881), when drug efficacy was defined by manual literature curation. Targets whose model prediction was greater than a pre-determined binary threshold were 5.4 times more likely to be approved than a typical drug. When applying the threshold to all possible 55,896 target-disease combinations we find 318 promising, possibly efficacious targets that have never been trialed previously. Considering eczema, both CALCA and S100A14 are both above the efficacy threshold and are known epidermal agonists while IL12B's probability was far below the threshold and was previously unsuccessfully targeted by ustekinumab, a drug that failed in a second clinical trial for lack of efficacy. Integrating the PRISM results into prospective drug developments could result in far more novel drugs that precisely target the proteins which cause disease, increasing clinical trial success, decreasing development cost and benefiting patients suffering from disease.

Session Title: Omics Technologies Poster Thursday Session

Board 1164T: Retrieval Augmented Generation Greatly Improves the Ability of Large Language Models to Precisely Assign Human Phenotype Ontologies to Clinical Data

Authors:

B. Garcia, P. Yelemali, N. Gogate, E. Rivera-Munoz, A. Jolly, M. Dawood, H. Du, J. Posey; Baylor Coll. of Med., Houston, TX

Abstract:

Accurate deep phenotyping of patient clinical data is essential for unraveling the complex relationship between genotype and phenotype in rare genetic diseases and for understanding the functionality of the almost 20,000 protein-encoding genes of the human genome. Central to these goals and to the mission of the Genomic Research to Elucidate the Genetics of Rare (GREGoR) Consortium is the use of Human Phenotype Ontology (HPO) terms, which provide a contextualized, standard vocabulary essential for both human and machine interpretation of phenotypic data. However, the need for refined phenotyping tools applicable to an international referral cohort that lacks a uniform, standardized electronic medical records (EMR) source remains unaddressed.To address this need, we developed a novel application employing Retrieval Augmented Generation (RAG) and Large Language Models (LLMs) to facilitate the precise assignment of HPO terms from medical free text. This user-friendly tool, RAG-HPO, was developed using Langchain and designed to interact with a locally stored LLM to recognize phenotypes and assign HPO terms. RAG-HPO utilizes a synthetic database composed of information collected from the HPO project and publicly available clinical case reports. This database is both shareable and easily modifiable for continual improvement of the tool's capability to accurately recognize phenotypes and assign appropriate HPO terms.We evaluated the capability of RAG-HPO paired with Llama3-70b in assigning HPO terms to a diverse set of published clinical cases from BMJ Case Reports and Oxford Medical Case Reports. These results were then compared to a manually annotated standard as well as outputs from Doc2HPO, ClinPhen, and Llama-70b alone. RAG-HPO demonstrated substantial improvement in precision and recall over other methods. Specifically, RAG-HPO achieved a precision of 0.85, recall of .073, and an F1 score of 0.78. In comparison, Doc2HPO had a precision of 0.67, recall of 0.40, and an F1 score of 0.49, while ClinPhen showed a precision of 0.62, recall of 0.36, and an F1 score of 0.43. At baseline, Llama-3 had much lower performance, with a precision of 0.18, recall of 0.13, and an F1 score of 0.15.RAG-HPO is an easy-to-use tool that ensures the safe and private evaluation of medical free text by clinical and scientific users. It can be downloaded, adapted for use with many LLMs, and improved over time through user feedback. With its increased precision and recall, RAG-HPO is a significant contribution to the field, poised to expedite the exploration of genetic mechanisms underlying rare diseases and catalyze discoveries in genetic research.

Session Title: Omics Technologies Poster Thursday Session

Board 1165T: Saliva DNA can support long read sequencing technologies to improve resolution of high complexity genomics regions

Authors:

M. Tayeb[1], S. Kelly[2], A. Carroll[2], N. Strom[3], E. Smith[3], S. Murugapiran[1], D. Gerhardt[2], A. Halpern[2], C. Davidson[2], B. Le François[1]; [1]DNA Genotek Inc, Ottawa, ON, Canada, [2]Illumina Inc, San Diego, CA, [3]Diversigen Inc, New Brighton, MN

Abstract:

Saliva is an easy to collect, non-invasive specimen which has displaced blood as the primary source for human DNA (hDNA) across many genomics workflows. Long read sequencing (LRS) technologies have gained traction in the field in recent years because of their ability to generate sequencing data from higher complexity genomic regions, which tend to be poorly covered in traditional short read datasets. LRS technologies have enabled the identification of several disease-causing genetic variants, such as an intergenic duplication event in a patient with Kleefstra syndrome. In this study, saliva samples were collected from 4 individuals using Oragene saliva collection devices and the extracted DNA was sequenced using Illumina Complete Long Read (ICLR) technology. Despite a fairly broad DNA size distribution (~2kb to >50kb), saliva samples generated robust libraries and high quality long reads with N50 ranging from 6.5-7.5kb, which was marginally better than N50 values for matching DNA extracted from blood samples (5.3-6.9kb). Phase block N50 values were also higher for saliva DNA processed with ICLR (370kb-400kb) than paired blood samples (310kb-380kb). We observed a strong concordance between SNPs identified in saliva samples vs paired blood samples (~95%), with saliva and blood each contributing ~2.5% unique SNPs in our dataset. Most of these differences were confined to low confidence regions, and the concordance between blood and saliva was ≥99.68% when only high confidence regions were compared. As expected, ICLR significantly improved the coverage of key genomic regions, from genes of the major histocompatiblity complex to important diagnostics markers for human genetic traits and diseases. Together our data demonstrates that salivary DNA is well-suited for LRS applications, particularly ICLR given its lower input requirement compared to other methods. Overall, this further highlights the value and versatility of saliva as a specimen in a wide range of human genomics assays.

Session Title: Omics Technologies Poster Thursday Session

Board 1166T: Sample and Cell-type specific inference from bulk RNA-sequencing data via EPIC-unmix

Authors:

C. Tang[1], Q. Sun[1], X. Zeng[1], X. Yang[2], F. Liu[3], J. Zhao[4], Y. Shen[5], B. Liu[3], J. Wen[1], Y. Li[1]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Univ. of California, San Francisco (UCSF), San Francisco, CA, [3]Natl. Univ. of Singapore, Singapore, Singapore, [4]Univ Florida, Gainesville, FL, [5]Inst. for Human Genetics, San Francisco, CA

Abstract:

Cell type specific (CTS) analysis is essential to reveal biological insights obscured in bulk tissue data. However, single-cell (sc) or single-nuclei (sn) resolution data are still cost-prohibitive for a large number of samples. Thus, computational methods to perform deconvolution from bulk tissue data are highly valuable. We here present EPIC-unmix, a novel two-step empirical Bayesian method integrating reference sc/sn RNA-seq data and bulk RNA-seq data from target samples to enhance the accuracy of CTS inference. We also couple EPIC-unmix with a gene selection strategy to boost CTS inference performance. Specifically, genes are pre-selected based on cell-type marker genes and agreement in expression among sc/snRNA-seq datasets as well as CTS bulk RNA-seq data. We demonstrate through comprehensive simulations across three tissues, EPIC-unmix achieved 4.6% - 109.8% higher accuracy compared to alternative methods. By applying EPIC-unmix to human brain tissue RNA-seq data from the Religious Orders Study/Memory and Aging Project (ROSMAP) and Mount Sinai Brain Bank (MSBB) cohorts, we identified multiple genes differentially expressed between Alzheimer's disease (AD) cases versus controls in a CTS manner for four major brain cell types, including 57.4% novel genes not identified using similar sample size sc/snRNA-seq data, indicating the power of our *in-silico* approach. Among the 42.6% overlapping CTS DEGs, 83%-100% have effect sizes in consistent direction with those from sc/snRNA-seq data, supporting the reliability of our findings. EPIC-unmix inferred CTS expression profiles similarly empowers CTS eQTL analysis. Our results replicated many CTS eQTLs identified in previous studies based on sc/snRNA-seq or cell sorting, with highly consistent (Pearson Correlation Coefficient ranging from 0.792-0.886) effect sizes. Among the novel eQTLs, we highlight a microglia eQTL for AD risk gene *AP3B2,* obscured in bulk and missed by sc/snRNA-seq based eQTL analysis. The variant resides in a microglia-specific candidate cis-regulatory region, forming chromatin loop with *AP3B2* promoter in microglia. Taken together, we believe EPIC-unmix will be a valuable tool to enable more powerful CTS analysis.

Session Title: Omics Technologies Poster Thursday Session

Board 1167T: SCUBA enables consistent visualization of single-cell sequencing data across storage formats and supports the development of more flexible analysis tools

Authors:

J. Desai[1], W. M. Showers[1,2], C. Smith[1,2], C. T. Jordan[2], A. E. Gillen[2,3]; [1]RefinedSci., Aurora, CO, [2]Div. of Hematology, Univ. of Colorado Sch. of Med., Aurora, CO, [3]Rocky Mountain Regional VA Med. Ctr., Aurora, CO

Abstract:

While robust tools exist for the analysis of single-cell datasets in both Python and R, interoperability is limited, and analysis tools generally only accept one object class. Considerable programming expertise is required to integrate tools across package ecosystems into a comprehensive analysis, due to their differing languages and internal data structures. This complicates validation of results and leads to inconsistent visualizations between analysis suites. Conversion between object formats is the most common solution, but this is difficult and error-prone due to the rapid pace of development of the analysis suites and their underlying data structures. To address this, we created SCUBA (Single-Cell Unified Backend API), an R package that implements a unified data access API for all common R and Python single-cell object formats. SCUBA extends the data access approach from the widely used Seurat package to SingleCellExperiment, anndata, and mudata objects. SCUBA also implements new data-specific access functions for all supported object types. Data are returned by SCUBA in a single R data.frame, with requested variables as columns, and cells as rows. We benchmarked SCUBA on sample datasets with varying numbers of cells and found that maximum performance is achieved with anndata objects, despite the use of reticulate. Performance scales well across all supported formats, even for large objects. In addition to performance, SCUBA offers several advantages over object conversion for the visualization and further analysis of pre-processed single-cell data. First, SCUBA extracts only data required for the operation at hand, leaving the original object unmodified. This process is simpler, less error prone, and less memory intensive than object conversion, which operates on the entire dataset. Second, code written with SCUBA can use any supported object class as input, with simple and consistent syntax across object formats. This allows a single analysis script or package to work seamlessly with multiple object types, reducing the complexity of the code and improving both readability and reproducibility. For example, scExploreR, our single-cell interactive browser, utilizes SCUBA to empower users to easily generate identical visualizations from any supported object class. We envision SCUBA as the basis of many such packages. Adoption of SCUBA will ultimately improve collaboration and

reproducible research in single-cell analysis by lowering the barriers between package ecosystems.

Session Title: Omics Technologies Poster Thursday Session

Board 1168T: Sex differences in the human plasma proteome ⭐

Authors:

E. Viippola[1], T. Hartonen[1], T. Nakanishi[1,2,3,4], M. Zhu[1], S. May-Wilson[1], M. Kanai[5,6,7,8], M. Argentieri[6,9,10], R. S. Rodosthenous[1], FinnGen, Biobank Japan, Japan COVID-19 Task Force, H. Namkoong[11], K. Fukunaga[12], K. Matsuda[13,14], Y. Okada[2,3,15,16], M. Daly[1,10], T. Tukiainen[1], X. Shen[17,18,19,20], A. Ganna[1,6]; [1]Inst. for Molecular Med. Finland (FIMM), HiLIFE, Univ. of Helsinki, Helsinki, Finland, [2]Dept. of Genome Informatics, Graduate Sch. of Med., Univ. of Tokyo, Tokyo, Japan, [3]Lab. for Systems Genetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [4]Japan Society for the Promotion of Sci., Tokyo, Japan, [5]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [7]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [8]Ctr. for Computational and Integrative Biology, Massachusetts Gen. Hosp., Boston, MA, [9]Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [10]Broad Inst. of MIT and Harvard, Cambridge, MA, [11]Dept. of Infectious Diseases, Keio Univ. Sch. of Med., Tokyo, Japan, [12]Div. of Pulmonary Med., Dept. of Med., Keio Univ. Sch. of Med., Tokyo, Japan, [13]Lab. of Clinical Genome Sequencing, Dept. of Computational Biology and Med. Sci., Graduate Sch. of Frontier Sci., Univ. of Tokyo, Tokyo, Japan, [14]Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan, [15]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Japan, [16]Lab. of Statistical Immunology, Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka Univ., Suita, Japan, [17]Ctr. for Intelligent Med. Res., Greater Bay Area Inst. of Precision Med. (Guangzhou), Fudan Univ., Guangzhou, China, [18]State Key Lab. of Genetic Engineering, Ctr. for Evolutionary Biology, Sch. of Life Sci., Fudan Univ., Shanghai, China, [19]Dept. of Med. Epidemiology and Biostatistics, Karolinska Inst.t, Stockholm, Sweden, [20]Ctr. for Global Hlth.Res., Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract:

RNA sequencing studies have revealed extensive sex differences in gene expression (GEX), but sex differences in protein expression (PEX) have been less explored.
We analyze sex-biased PEX in 2,919 proteins measured in plasma from 42,979 UK Biobank participants. We found that 69 % of the proteins had sex-specific PEX ($p<2e-5$). Unlike what has been observed for GEX, genes coding these proteins are not more likely to be on the X

chromosome. Genes encoding male and female-biased proteins are more likely to be expressed in the testis and pituitary gland, respectively.
As previously reported, we found a low correlation between GEX and PEX levels. However, we found that genes with the most sex-biased GEX were also more likely to be sex-biased for PEX (r=0.55 for liver GEX). We confirm this relationship in 277 FinnGen participants with both single-cell RNA sequencing and PEX.
We explored sex-specific protein-protein correlations and found 187 protein pairs with large correlation differences between sexes (difference in r>0.3). Notably, 55 % of them involved *PAEP* or *RLN2* proteins, both expressed in female-specific tissues. In the FinnGen individuals with both GEX and PEX, sex-specific correlations were more common for PEX than GEX.
Menopause substantially impacts hormonal regulation in women. We found that 57 % of proteins were differently associated with sex before vs. after menopause and that, on average, male-biased proteins were more similar between men and post-menopausal women compared to women before menopause.
We developed a machine learning model to predict sex based on proteomic data and explore intra-sex variability. We identified 331 individuals showing PEX patterns resembling the opposite sex and confirmed these were not due to technical artifacts or sample swaps. Sex-proteome-atypical individuals were more likely to have sex aneuploidies, particularly XXY (p<1e-16). Females with male-typical proteomic profiles were taller (RR=1.2, p=8e-16) and had higher testosterone levels (RR=1.1, p=7e-131), while males with similar profiles were shorter (RR=0.9, p=0.011) and had lower testosterone levels (RR=0.7, p=2e-29).
The proteomics-based sex predictor was replicated in Biobank Japan (N=2,553, AUC=0.995) and Japan COVID-19 Task Force (N=1,367, AUC=0.976) where we found that females with male-typical proteomic profiles were more likely to have severe COVID-19 (OR=3.0, p=4e-5).
These findings reveal complex patterns of sex-biased proteomic variability, providing insights into sex-specific health risks.

Session Title: Omics Technologies Poster Thursday Session

Board 1169T: ShinyMagnifying: A Comprehensive Shiny-Based Toolbox for Single-Cell RNA-Seq Analysis

Authors:

A. Chamorro-Parejo, S. Moeng, B. Hoden, K. Ramos; Texas A&M Inst. of BioSci.s and Technology, Houston, TX

Abstract:

Single-cell RNA sequencing (scRNA-seq) has transformed biomedical research by providing high-resolution insights into cellular heterogeneity and gene expression variability. Despite its groundbreaking impact, scRNA-seq data analysis remains a challenge due to the complexity, sparsity, and high dimensionality of data and the limited accessibility to user-friendly interfaces for data analysis. To address these challenges, we developed a computational toolbox in R using a common web-based framework, namely, Shiny. This toolbox facilitated analyses by creating an interface that ensured accessibility, adaptability, and configurability of Seurat and Scanpy data types. The new toolbox incorporated a comprehensive set of functionalities for scRNA-seq data analysis (basic analysis workflow, gene ontology (GO) analysis, cell type annotation, trajectory/pseudo-time analysis, network construction, and alignment-free expression quantification) in the evaluation of public databases as well as data from our own laboratory to examine lung epithelial cell heterogeneity and their response to carcinogen challenge. The toolbox was used to convert complex scRNA-seq datasets into interactive and shareable interfaces that were user-friendly and customizable. Further, the tool allowed seamless use of objects that store all the information associated with a dataset, facilitated addition of new features, incorporated GO analysis, and refined cell annotations. The web-based nature of the toolbox enabled deployment on a web portal based-data analysis without additional licensing costs and provided an engaging and intuitive user experience. Thus, ShinyMagnifying can provide the computational power necessary for advanced data analyses, while democratizing access to advanced genomic analysis.

Session Title: Omics Technologies Poster Thursday Session

Board 1170T: Simultaneous single cell sequencing of genetic and epigenetic bases

Authors:

W. Gosal[1], A. Ciau-Uitz[1], C. Shi[1], I. Valent[1], J. Monahan[1], J. Vivian[1], M. Taipale[1], P. Creed[1], P. Golder[1], P. Kokko-Gonzales[1], R. Tavares[2], S. Balasubramanian[2]; [1]biomodal, Cambridge, United Kingdom, [2]Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

A body of evidence indicates that epigenetic modifications in DNA comprises a fundamental pathway by which genes can be silenced or activated, determining cell fate and function. In mammalian genomes, the fifth carbon of cytosine is one major target for epigenetic modification. 5-methylcytosine (5mC) is associated with gene silencing and patterns of this modification are altered in diseases such as cancer. In contrast, 5-hydroxymethylcytosine

(5hmC) has only recently been suggested to play a role in gene regulation and is generally considered a marker of active genes and enhancers. Currently, reading these two modifications remains elusive at the single cell level, making the function and relationship between these modifications difficult to precisely ascertain. Traditionally, a large number of cells are combined and analysed as a bulk, which produces average methylome profiles. This can hide cryptic heterogeneity such as distinct subpopulations with distinct methylation patterns.
Here we present a single nuclei workflow to simultaneously determine 5mC and 5hmC at the single molecule level. Our method utilises a transposase to simultaneously tag and fragment DNA with a hairpin sequence containing a molecular barcode unique to the cell of origin. After tagmentation, we use an armoury of enzymes, including a high-fidelity maintenance methyltransferase together with a methylcytosine dioxygenase, a helicase and a deaminase to unequivocally read 5mC and 5hmC in addition to the four canonical genetic bases from the same molecule.
Applying this method, we determined the 5mC and 5hmC patterns in the ES cells line (ES-E14TG2a) cultured under two conditions (Serum+LIF+2i and Serum+LIF) at high sensitivity (>95% for both modification) and specificity (>98%). The method successfully recovers distinct populations that have distinctive (hydroxy)methylome trajectories at regulatory regions. In addition, we demonstrate consistency between pseudo-bulk and bulk measurements for these different cell types. Furthermore, we extend this method to cells isolated from complex tissues such as mouse cortex, where we identify multiple cell populations based on their respective (hydroxy)methylome patterns.
In summary, we demonstrate simultaneous, phased reading of all four genetic bases and the two most biologically relevant cytosine modifications at the single cell level. This provides a more complete picture of the information stored in cellular genomes and will have applications throughout biology and medicine.

Session Title: Omics Technologies Poster Thursday Session

Board 1171T: Single-cell based integrative analysis of transcriptomics and genetics reveals robust associations and complexities for inflammatory diseases

Authors:

H. Townsend[1], K. Rosenberger[1], L. Vanderlinden[2], J. Inamo[2], F. Zhang[2,1]; [1]Univ. of Colorado - Boulder, Boulder, CO, [2]Univ. of Colorado - Anschutz Med. Campus, Aurora, CO

Abstract:

Background: Understanding genetic underpinnings of immune-mediated inflammatory diseases is crucial for improving treatments. Single-cell RNA sequencing (scRNA-seq) identifies cell states expanded in diseased tissue, but often overlooks genetic causality due to cost and small genotyping cohorts. Conversely, large genome-wide association studies (GWAS) are commonly accessible.
Methods: We present a 3-step robust benchmarking analysis of integrating GWAS and scRNA-seq to identify genetically relevant cell states and genes in inflammatory diseases. First, we applied and compared the results of two recent algorithms, based on networks (scGWAS) or single-cell disease scores (scDRS), according to accuracy/sensitivity and interpretability. While previous studies focused on coarse cell types, we used disease-specific, fine-grained single-cell atlases (183,742 and 228,211 cells) and GWAS data (Ns of 97,173 and 45,975) for rheumatoid arthritis (RA) and ulcerative colitis (UC). Second, given the lack of scRNA-seq for many diseases with GWAS, we further tested the tools' resolution limits by differentiating between similar diseases with only one fine-grained scRNA-seq atlas. Lastly, we provide a novel evaluation of noncoding SNP incorporation methods by testing which enabled the highest sensitivity/accuracy of known cell-state calls.
Results: We first found that single-cell based tool scDRS called superior numbers of supported cell states, like MERTK+ myeloid cells in RA, which were overlooked by network-based scGWAS. While scGWAS was advantageous for gene exploration, scDRS captured cellular heterogeneity of disease-relevance without single-cell genotyping. For noncoding SNP integration, we found a key trade-off between statistical power and confidence with positional (e.g. MAGMA) and non-positional approaches (e.g. chromatin-interaction, eQTL). Even when directly incorporating noncoding SNPs through 5' scRNA-seq measures of regulatory elements, non disease-specific atlases gave misleading results by not containing disease-tissue specific transcriptomic patterns. Despite this criticality of tissue-specific scRNA-seq, we showed that scDRS enabled deconvolution of two similar diseases with a single fine-grained scRNA-seq atlas and separate GWAS. Indeed, we identified supported and novel genetic-phenotype linkages separating RA and ankylosing spondylitis, and UC and crohn's disease. Overall, while noting evolving single-cell technologies, our study provides key findings for integrating expanding fine-grained scRNA-seq, GWAS, and noncoding SNP resources to unravel the complexities of inflammatory diseases.

Session Title: Omics Technologies Poster Thursday Session

Board 1172T: Single-cell multi-omics analysis reveals mucosal immune cell profiles in ulcerative colitis

Authors:

S. Feng[1,2], E. Heidrich-O'Hare[1], Z. Xu[1,2], H. Hu[1], S. Tao[1], W. Chen[1], R. Duerr[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Tsinghua Univ., Beijing, China

Abstract:

INTRODUCTION: Ulcerative colitis (UC) and Crohn's disease (CD) are the two forms of inflammatory bowel disease (IBD). When UC involves only part of the colon, a sharp demarcation usually separates inflamed distal mucosa and uninflamed proximal mucosa. GWAS have successfully identified genetic risk loci associated with UC; however, the functional interpretation of these loci and their implications in disease pathogenesis remain incompletely understood. **RATIONALE:** Cells and molecules enriched in inflamed mucosa are candidate targets for UC therapy. We compared single-cell assays for gene expression (GEX) and antibody-derived cell surface protein tags (ADTs) (CITE-seq), or GEX, ADTs and transposase accessible chromatin (ATAC) (DOGMA-seq) in immune cells from macroscopically inflamed and non-inflamed mucosa in UC patients. **METHODS:** Biopsies of inflamed and non-inflamed mucosa in UC patients were cryopreserved, thawed in batches, dissociated, and flow-sorted to enrich live immune cells. ~45,000 cells from 13 UC patients were profiled with CITE-seq, and ~9,000 cells from 7 UC patients were profiled with DOGMA-seq. Downstream analyses included trajectory inference by RNA velocity, polygenic and heritability enrichment analysis, and transcription factor (TF)-gene regulatory network construction. **RESULTS:** By leveraging ATAC peaks and their associated TF binding motifs, which provided complementary information to GEX and ADT modalities, we distinguished credible T cell subtypes such as Th1, Th17, and atypical functional CD4+ T cell subtypes. To prioritize cell types related to polygenic risk and heritability of UC, we applied the single-cell disease relevance score (scDRS) and stratified linkage disequilibrium score regression (S-LDSC) to GEX or ATAC modality. Both analyses revealed significant enrichment of UC-associated genetic variants in CD4+ T cells, particularly in highly activated CD279+CD103- effector memory T cell, CD103+CD279+ resident memory Th17, and central memory Treg subtypes of CD4+ T cells. These findings guided our downstream analyses, where we identified GEX and ADT dynamics and the key TFs driving the trajectories towards these UC-associated subtypes. To further elucidate the potential driving regulatory mechanisms in UC mucosal CD4+ T cells, we constructed TF-gene regulatory networks involving fine-mapped UC GWAS SNP-associated genes. **CONCLUSION:** Using single-cell multi-omics data, we provide a comprehensive characterization for the heterogeneous UC mucosal immune cells, and highlight the potential functional consequences of UC genetic variants, with a particular focus on CD4+ T cells.

Session Title: Omics Technologies Poster Thursday Session

Board 1173T: Single-cell RNA sequencing analysis of malignant transformation and carcinogen response heterogeneity in lung cancer

Authors:

S. Moeng[1], A. D. Chamorro-Parejo[1], J. J. Cai[2], K. S. Ramos[1]; [1]Texas A&M Inst. of BioSci.s and Technology, Houston, TX, [2]Texas A&M Univ., Dept. of Vet. Integrative BioSci.s, College Station, TX

Abstract:

Lung cancer, the leading cause of cancer mortality worldwide, displays high degrees of inter- and intra- tumor heterogeneity that pose significant challenges to our understanding of the biology of cancer and the development of molecular biomarkers and innovative therapies. In this study, we used single-cell RNA sequencing (scRNA-seq) to examine the transcriptomic heterogeneity of transformed (NCI-H460) and non-transformed (BEAS-2B) lung epithelial cell lines, along with their response to carcinogen challenge. Discretization of four primary clusters by K-means clustering resolved 13 subclusters, including three subclusters each for control and carcinogen-challenged H460 cells, and three and four subclusters, respectively, for vehicle and carcinogen-challenged BEAS-2B cells. Marker genes for each cluster and subcluster were identified to define the uniqueness of each group. Salient genes identified in our analysis included ALDH3A1, TP53I3, LIAS, MXD3, ANTXR1, PRRX2, ADGRG6, with a vast degree of transcriptional heterogeneity seen in several of the subclusters. We next characterized functional annotations exhibiting major changes along cell trajectories and pseudo-time analyses. A clear transition was seen in H460 cells challenged with the carcinogen compared to vehicle, with varying transition states marked by changes in the expression of ALDH3A1, CDKN1A, MDM2, FDXR, and others. While transformed cells exhibited significant transitional states, non-transformed cells maintained considerably stable patterns of gene expression. Gene Ontology analysis identified DNA replication, DNA repair, regulation of DNA binding, and protein transmembrane transport as major targets of transcriptional deregulation. Together, these findings highlight the differential transcriptional states of transformed and non-transformed lung epithelial cells and their variable response to carcinogen challenge. Thus, scRNA-seq approaches can be successfully used to resolve cancer biology networks and identify putative targets and pathways for precise diagnosis and therapeutic intervention.

Session Title: Omics Technologies Poster Thursday Session

Board 1174T: Single-cell RNA sequencing of plasmacytoid dendritic cells implicates inflammatory pathways in primary graft dysfunction after heart transplantation

Authors:

L. C. Kwee[1], J. B. Lerman[2], R. T. Gross[3], P. N. Patel[3], M. Mendiola Pla[3], K. Dewan[3], K. Corey[1], M. Ramaker[1], C. L. Holley[2], A. D. DeVore[2], J. N. Schroder[3], C. A. Milano[3], K. J. Weinhold[3], L. K. Truby[4], D. E. Bowles[3], S. H. Shah[1,2]; [1]Duke Molecular Physiology Inst., Durham, NC, [2]Dept. of Med., Div. of Cardiology, Duke Univ. Med. Ctr., Durham, NC, [3]Dept. of Surgery, Duke Univ. Med. Ctr., Durham, NC, [4]Univ. of Texas Southwestern Med. Ctr., Dallas, TX

Abstract:

Primary graft dysfunction (PGD) is a leading cause of morbidity and mortality early after heart transplantation and is characterized by acute failure of the cardiac allograft. In prior work, we identified CLEC4C as a circulating pre-transplant biomarker in the recipient that predicted PGD. CLEC4C is a membrane protein in plasmacytoid dendritic cells (pDCs), which participate in response to pathogen-associated molecular patterns, suggesting a role for the innate immune response in the development of PGD. Here, we performed single-cell RNA sequencing (sc-RNAseq) to better understand the potential role of pDCs in the pathogenesis of PGD. Peripheral blood mononuclear cells were prospectively isolated from four recipients immediately prior to cardiac transplant: two who experienced PGD as defined by consensus criteria (cases), and two who did not (controls). Sc-RNAseq libraries were prepared using standard 10x Genomics protocols, followed by paired-end NovaSeq 6000 sequencing. After the removal of ambient RNA and QC filtering, samples were normalized and integrated. Cells were clustered using Seurat and annotated with singleR. Differential expression analysis compared transcript levels between PGD cases and controls in pDCs using the Wilcoxon rank sum test, and pathway analysis was performed using Gene Set Enrichment Analysis (GSEA). After processing, data from 46,648 cells (16,054 genes) were used for analysis. The pDC cluster exhibited high expression of two canonical transcripts: *CLEC4C* was expressed in 77% of cells in the pDC cluster vs. 0.3% in other clusters, with log2 fold change (log2FC) = 9.2; *LILRA4* was in 85% of putative pDCs vs. 0.4% of other cells, with log2FC = 9.9. One hundred twenty-five transcripts were differentially expressed in pDCs in PGD cases vs. controls (FDR < 5%), including *MAP2K1*, part of the MAP kinase signal transduction pathway, transcription factors including *CREB3L2* and IKZF1, and the chemokine receptor *CXCR4.* GSEA revealed 49 associated KEGG pathways, including inflammatory pathways such as chemokine signaling (FDR-adj p=0.0005) and Fc gamma R-mediated phagocytosis (FDR-adj p =0.005), along with several cancer pathways. pDCs are a relatively small subset of cells, and were more abundant in controls (0.6% vs. 0.3%, $p=10^{-6}$). These results suggest that

inflammatory pathways activated in pDCs before cardiac transplantation may impact the development of PGD after transplant. Although additional studies are needed to better define how these pathways contribute to the pathogenesis of PGD, leveraging this for risk stratification could improve post-transplant outcomes.

Session Title: Omics Technologies Poster Thursday Session

Board 1175T: SLACS (Spatially-resolved Laser Activated Cell Sorting) Empowers Primary Template-Direct Amplification

Authors:

J. Hong, S. Lee, A. Lee; Meteor Biotech, Seoul, Korea, Republic of

Abstract:

Cancer genomes often contain rare sub clonal mutations critical for understanding tumor heterogeneity and evolution. Somatic variant detection, including single nucleotide variants (SNVs), copy number variants (CNVs), and structural variants (SVs), can uncover rare subclones and unravel mysteries about tumor clonal evolution. Current WGA techniques often result in uneven amplification of different genomic regions. If coverage is not uniform, these low-frequency variants may not be detected. Inevitably, a WGS technique that shows consistent coverage is required. Biases can lead to false negatives (missed mutations) or false positives (artifacts interpreted as mutations). On top of that, there is a great deal of spatial heterogeneity within tumors, which can make it difficult to recognize these rare tumor mutations. Spatial DNA sequencing allows us to map genetic variations in relation to their physical locations within the tumor, providing a more detailed and accurate understanding of tumor heterogeneity and evolution. SLACS is a direct laser punching ROI Isolation system that enables high-resolution analysis of molecular information within tissues. This technology precisely capturing them using a pulsed infrared (IR) laser without requiring any microdissection steps. To address spatial heterogeneity and find true mutations with uniform WGS, we applied Primary Template directed Amplification (PTA), which provides uniform coverage, to a technique called Spatially-resolved Laser Activated Cell Sorting (SLACS). In this study, we employed SLACS technology to meticulously isolate targets from a complex heterogeneous population in cancer cell line & cancer tissues. These isolated cells conducted WGS by optimized PTA protocols for SLACS. The data in the HL60 cell line got good quality of data similar to the existing data at the level of 1 cell, and then we applied it in cancer tissue to find existing mutations and de novo mutations between groups. We were able to see more deeply into the cancer genome, including SNVs, CNVs, SVs, and more. By leveraging spatial DNA sequencing, we were able to map the

distribution of these genetic variations within the tumor, offering insights into the spatial dynamics of tumor evolution and heterogeneity. This study underscores the transformative impact of integrating SLACS with PTA in cancer genomic research. Spatial DNA sequencing provided a deeper understanding of tumor architecture and genetic diversity, highlighting its importance in the accurate detection of somatic variants. Future investigations will aim to optimize this protocol further and explore its broad applicability across diverse biological fields.

Session Title: Omics Technologies Poster Thursday Session

Board 1176T: ST-COS: Simulator of Spatial Transcriptomic Data Preserving Gene-Gene Co-expression Patterns

Authors:

Y. Zhang, Z. Yu, W. Liu, J. Chen, Y. Li; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

A wide variety of computational methods designed for spatial transcriptomic (ST) data have been recently developed to enable gene expression profiling with spatial information. ST data with known spatial domain or cell type annotation by pathologists is usually considered as the “gold standard”. Nevertheless, the number of benchmark datasets and their size are insufficient for validating statistical and computational efficiency of various methods. The rapid advancement of analytical methods developed for ST data demands comprehensive and impartial re-evaluation and benchmarking, requiring large amounts of annotated ST data than currently available. Synthetic datasets can serve as promising and powerful alternatives. Current ST simulators, however, require reference single cell RNA-seq or ST data, offer limited choices of spatial patterns, neglect batch effects across slides, and do not consider gene-gene co-expression patterns. Here, we present ST gene-gene CO-expression patterns preserving Simulator (ST-COS). ST-COS generates gene expression profiles based on the negative binomial model coupled with gaussian copula to flexibly simulate gene co-expression. ST-COS allows both reference-free and reference-based realistic simulations, diverse spatial variation patterns, batch effects, and most importantly, correlation among genes. Genes naturally synergize with each other through biological networks, leading to complex correlation structure that potentially affects the robustness of various key downstream analyses including spatial variable gene and spatial co-expression gene set detection, spot deconvolution, and domain segmentation; our ST-COS now provides “silver standard” realistic datasets to systematically evaluate all these methods. Deploying ST-COS, we simulated data using a wide range of

parameter settings to characterize the benefits and limitations of the state-of-arts methods for ST data across different scenarios.

Session Title: Omics Technologies Poster Thursday Session

Board 1177T: Succinic semialdehyde dehydrogenase deficiency: Age-dependent metabolomic profile of diagnostic biomarkers in a cross-sectional study

Authors:

S. Elsea[1], C. Gijavanekar[1], A. Rajagopal[1], V. R. Sutton[1], T. Wilson[1], P. Pearl[2], K. Glinton[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Boston Children's Hosp., Boston, MA

Abstract:

Succinic semialdehyde dehydrogenase deficiency (SSADHD, OMIM #271980) is a neurometabolic disorder in the GABA catabolism pathway. The disorder is due to bi-allelic pathogenic variants in *ALDH5A1* and is usually characterized by moderate-to-severe developmental delays, hypotonia, intellectual disability, ataxia, seizures, hyperkinetic behavior, aggression, psychiatric disorders, and sleep disturbances. GABA is catabolized by a 2-step enzymatic pathway involving GABA-transaminase (GABA-T) and succinic semialdehyde dehydrogenase (SSADH) in the astrocyte mitochondria. Deficiency of either enzyme leads to accumulation of GABA. Accurate genetic diagnosis needs comprehensive disease knowledge, including pathogenic genetic variants and biomarkers. Metabolite and genomic testing may be essential but are underutilized due to the invasiveness of cerebrospinal fluid analysis and poor characterization of genomic variants. We previously reported a clear metabolomic signature of plasma biomarkers for GABA-TD (2-pyrrolidinone, succinamic acid, 4-guanidinobutanoate) and SSADHD (2-pyrrolidinone and 4-guanidinobutanoate) that also distinguish individuals treated with vigabatrin, an anti-seizure medication that blocks GABA-T. We performed plasma metabolomic analysis in a larger cohort of SSADHD individuals (n=24) with ages ranging from 6 months to 41 years at sample collection. 2-pyrrolidinone levels (mean z-score +0.9, n=7) were significantly lower in the adult age group (>=19 years) compared to the pediatric age group (<=18 years) (mean z-score +3.2, n=17), $p<0.0005$. Although significantly elevated compared to control population, 4-guanidinobutanoate levels were also significantly lower in the adult cohort (mean z-score +2.3, n =17) compared to the pediatric cohort (mean z-score +3.01, n=17), $p<0.005$. Reduction in GABA and GHB levels in patients has been reported in longitudinal studies. This observation is consistent with our findings with 2-pyrrolidinone and 4-guanidinobutanoate in this cross-sectional study. The elevation of 4-guanidinobutanoate is

hypothesized to be due to increased conversion of GABA by glycine amidinotransferase due to accumulation of succinic semialdehyde and GABA. Unlike GABA, 4-guanidinobutanoate is a stable compound and is reliably and consistently detected in our assay. Additionally, analytical ratios of these metabolites may also be investigated, particularly in adults, to investigate their potential as biomarkers of SSADHD. These findings are critical to consider regarding biochemical analysis and *ALDH5A1* variant interpretation for SSADHD with age.

Session Title: Omics Technologies Poster Thursday Session

Board 1178T: Transgenic Humanized Mouse Models for Disease Research: Leveraging Bxb1 Integrase and New Safe Harbor Loci

Authors:

A. Zuberi, J. Rollins, X. Li; The Jackson Lab., Bar Harbor, ME

Abstract:

Humanized mouse models containing low to single copies of human transgenes are pivotal in biomedical research, offering a translational bridge from preclinical studies to clinical trials. Traditionally, CRISPR/Cas9 has been used to insert human DNA in single copy into the corresponding mouse gene or into safe harbor loci, like ROSA26 on Chr 6. However, size limitations have often resulted in intronless coding regions of specific cDNAs being used as surrogates for the intact human gene, and these fail to accurately represent human gene function, particularly if the gene is expressed as alternatively-spliced transcripts. Also, cDNA transgenesis often utilizes non-native promoter and regulatory elements that alter temporal and tissue-specific expression profiles. Serine integrases, such as Bxb1 integrase, can also facilitate genome editing via RMCE (Recombination Mediated Cassette Exchange) between donor DNA flanked by attB sequences and previously inserted attP sequences in the mouse genome (e.g., ROSA26). As the central dinucleotides of both attB and attP can be altered, the use of paired attP with paired attB differing in this dinucleotide can confer directionality to the single copy integration event. The major advantage of Bxb1 integrase in genome editing is that large DNA sequences can be incorporated. Indeed, our largest transgenic mouse line generated to date contains 78-kb of contiguous human genomic DNA. However, despite its potential, the *in vivo* Bxb1 RMCE efficiency in zygotes remains modest (~10%) for transgenes ~10-kb in size or less. We are exploring alternative means of Bxb1 integrase delivery and expanding our editing strategies to include genetically modified ES cells to improve RMCE efficiency. Currently, codon-optimized Bxb1 integrase mRNA is microinjected into single cell zygotes of a dual attP-containing strain along with attB-flanked donor DNA. We are testing the effects of replacing

Bxb1 integrase mRNA with recombinant Bxb1 integrase protein and the potential benefits of using a Bxb1 integrase expressing transgenic mouse. In parallel, due to the increasing utilization of ROSA26 for new transgene insertion, we are seeking new safe harbor loci within the mouse genome. We report on one potential intergenic safe harbor on Chr 12, where an attP-flanked CAG-GFP transgene has been inserted. Because of the insertion of dual Bxb1attP landing pads into new safe harbor loci in the mouse as part of our validation procedure, the utility of using Bxb1 integrase for large DNA single copy transgenesis into other loci is enhanced.

Session Title: Omics Technologies Poster Thursday Session

Board 1180T: Unbiased multiplex antigen screening of Cerebrospinal Fluid detects microbial and autoantigenic epitopes associated with Multiple Sclerosis ⭐

Authors:

Y. Chan, N. J. Barton, K. Tran, M. N. Olson, M. Deshpande, I. Radu, N. Francis, M. Kurban, A. R. Orszulak, S. M. Chigas, J. Sundstrom, P. Dawes, L. F. Murray, C. Ionete, C. C. Hemond, E. Lim; UMass Chan Med. Sch., Worcester, MA

Abstract:

There is evidence that Multiple Sclerosis (MS) is an autoimmune neurodegenerative disease partially due to cross reactive autoantibodies against central nervous system (CNS) epitopes caused by common viral infections. 100% of all MS patients are Epstein Barr virus (EBV) seropositive and infections by other viruses such as human herpesvirus have been shown to be associated with MS. To further understand MS pathology due to cross reactive autoantibodies, we investigated the cerebrospinal fluid (CSF) samples of 92 MS and 103 non-MS patients for intrathecal antibodies with a multiplex unbiased bacteria peptide library. These patient CSF samples were incubated with a bacteria library expressing random surface antigen peptides. CSF antibodies will bind to the antigen and are then selected, lysed and deeply sequenced. Using innovative bioinformatics, we mapped the data onto relevant protein sequences to discover antibody epitopes targeting a region of the protein and the strength of the epitope signal is expressed as an area-under-curve (AUC) score. We then use an outlier-sum (OS) statistical test to determine if the AUCs are significantly enriched within the MS samples compared to the non-MS samples. Using this procedure, we tested EBNA1, an EBV antigen for epitopes and found a significant enrichment at position 407-419. We subsequently multiplex tested the data against a library of 129 human viruses and discovered other epitopes from viruses such as Hantaan virus, human herpesvirus 6A (HHV6A), hepatitis B virus (HBV), hepatitis C virus (HCV) and human

herpesvirus 8 (HHV8) to be associated with MS. Then, we screened for potential autoantigens of the CNS by mapping the data to 2438 CNS surface proteins. From this, we discovered several MS-associated autoantigenic epitopes in proteins such as ADRB3, HTR3A and MPO. Because of previous associations of MS with toxoplasma gondii infection, we also mapped the data to the toxoplasma gondii proteome and discovered additional epitopes enriched for MS. We further assessed epitope-epitope correlations within the patient samples and found distinct patterns of association between these microbial and auto antigenic epitopes like toxoplasma gondii-ADRB3, HHV6A-HBV and HCV-HHV8. Finally, we performed machine-learning to determine if these epitopes are predictive for MS and found that the model incorporating all the epitopes could most effectively discriminate between MS and non-MS (ROC-AUC score = 0.91). Our results demonstrate the effectiveness of multiplex unbiased screens to detect the identity of potentially cross-reactive antibodies targeting MS CNS epitopes and they can also be used as effective biomarkers for MS.

Session Title: Omics Technologies Poster Thursday Session

Board 1181T: Unraveling alveolar macrophage plasticity and metabolism in hypoxia-induced pulmonary hypertension.

Authors:

N. Campbell[1], I. Campbell[2], H. Zhang[1], M. Li[1], C. Delaney[1], R. Kumar[3], D. Irwin[1], B. Graham[3]; [1]Univ. of Colorado, Aurora, CO, [2]Stanford Univ. (Alumni), Stanford, CA, [3]Univ. of California, San Francisco, San Francisco, CA

Abstract:

Few studies have probed lung alveolar macrophage (AM) molecular phenotypes after exposure to hypoxia despite macrophages' known contribution to pulmonary hypertension (PH). In this study, we aimed to identify AM subtypes and their associations with hypoxia-induced PH. We hypothesized that hypoxia induces metabolic dysregulation, leading to significant phenotypic alterations in AM subtypes. We exposed $Cx3cr1^{GFP/+}$ (C-X3-C motif chemokine receptor 1, green fluorescent protein) reporter mice to normoxia (~21% O_2), or to one day (D1), seven days (D7), or 21 days (D21) of hypoxia (~10% O_2). Flow cytometry was used to isolate $Cx3cr1^-$ AMs, followed by single-cell RNA sequencing using the 10x Genomics Chromium platform. Data analyses were performed using Cell Ranger, Seurat, ClusterMap, and Ingenuity Pathway Analysis (IPA). Our final analyses included n=1272 (normoxia), n=1081 (D1), n=898 (D7), and n=19096 (D21) AMs. Statistical significance for functions, pathways, and upstream regulators was determined using Fisher's exact test (Benjamini-Hochberg adjusted p-value < 0.05).We identified

two putative-untrained AM subtypes that constituted 97% of the total population in normoxia, 92% of which disappeared by D1. The remaining population transitorily expanded at D7 but significantly decreased at D21. These subtypes were enriched with Granzyme A Signaling and the inhibition of the Pentose Phosphate, Ketogenesis, Gluconeogenesis I, Super pathway of Geranylgeranyl Diphosphate Biosynthesis I, and Glycolysis I. These pathways were enriched at D21 but not in normoxia, D1, or D7. Additionally, these subtypes were overrepresented by Protein Metabolism and VDR/RXR Activation in normoxia but not hypoxia.Two putative-trained subtypes emerged at D1, constituting the same population size as those that disappeared by D1. By D7, these populations transiently decreased, but by D21, they recovered, becoming the most abundant. They were enriched with Cholesterol Biosynthesis I, II, and III, and the Mevalonate Pathway I at D7 but not D1 or D21. The Super pathway of Geranylgeranyl Diphosphate Biosynthesis I pathway was enriched at D1, D7, and D21. Mitochondria Dysfunction was inhibited at D1 but activated at D21, implicating these pathways in PH development.In summary, this study redefines the Macrophage Disappearance Reaction (MDR) as the point where transcriptomes reprogram and macrophages transition from untrained to trained states, integrating the MDR, niche model, and macrophage plasticity phenomena. This unified framework underscores how these changes increase heterogeneity and make AMs more plastic and adapted to hypoxia.

Session Title: Omics Technologies Poster Thursday Session

Board 1182T: Unraveling neurodevelopmental phenotypes in individuals with complex copy number and structural variants via integrated genomic, transcriptomic and epigenomic analyses

Authors:

Y. Zhuang[1], R. Wang[1], D. Ellershaw[1], E. Chekalin[1], J. Albay[1], D. Hernandez[1], Q. Zhang[2], C. Glidewell-Kenney[1], I. Khrebtukova[1], K. Wentworth[3], Z. Talebizadeh[3], Y. Dewoody[4], B. Shah[5], V. Vogel-Farley[3], A. Crawford[1]; [1]Illumina Inc., San Diego, CA, [2]Illumina, Inc, San Diego, CA, [3]Global Genes, Aliso Viejo, CA, [4]Ring14 USA, Midland, TX, [5]Project8p Fndn., New York, NY

Abstract:

Background: Large pathogenic copy number variants (CNVs) and structural variants (SVs) such as deletions/duplications on chromosomes 8 or 15, and ring chromosome 14 are associated with neuro-developmental disorders (NDDs) and birth defects. Despite clinical and cytogenomic efforts, gaps remain in fully understanding these complex disorders. Our study adopts a multiomic approach to investigate the molecular basis of these disorders and provides new

insights into biological mechanisms.*Methods:* DNA and RNA were isolated from blood samples provided by families of 8p inversion duplication/deletion syndrome (8p, N=12), duplication 15q syndrome (dup 15q, N=14), and ring 14 syndrome (Ring14, N=3). We generated a multiomic dataset from these samples using short read whole genome sequencing (srWGS), Illumina Complete Long Read (ICLR) sequencing, whole transcriptome sequencing (WTS), and methylation profiling on EPIC v2 arrays. A multiomic analysis framework was established including AI-enabled WGS interpretation, differential gene and allele specific expression, differential methylation, and parent of origin analyses.*Results:* srWGS identified CNVs and SVs consistent with previous karyotypic or microarray findings. ICLR provided additional breakpoint resolution when using the T2T reference genome. Gene expression changes were accordant with regions of copy number change. Duplicated regions in both 8p and 15q samples were maternally derived in agreement with other studies. Allele specific expression of the paternal allele in dup 15q syndrome implies a mechanistic difference between dup 15q syndrome and Prader-Willi syndrome.*Conclusion:* This study provides a multiomic view into a class of neurodevelopmental disorders using integrative WGS, WTS, and methylation data analysis. It demonstrates the potential of using multiple omics data to facilitate research and clinical findings.

Session Title: Omics Technologies Poster Thursday Session

Board 1183T: Unsupervised evaluation of pre-trained DNA large language models embeddings

Authors:

R. Awasthi, X. Zhu; Case Western Reserve Univ., Cleveland, OH

Abstract:

DNA large language models(DNA-LLMs) have generated a lot of hope and hype for solving complex genetics tasks. These models have demonstrated remarkable performance in tasks like gene finding, enhancer annotation, and histone modification. However, they have struggled with tasks such as learning individual-level personal transcriptome variation, highlighting the need for robust evaluation approaches. Current evaluation approaches assess models based on multiple downstream tasks, which are computationally demanding and fail to evaluate their ability to learn as generalist agents.This study proposes unsupervised metrics (RankMe, NeSuM, and StableRank) to evaluate the embedding quality of DNA-LLMs based on linear algebra principles. RankMe estimates the effective rank of a matrix. NeSuM analyzes the eigenspectrum of the covariance matrix of representations. StableRank determines how difficult it is to

estimate the matrix by subsampling its rows. These approaches were previously used for evaluating natural language models but not DNA-LLMs. We also propose a network-based approach to qualitatively evaluate the embeddings with biological insights.We selected four state-of-the-art pre-trained DNA-LLMs: Nucleotide Transformer, DNA-BERT2, HyenaDNA, and MistralDNA, and 50k coding gene sequences and 50k non-coding sequences from the Genomic Benchmark dataset. First, we generated embeddings of all 100k sequences using all four models and benchmarked them with the metrics RankMe, NeSuM, and StableRank scores. Next, we generated gene-gene networks of 50k coding gene sequences and learned communities, and for the largest community, we performed gene set enrichment analysis.From the embedding quality results, we found that the Nucleotide Transformer had the best results for RankMe and NeSuM metrics, achieving RankMe scores of 6.19 (coding) and 6.29 (non-coding), and NeSuM scores of 6.79 (coding) and 6.55 (non-coding). DNA-BERT2 followed closely behind, while HyenaDNA and MistralDNA lagged. For StableRank, DNA-BERT2 had the top score with 2.09 for coding sequences and 2.39 for non-coding sequences. Gene set enrichment analysis on coding sequences within the largest community uncovered interesting insights. The Nucleotide Transformer embeddings exhibited the most significant enrichment for the Cytoplasm (P-value = 0.000069), supported by DNA-BERT2 (P-value = 0.0139), highlighting their capabilities in capturing biologically relevant patterns. We hope the proposed quantitative and qualitative approaches drive a new paradigm of DNA-LLMs evaluation, optimizing computational efficiency, a significant challenge in utilizing DNA-LLMs.

Session Title: Omics Technologies Poster Thursday Session

Board 1184T: Using donor-specific assemblies as a benchmark for short- and long-read sequencing platform comparisons within the SMaHT network

Authors:

S. Mcgee[1], M. Vollger[1], Y. Kwon[1], K. Munson[2], K. Hoekzema[2], C. Frazar[1], N. Parmalee[3], W. Harvey[1], J. Bennett[4], A. Stergachis[1], E. Eichler[2], C-L. Wei[1]; [1]Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA, [3]Seattle Children's Res. Inst., Seattle, WA, [4]Seattle Childrens Res. Inst., Seattle, WA

Abstract:

With the increasing focus on somatic mosaicism and the need for an accurate determination of low allele fraction variations, we have developed a technique to directly benchmark and compare the accuracy of short-read sequencing (SRS) and long-read sequencing (LRS)

technologies by pairing sequence data with a high-quality donor-specific assembly (DSA) generated from the same cell line sample. Specifically, as part of the Somatic Mosaicism across Human Tissues (SMaHT) network, we generated a high-quality diploid genome assembly of the COLO829BL lymphoblastoid cell line using paired PacBio HiFi Fiber-seq, ultra-long Oxford Nanopore (ONT), and Hi-C data. This assembly has a hap1/hap2 QV of 56.3/60.0, with 280 Mbp of diploid genome not syntenic with GRCh38. In addition, we generated sequence data on the same cell line sample using Element AVITI, Ultima Genomics, Illumina NovaSeq X Plus, Oxford Nanopore, and PacBio HiFi. Sequencing reads were mapped to the single diplotype reference fasta for COLO829BL and mismatches within each mapped read were identified, under the assumption that any mismatch should represent a sequencing artifact. We characterized read mismatches using Best (v0.1.0), which tallies mismatch rates for homopolymer and non-homopolymer insertions and deletions individually, as well as a total mismatch rate. In addition, it enables the benchmarking of sequencer reported base quality score values against a ground truth genome. We demonstrate marked heterogeneity in the accuracy of these sequencing technologies and show that using a DSA for alignment reduces the fraction of unaligned bases across technologies by ~20%. Specifically, we find that Element AVITI has by far the lowest rate of indel mismatches, with an error rate of 0.004 homopolymer deletions per kbp of sequence data and Ultima ppmSeq has the lowest mismatch rate among all short-read sequencing technologies. In addition, we find that PacBio HiFi, as circular consensus sequences, has the lowest rate of sequence mismatches, with a rate of 0.148 mismatches per kbp of sequence data, which is ~10-fold lower than all of the tested single-pass sequencing approaches. Finally, we show that base score recalibration using GRCh38 actually worsens the base quality scores for some sequencing technologies. The results provide important insights into the optimal design to enhance somatic variant discovery in human genomes and benchmark for accuracy for emerging sequencing technologies.

Session Title: Omics Technologies Poster Thursday Session

Board 1185T: VariLink - a comprehensive assay for cancer genome characterization and its comparison to WGS

Authors:

C. Padilla[1], J. Torchia[1], D. Hwang[1], E. Danwanichakul[1], L. Munding[2], Z. Sanborn[1]; [1]Dovetail Genomics, Scotts Valley, CA, [2]Dovetail Genomics, SCOTTS VALLEY, CA

Abstract:

Cancer is a disease caused by genetic and/or epigenetic disruption. Currently, no single technology can capture the wide range of variation known to drive tumorigenesis with high sensitivity. Deciding how comprehensively to characterize a tumor genome is influenced by assays' costs, lab time, as well as the utility and value of the data generated. To address this, we developed VariLink - a multi-purpose proximity ligation assay that yields high-quality, unbiased Hi-C libraries that capture not only chromosomal topological features linked to the tumor's epigenetic state, but capable of highly sensitive discovery of structural variants (SVs). Importantly, VariLink data displays uniform coverage and thus enables similar performance to whole genome shotgun (WGS) for single nucleotide variants (SNVs), insertions / deletions (InDels), and copy number variants (CNVs) discovery. The library is generated from one single-shift assay fully compatible with Illumina sequencers.
To assess the real-world performance of VariLink relative to standard NGS, we selected a clinical ovarian serous adenocarcinoma to generate WGS and VariLink libraries and sequence to ~30x coverage. To find gene fusions induced by SVs, RNA-seq libraries were sequenced to yield over 100M reads. Sequencing data were aligned by BWA (VariLink and WGS) and STAR (RNA-Seq). SVs were called by Hi-C breakfinder (VariLink) and Manta (WGS), and Arriba was used to call gene fusions from RNA-Seq. For both VariLink and WGS, DeepVariant was used to detect SNVs & InDels and CNVs were called by CNVkit. VariLink data discovered three translocations (tlocs) between chromosomes 1, 3, 4, and 12 with strong supporting evidence (569 - 3,232 read pairs). These tlocs are supported by changes in copy number and are located at CNV boundaries. Furthermore, all 3 tlocs are validated by PCR and Sanger. By comparison, RNA-Seq/Arriba called only one fusion consistent with tloc(1;12), while WGS/Manta detected 2 of these 3 tlocs with low precision (ranked 538 & 946 of 1,018 putative tlocs sorted by decreasing read pair support). An overlap analysis of known variants demonstrate that both libraries produce concordant callsets, where 95.7% (89.7%) of WGS calls and 95.8% (87.8%) of VariLink calls overlap the set of all known SNVs (InDels) detected in both libraries. Similarly, copy number between VariLink & WGS share a high degree of similarity (> 98%). We find VariLink outperforms WGS and RNA-seq in SV discovery with both higher sensitivity and significantly higher precision. This result, coupled with VariLink's similar performance to WGS across other variant types, indicate VariLink is a powerful tool for genetic variation discovery.

Session Title: Omics Technologies Poster Thursday Session

Board 1186T: When Proteins Mislead: The Epitope Effect in pQTL Analysis

Authors:

J. Kuliesius[1], J. A. Marsh[1], P. Navarro[1], L. Klaric[2], J. F. Wilson[1]; [1]Univ. of Edinburgh, Edinburgh, United Kingdom, [2]BioAgeLabs, Richmond, CA

Abstract:

Quantitative trait loci (QTL) studies, enriched by proteomics data, have revolutionized our understanding of genetic regulation of protein expression. This research assesses the potential influence of missense variant-driven epitope effects on protein QTL (pQTL) detection in genetic studies. Specifically, altered protein 3D structure could affect the binding of detection molecules such as antibodies or aptamers, which would then be misrepresented as change in protein concentration. We hypothesize that for the epitope effect to be widespread, pQTL reported using affinity-based proteomics platforms would be enriched in missense variants affecting surface amino acids.

We compared two high-throughput proteomics platforms - Olink and SomaLogic, which analyse up to 2923 and 4719 human proteins, respectively. Our methodology involves a robust alignment of sentinel pQTL reported in the largest proteomics studies to date, creating a rich database of 11805 pQTL identified using the antibody-based Olink technology and 21027 pQTL with the aptamer-based SomaLogic platform. By leveraging machine learning structural predictions from AlphaFold and empirical data from the Protein Data Bank (PDB), we systematically evaluate the pQTL that may be influenced by epitope effects as reported by the two proteomics platforms.

Our analysis identified 666 and 1539 missense pQTL in close linkage disequilibrium (LD > 0.6) with reported *cis* variants in Olink and SomaLogic studies, respectively. Preliminary findings suggest that while the SomaLogic platform detects more cis missense variants, both platforms are generally robust against or similarly affected by epitope effects. This study not only underscores the complexity of interpreting proteomic data but also sets the stage for future advancements in proteogenomic research and technology development.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7001T: Anthracyclines induce global changes in chromatin accessibility in cardiomyocytes that overlap with cardiovascular disease loci

Authors:

E. Matthews, M. C. Ward; Univ. of Texas Med. Branch, Galveston, TX

Abstract:

Breast cancer drugs can induce adverse effects on the heart in some women. Treatment with anthracyclines or Trastuzumab increase the risk for cardiovascular diseases (CVD) such as arrhythmia (ARR) and heart failure (HF) that ultimately affect the heart muscle. These CVD are associated with hundreds of genetic variants in non-coding regions of the genome. However, how these drugs affect the non-coding genome of the heart and CVD risk loci is unknown. We therefore measured chromatin accessibility across iPSC-derived cardiomyocytes from four individuals treated with the anthracyclines, Doxorubicin, Epirubicin, and Daunorubicin, a related non-anthracycline, Mitoxantrone, and the monoclonal antibody Trastuzumab, or a vehicle control for three and 24 hours. We identified 153,348 high-confidence regions of open chromatin across 48 samples where the major sources of variation are associated with treatment and time. Jointly modeling the data across treatments and time revealed four accessibility signatures which we define as early acute response regions (EAR: 7,321), early-sustained response regions (ESR: 16,667), late response regions (LR: 41,767) and no response regions (NR: 83,741). While accessible chromatin is enriched around transcription start sites (TSS), there is a significant depletion of TSS-associated regions in the ESR and LR categories suggesting a shift to distal regulation over treatment time. To determine the potential impact of these regions, we measured the expression response of the closest gene to each region. 70% of drug response genes are associated with at least one drug-responsive chromatin region. Drug-responsive chromatin regions are enriched for motifs for 21 transcription factors, including PRDM1, ZSCAN3 and FOSL1 that are also drug-responsive genes. Finally, we overlapped our responsive chromatin regions with SNPs associated with ARR (661) and HF (398). Seventeen HF SNPs and thirty-three ARR SNPs are located within drug-responsive regions. A shared HF and ARR SNP, rs3176326, located in an intron of the *CDKN1A* gene, shows increased accessibility in response to all anthracyclines and a concomitant increase in expression of the *CDKN1A* gene. Our results demonstrate large-scale changes in chromatin accessibility in cardiomyocytes treated with anthracyclines, which correspond to several regions harboring CVD risk loci.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7002T: Building Gene-drug Clinical Validity Framework for integrating Pharmacogenomics into Genomic Medicine

Authors:

L. Gong[1], M. Whirl-Carrillo[1], S. Scott[1], K. Caudle[2], A. Moyer[3], C. J. Klein[1], T. Klein[1]; [1]Stanford Univ., Stanford, CA, [2]St. Jude Children s Res. Hosp., Memphis, TN, [3]Mayo Clinic, Rochester, MN

Abstract:

Pharmacogenomics (PGx) studies the associations between germline genetic variation and drug response variability to guide prescribing decisions and optimize drug efficacy. Despite the availability of multiple PGx resources to facilitate implementing PGx into patient care, adoption of PGx currently remains limited, in part due to unique nuances specific to PGx testing and the disconnected standards/terminologies between PGx and the broader medical genetics community. To address this gap, the Clinical Genome Resource (ClinGen) PGx Working Group (PGxWG) was formed to create frameworks for evaluating gene-drug response clinical validity and actionability aligned with ClinGen's gene-disease evaluation methods and to develop a classification system for interpreting PGx variants.

The PGxWG has developed a standardized framework that defines and evaluates the clinical validity of gene-drug pairs to assign a level of association between a gene and drug response phenotype. This semi-quantitative framework considers both clinical and experimental evidence and assigns a consensus score across drug response outcomes (efficacy, toxicity, and dosage), pharmacokinetic measures, and functional evidence. It uses qualitative classifications (Definitive, Strong, Moderate, Limited, Disputed, Refuted, and No Known PGx Relationship), which intentionally mirror ClinGen's gene-disease validity terms but with PGx-specific definitions. The framework leverages efforts from leading PGx knowledge resources such as PharmGKB, CPIC, DPWG, and the FDA to systematically evaluate and consolidate all supporting and refuting evidence. We have validated and refined the framework using representative gene-drug pairs through an iterative process. This evidence-based, systematic method to assess the strength of gene-drug response relationships will lay the groundwork for assessing the clinical actionability of gene-drug pairs and enable more consistent interpretation of PGx genes and variants.

For over 20 years, the medical genetics and PGx communities have worked independently, leading to unaligned standards and terminologies that ultimately has hindered clinical PGx integration. The PGxWG is dedicated to creating standardized frameworks to define the clinical validity and actionability of PGx genes and variants and providing an accessible nomenclature system for use across genomic medicine resources such as ClinGen, ClinVar, and ClinPGx. This initiative is critical for integrating PGx into ClinGen and more efficiently facilitating the inclusion of PGx into the larger context of genomic medicine.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7003T: Characterizing UGT1A1 promoter TA repeat polymorphisms and searching for additional polymorphisms that regulate hepatic UGT1A1 expression

Authors:

M. Taylor[1], A. Montalvo Lopez de Victoria[2], J. Collins[3], D. Wang[4]; [1]Univ. of Florida, Gaiesville, FL, [2]Univ. of Florida Coll. of Pharmacy, Gainesville, FL, [3]Univeristy of Florida, Gainesville, FL, [4]Univ. of Florida, Gainesville, FL

Abstract:

UGT1A1, a member of the phase II drug-metabolizing enzymes in the UDP-glucuronosyltransferase family, metabolizes pharmaceutics such as irinotecan and bilirubin. Its genetic polymorphisms, the TA repeats in the TATA promoter region, are associated with UGT1A1-related phenotypes and used for guiding irinotecan and atazanavir therapy. However, the results are inconsistent, especially in individuals of African ancestry. Our prior results showed an association between *UGT1A1* TA repeat polymorphisms and *UGT1A1* expression in liver samples from European American donors, but the association was less pronounced in samples from African American donors, questioning the direct impact of TA repeat polymorphisms on *UGT1A1* expression. This study evaluates the transcriptional activity of *UGT1A1* TA repeat polymorphisms (UGT1A1*1 [6TA], *28 [7TA], *37 [8TA], *36 [5TA]) using reporter gene assays. We PCR amplified ~300bp promoter DNA with varying TA repeats and cloned it into the pGL3 vector. A notable single nucleotide polymorphism (SNP), rs34547608, ~50 bp from the TA repeat, in linkage disequilibrium (LD) with *36, was found. To isolate its effect from 5TA, we altered C to T at rs34547608 via site-directed mutagenesis. Another SNP, rs887829 (~300bp from TA repeat), is in high LD with *28 and *37 and often serves as a surrogate marker for *28 and *37 in association studies. We tested its individual effects by amplifying and cloning ~700bp promoter fragments with or without rs887829 into pGL3. All constructs were sequenced to verify correct genotypes and the absence of random mutations. HepG2 cells were transfected with the constructs containing various TA repeats and/or SNPs (rs34547608 for *36 and rs887829 for *28 and *37) to assess their transcriptional activity using luciferase assays. Our preliminary results with long construct showed *36 (containing rs34547608) has higher activity than *1 (232.59% ± 100.66 vs 100% ± 8.01, $p<0.0001$) while *28 and *37 (both containing rs887829) has lower activity than *1 (59.83 ± 8.04 or 76.56 ± 14.86 vs 100% ± 8.01, $P<0.0001$). More reporter gene assays are ongoing to separate the effects from TA repeats (5TA, 7TA or 8TA) or SNPs (rs34547608 or rs887829) and to test their interactions. Our results indicate different TA repeats affect the *UGT1A1* promoter activity, but TA repeat polymorphisms cannot fully explain *UGT1A1* expression in those of African genetic ancestry.

Genome-wide approaches will be employed to identify other potential cis- or trans-regulatory SNPs that may explain the variability of *UGT1A1* expression in liver, particularly in individuals with African ancestry.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7004T: Chrono therapeutic effect of dietary intervention on modern lifestyle and association with circadian gene expression

Authors:

N. Fatima, G. Sonkar; King George's Med. Univ., Lucknow, India

Abstract:

Background: The circadian timing system is our body's built-in clock and controls our physiology on a daily basis. Disruptions to this system are linked to many health problems. **Aims:** We aimed to investigate effect of time restricted feeding on circadian rhythm regulating gene in diet induces obesity. **Method:** The eighteen Wistar rats in our study were split into three groups i.e control, TRF and high-fat diet (HFD) group. After being exposed to High fat diet (HFD) for two months to induced obesity, the HFD group kept three-month TRF treatment. The TRF group rats were allowed to eat chow diet at TRF for three months. At the end of the study, the remaining rats were sacrificed, and blood samples were collected. In order to analyze blood glucose, insulin, lipid profile, and gene expression. **Results:** Results revealed that the monthly percentage weight gain in the control group and the HFD-fed group was 25.8% and 48.5%, respectively. Rats on a high-fat diet displayed decreased insulin levels, a notable reduction in melatonin levels (p=0.006), and a decline in high-density lipoprotein (HDL) levels, coupled with elevated total cholesterol, triglycerides (TG), and low-density lipoprotein (LDL) levels. The TRF intervention exhibited a favorable impact, reducing body weight, blood glucose levels, TG, LDL, and Per1, Cry1, and Bmal1 genes showing upregulation in the HFD group. However, rats subjected to a three-month TRF intervention exhibited downregulation of these genes. **Conclusions:** These results demonstrate that TRF is a feasible and promising behavioral intervention with potential for both the prevention and treatment of obesity and related metabolic diseases. It can be easily included into lifestyle modifications.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7005T: Cross platform comparison of pharmacogenomics reporting from whole-genome sequencing

Authors:

S. Garikipati, A. El-Khani, C. Minas, D. Sanchez, H. Wu, F. Aldhuhoori, J. Quilez, J. Mafofo, J. Arres, M. Shaikh, P. Kandavel, S. Elavalli, S. Chandrashekar, T. Magalhaes; M42, Abu Dhabi, United Arab Emirates

Abstract:

Pharmacogenomics (PGx) studies the effect of genetic variation on drug response. PGx is increasingly used in biomedical research and clinical practice - primarily using commercial or custom microarrays targeting specific genes linked to drug response and a few attempts conducting whole-exome sequencing (WES). While whole-genome sequencing (WGS) has the potential to overcome some of the limitations of microarrays and WES, it has been less explored. Moreover, most WGS-based attempts have relied on Illumina short-read sequencing (SRS) and only few have used long-read sequencing (LRS), which may result in better resolved star alleles.

Here we assessed the accuracy and consistency across sequencing platforms of PGx profiling from WGS. For each of 7 Coriell reference samples, we performed ~30X WGS with Illumina and MGI (both SRS) as well as with long-read Oxford Nanopore Technologies (ONT) - the latter using both the R9 and R10 chemistries. We selected those samples as array-derived truth datasets existed for the 7 genes profiled in our PGx reporting analysis pipeline. We rigorously evaluated the precision of our PGx pipeline, focusing specifically on its ability to accurately identify star alleles across each sample and gene, as well as the accuracy of PGx interpretation (phenotype). Results were compared with a golden truth dataset, achieving a precision greater than 83.46% in sample identification and over 85.00% in gene profiling, with the lowest values observed in ONT R10. Moreover, the interpretation of these alleles matched the expected pharmacogenomic significance from the golden truth with a precision above 94.10% for samples (lowest in Illumina) and 94.81% for genes (lowest in MGI). Notably, certain genes like DPYD presented significant challenges, with precision as low as 28.57% across all platforms, highlighting the known difficulties in accurately resolving star alleles in some pharmacogenetic genes. Overall, our study underscores the integral role of advanced sequencing technologies in enhancing the accuracy and utility of pharmacogenomics. Besides, the increasing number of population genome programs using WGS enables opportunistic PGx reporting for their participants. Confidence in PGx reporting across sequencing platforms will allow aggregating PGx results across datasets.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7006T: CYP2D6*71 is a poor metaboliser allele common in Polynesian and Māori people and absent from Europeans

Authors:

M. Kennedy[1], L. Hitchman[2], T. A. Kerekere[1], A. L. Miller[2], E. Goodin[3], C. Koia[4], H. Watson[4], S. Robertson[5], P. Wilcox[6], N. Magon[1], T. Kettle[1]; [1]Univ. of Otago, Christchurch, New Zealand, [2]Univ. of Otago, Christchurch, New Zealand, [3]Univ. of Otago, Dunedin, New Zealand, [4]Ngati Porou Oranga, Gisborne, New Zealand, [5]Otago Univ., Dunedin, New Zealand, [6]Univ. of Otago, Dunedin, New Zealand

Abstract:

CYP2D6 is one of the most important pharmacogenes responsible for metabolising a wide range of medications. The gene is extremely polymorphic, which can lead to variable activity of the CYP2D6 enzyme, and elevated risks of adverse drug reactions and treatment failure. However, the full extent of variability in *CYP2D6* is not yet known, particularly for understudied populations.

We employed nanopore sequencing of 6.6kb amplicons encompassing *CYP2D6* (Liau et al., 2019) to detect all genetic variants within a group of Māori volunteers, largely affiliated with the Ngāti Porou iwi, in Aotearoa New Zealand. We confirmed the prevalence of an allele called CYP2D6*71 that was initially discovered in Han Chinese people and not observed in Europeans, which constitutes close to 10% of alleles in this Māori cohort. This confirmed and extended our earlier findings in a cohort of Pacific people (Hitchman et al., 2022).

Understanding the functional impact of the CYP2D6*71 allele is crucial to allow accurate inference of metabolizer phenotypes, and to determine whether it should be included in routine *CYP2D6 tests*.

The key CYP2D6*71 variant (rs118203758) causes a G42E substitution in the N-terminal membrane insertion region of CYP2D6, and it has been previously suggested that this may inactivate the enzyme. In order to establish the functional relevance of CYP2D6*71, we have employed two distinct approaches. First, we are using a recombinant expression system in mammalian cells, with preparation of microsomes and incubation with the CYP2D6 substrates metoprolol and solanidine, followed by mass spectrometric analysis. This work is in progress. Second, we identified several individuals with CYP2D6*71 in combination with known nonfunctional *CYP2D6* alleles, then used mass spectrometry to analyse the metabolic products of solanidine, a recently described biomarker for CYP2D6 activity, in stored plasma samples. In these cases, no evidence for metabolism of solanidine could be seen, in contrast to plasma from people with functional *CYP2D6*.

This preliminary, but fairly compelling, observation indicates that CYP2D6*71 is most probably a nonfunctional, poor metabolizer allele. Given the prevalence of this allele in Polynesian and Māori people, *CYP2D6* testing in these populations must include the key CYP2D6*71 variant (rs118203758) to ensure phenotypes are correctly inferred. Furthermore, CYP2D6*71 occurs at an appreciable frequency in Han Chinese people, so these findings may also be relevant to this population.

References

Hitchman et al. (2022) Front Genet 13:1016416.
Liau et al. (2019) Pharmacogenomics 20:1033-1047.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7007T: Diagnosis and Treatment of rare cases of PH1 in elderly patient with chronic kidney disease

Authors:

A. Chauhan[1], R. Chauhan[2]; [1]Univ. of Georgia, Columbus, GA, [2]Renal Associates LLC/Med. Coll. of Georgia, Columbus, GA

Abstract:

Primary hyperoxaluria type 1 (PH1) is a rare disorder that mainly affects the kidneys. The majority of affected individuals develop symptoms of the condition between birth and the age of 25. It results from buildup oxalates, which are normally filtered through the kidneys and excreted in the urine. In people with PH1, the oxalate is deposited in the kidneys and urinary tract and combines with calcium to form calcium oxalate. People with PH1 may experience recurrent kidney stones, blood in the urine, and urinary tract infections. PH1 can result in end-stage kidney disease, which is life-threatening, and the need for dialysis. As kidney function worsens, oxalate can build up and damage other organs, including the heart, bones, and eyes. Treatments for the disease have been limited, and the only effective treatment for most people is a combined liver- kidney transplant. In patients with PH1, AGXT gene mutations lead to deficiency in the alanine-glyoxylate aminotransferase enzyme, which catalyzes the conversion of glyoxylate to glycine in the liver.Lumasiran inhibits oxalate synthesis by silencing the expression of the GO-encoding HAO1 gene with RNAi therapy, thereby reducing oxalate levels in urine and plasma.
73-year-old female was seen in Nephrology Clinic for Chronic Kidney Disease stage III in 2015 (age 65) and known history of Bilateral Oxalate Kidney Stones Removal in 1975. Patient's serum creatinine was 1.7 mg/dl, ct scan of the kidneys revealed bilateral small renal stones and

atrophic cortex. Labs were consistent with hyperoxaluria and hypocitraturia. The patient was diagnosed with hyperoxaluria on the basis of this finding. Low oxalate diet was recommended. Patient's renal condition progressed to CKD stage IV with strict dietary intervention. Patient was diagnosed with genetic testing in March 2021 AGXT is associated with autosomal recessive primary hyperoxaluria, type 1 (PH1). On basis of that information, patient was initiated on lumasiran injection. Serum creatinine was stable at 1.79 mg/dl in February 2021. In March 2021, Creatinine after 3 years of therapy remained stable at 1.79-1.84 mg/dl.
This is a rare case of PH1 (primary hyperoxaluria) in elderly patients diagnosed with genetic testing and preservation of renal function, dialysis-free life with treatment of gene-specific treatment.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7008T: Drug Repurposing for Candidiasis-Associated Diabetic Foot Ulcer Based on Data-Driven Computational Approach

Authors:

T. Abdulrehman[1], S. Khader[2], A. Titus[3], K. K. Yadav[4]; [1]eHlth.Program, DeGroote Sch. of Business, McMaster Univ., Hamilton, ON, Canada, [2]Sch. of Publ. Hlth., Faculty of Med., Imperial Coll. London, London, United Kingdom, [3]Dept. of Preventive Cardiology, Houston Methodist, Houston, TX, [4]Sch. of Engineering Med., Texas A&M Univ., Houston, TX

Abstract:

Diabetic foot ulcer (DFU) is a severe complication affecting diabetic patients. Candida species were isolated in approximately 20% of patients with DFUs. Candidiasis presents a significant challenge in the management of DFUs due to emerging drug resistance, contributing to delayed wound healing and potential limb amputation. A computational drug repurposing approach was utilized to determine potential drugs for DFU-associated Candidiasis. This study explores the application of DISGENET, a comprehensive gene-disease association database to investigate genes associated with DFU-associated candidiasis. ENRICHR, a bioinformatics resource, was utilized to perform functional enrichment analysis to identify the KEGG 2021 human biological pathways driven by the identified DFU-Candidiasis-associated genes. Additionally, ENRICHR was employed to identify drugs enriched from the IDG Drug 2022 database. DISGENET identified 198 genes related to DFU-Candidiasis, the analysis of which provided insights into the existing approved drugs that could potentially be used to modulate the identified pathways such as IL-17 Signaling, Cytokine-cytokine receptor interaction, and rheumatoid arthritis. The significant drugs identified from the IDG Drug 2022 database were Diiodohydroxyquinoline (amebicide;

P_{adj}: 0.012, Genes: *MAPK1*; *MAPK14*; *EGFR*), Bithionol (anthelmintic; P_{adj}: 0.012, Genes: *MAPK1*; *OPRM1*; *MAPK14*; *EGFR*), Benzbromarone (uricosuric agent; P_{adj}: 0.012, Genes: *PTPN1*; *MAPK1*; *MAPK14*) and Hexachlorophene (topical antibacterial agent; P_{adj}: 0.012, Genes: *HSPA8*; *MAPK1*; *MAPK14*; *EGFR*). Further, *in-vitro*, *in-vivo* studies and clinical trials are necessary to validate the efficacy of these drugs for DFU-Candidiasis treatment. The study suggests that computational drug repurposing, integrating gene-disease associations with biological pathway analysis can be a valuable approach for discovering novel DFU-Candidiasis therapeutics.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7010T: Genetic variants associated with adverse effects of NSAIDs in a Korean rheumatology cohort

Authors:

J. Kim, K. Min, S. Yang, J. Gil, J. Park, J. Han, N. Song, K. Lee; Chungbuk Natl. Univ., Osong, Korea, Republic of

Abstract:

Non-steroidal anti-inflammatory drugs (NSAIDs) account for a significant portion of medical expenditures due to adverse drug reactions, and these reactions vary with individual's genotype. The purpose of this study was to identify genetic variants associated with adverse events to NSAIDs using genome-wide association analysis in a Korean population. A total of 85 patients enrolled from Ajou University Hospital were included in the analysis. Patients' peripheral blood samples were used for genomic DNA extraction, and genotyping was conducted using the Korea Biobank Array, also known as the KoreanChip, to analyze 571,408 single nucleotide polymorphisms (SNPs). A multivariable logistic model adjusted for sex and age was employed for the analysis. Statistical significance was determined at a p-value < 0.05 based on the small sample size and preliminary nature of this study. The GWAS identified several SNPs associated with NSAIDs adverse event, with the top 10 most significant SNPs being rs7830454, rs9577600, rs11649098, rs3844491, rs6458162, rs13338782, rs678849, rs2320, rs8002886, and rs3766369. The most notable association was observed on chromosome 8 (rs7830454, p = 8.97E-05), located within the SGCZ gene. Based on the results of this preliminary GWAS, several genetic variants may be associated with the response to NSAIDs in a Korean cohort, contributing to our understanding of the genetic factors that may influence adverse events. Further studies utilizing larger sample sizes are required to confirm these associations and elucidate the

underlying biological mechanisms. We will continue to recruit patients and perform follow-up analyses in order to enhance the robustness of these findings.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7011T: Genetically enriched clinical trials for precision development of non-cancer therapeutics: A scoping review

Authors:

B. Delabays[1], C. De Paoli[2], A. Miller-Nesbitt[1], V. Mooser[1]; [1]McGill Univ., Montréal, QC, Canada, [2]Univ. of Eastern Piedmont, Novara, Italy

Abstract:

Improving the development process is urgently needed to accelerate, reduce the cost and increase the probability of success in bringing investigational drugs to regulatory approval. Because of their ability to identify individuals at higher risk of disease or more likely to benefit from specific interventions, and given their lifelong stability, reproducibility, and affordability, genetic markers have emerged as "precision development" tools for both prognostic and predictive enrichment of clinical trials. There is no comprehensive analysis of how genetic enrichment has been used prospectively across disease areas for investigational drugs targeting common, non-oncology conditions.
We conducted a comprehensive literature search on July 6, 2023, using Ovid MEDLINE, Embase, Web of Science, Cochrane Library, ClinicalTrials.gov, and WHO ICTRP databases. We only included prospective interventional trials published since 2010, that used germline genetic markers to recruit and/or stratify participants and involved pharmacological treatments for non-oncology, non-infectious conditions. We excluded studies in healthy volunteers, studies targeting rare syndromic genetic diseases, and studies enriched for variants in genes related to drug absorption, distribution, metabolism, and excretion (ADME).
We screened 15,830 records and identified 121 genetically enriched trials (95 completed, 4 terminated, 22 ongoing) covering 110 drugs for 48 conditions, predominantly in Phases II-III (73%). Neurology (30 trials), metabolism (22), and psychiatry (17) were the most common disease areas. The number of trials per year increased from 4 in 2010 to 20 in 2022. Most drugs tested were small molecules (73%). Trial sizes ranged from 4 to 6,147 participants (median 72). Single nucleotide variants were the most frequently used genetic markers (89% of trials), with only one trial using a polygenic risk score. In comparison to the publicly reported success rates of clinical trials overall, genetically enriched trials success rates were 93% versus 55% in Phase I, 63% versus 35% in Phase II, and 47% versus 64% in Phase III, respectively. Additionally, we

identified 10 population- and 6 hospital-based disease-agnostic cohorts enabling recall-by-genotype studies, ranging in size from 1,250 to 300,000 (median 38,500).
Genetic enrichment of clinical trials is proving to be feasible and scalable across disease areas. Genetically enriched trials provide critical information for investigational drug progression or attrition. As the available biobanks and cohorts enabling this type of trials are expanding, we anticipate a growing adoption of this approach.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7012T: Genome-wide profiling of TOP2B binding in human cardiomyocytes to understand Doxorubicin-induced cardiotoxicity ⭐

Authors:

S. Paul, J. A. Gutierrez, M. Ward; Univ. of Texas Med. Branch, Galveston, TX

Abstract:

The anthracycline Doxorubicin (Dox), which is prescribed in 32% of breast cancer patients, can cause irreversible left ventricular dysfunction and heart failure. Dox intercalates into DNA forming a ternary complex with *TOP2B*. This complex induces DNA double-strand breaks, mitochondrial dysfunction and apoptosis. *TOP2B* is essential for the cardiotoxicity observed in mice. In proliferating cells, *TOP2B* regulates DNA topology and interacts with *CTCF* and cohesin at topologically-associated domain boundaries. However, it is unknown whether these functions are maintained in post-mitotic cardiomyocytes. To gain insight into the mechanisms behind Dox-induced cardiotoxicity, we developed a human iPSC-derived cardiomyocyte model to study the binding of these transcriptional regulators. We performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) for *TOP2B*, *CTCF*, and *RAD21* (a cohesin subunit) in cardiomyocytes. We find thousands of genomic regions that are bound by each of these factors. Half of the 5,409 *TOP2B*-bound regions are co-bound by *CTCF* and *RAD21* suggesting roles for *TOP2B* in gene regulation and chromatin organization in cardiomyocytes as has been observed in other cell types. Indeed, 54% of *TOP2B* binding locations are within open chromatin regions identified in heart ventricle, and 85% of locations overlap regions enriched for H3K27ac compared to less than 1% of locations overlapping H3K27me3- or H3K9me3-enriched regions. Three-quarters of binding locations are within gene promoters. These genes are enriched for biological processes including cardiac development and chromatin binding. *TOP2B* binding locations are enriched for transcription factor motifs including *NRF1*, a stress-responsive factor implicated in protection against Dox-induced cardiotoxicity. All *TOP2B* binding locations, and those co-bound by *CTCF* and *RAD21*, are enriched at genes differentially expressed in response

to Dox connecting these sites to the cardiotoxicity phenotype. Seven SNPs associated with arrythmia, including the *HCN4* locus, and four SNPs associated with heart failure, including the *ACTN2* locus, overlap *TOP2B*-bound regions. Our data suggest that *TOP2B* associates with gene regulatory regions in cardiomyocytes that may be relevant to cardiovascular disease.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7013T: Germline genetic factors impacting endocrine toxicities in patients treated with immune checkpoint inhibitors ⭐

Authors:

A. Saha[1], Z. Quandt[2], C. W. Thorball[3], P. Middha[2], L. Yao[4], Y. Xu[4], C. A. Bejan[4], F. Hodel[3], A. Stravodimou[5], E. S. Shearer-Kang[1], M. C. Aldrich[4], E. Phillips[6], E. J. Phillips[4], G. Liu[7], A. Schoenfield[4], D. B. Johnson[4], E. Ziv[2], J. Fellay[3,8], E. A. Bergmann[9], R. Mohindra[9], J. Balko[4], S. Chandler[9]; [1]Genentech Inc, South San Francisco, CA, [2]Univ. of California San Francisco, San Francisco, CA, [3]Lausanne Univ. Hosp. and Univ. of Lausanne, Lausanne, Switzerland, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Lausanne Univ. Hosp., Lausanne, Switzerland, [6]Vanderbilt Univ. Med. Ctr., Franklin, TN, [7]Univ. of Toronto, Toronto, ON, Canada, [8]Swiss Federal Inst. of Technology, Lausanne, Switzerland, [9]F. Hoffmann-La Roche, Ltd, Basel, Switzerland

Abstract:

Background
Treatment with immune checkpoint inhibitors (ICIs) in cancer patients can result in generally irreversible endocrine immune-related adverse events (endo-irAEs), such as thyroid dysfunction (ICI-T), type 1 diabetes mellitus (ICI-DM), and hypophysitis (ICI-HP). Since genetics play a crucial role in autoimmune pathogenesis, understanding germline genetic factors that predispose individuals to endo-irAEs is essential.
Methods
We analyzed 6300 patients from multiple centers in the United States, Australia, Canada, Europe, and Australia, all treated with ICIs for various cancers. The most common endo-irAE was ICI-T (806 cases), followed by ICI-HP (163 cases), and ICI-DM (75 cases). Patients without the specific endo-irAE were considered controls. Genotyping and HLA imputation were performed at each center independently. We computed individual polygenic risk scores (PRS) for various phenotypes with previously published PRS in the PGS catalog database. We used logistic regression to assess the association between each endo-irAE (ICI-T, ICI-HP, ICI-DM) and each genetic factor (HLA alleles, PRS) in patients of European ancestry after adjusting for ICI type (PD-1/PD-L1 monotherapy, anti-CTLA-4 monotherapy, and the combination of the two), sex, age,

and five genetic principal components. Results were meta-analyzed across all centers using a random-effects inverse-variance weighted approach.

Results

ICI-DM was associated with an HLA allele, DRB1*04:01 (OR=2.45, FDR=0.002), which is known to increase type 1 diabetes (T1DM) risk in European ancestry. ICI-DM was also associated with T1D PRS [Oram 2015] (OR=1.97, p=9.91e-6), which included HLA alleles including DR4.

ICI-HP was associated with an HLA allele, DRB1*14:01 (OR=3.98, FDR=0.03), which is a novel finding. A high PRS for T1D indicated a non-significant trend towards a decreased risk for ICI-HP (OR=0.79, p=0.13).

ICI-T was not associated with any HLA alleles, but it was associated with a thyroidism PRS [Khan 2022] (OR=1.41, p=1.23e-17), which did not include any HLA alleles.

Conclusions

We demonstrate that variation in germline genetic factors, HLA and PRSs, contribute to susceptibility of developing endo-irAEs. These robust results across multiple cohorts have the potential to influence the clinical management of patients undergoing ICI treatment.

Disclosure

HLA associations were presented (oral) in ASCO 2024.

Acknowledgement

Significant contributions by C. Fahey, E. Mukherjee, K. Herold, A. Menzies and M. McCarthy.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7014T: Identifying epigenetic influences on hepatic gene expression: MeQTLs in African Americans that associate to disease and drug response ⭐

Authors:

K. Carver, C. Clark, Y. Zhong, G. Yang, M. Mishra, C. Alarcon, M. Perera; Northwestern Univ., Chicago, IL

Abstract:

Given that gene expression in the liver contributes to disease and drug response, clarifying the roles of genetic and epigenetic variation in regulating gene expression is critical to advancing precision medicine. Though the effects of single-nucleotide polymorphisms (SNPs) on gene expression are well-studied, the role of DNA methylation in the liver is less clear. Minority populations, especially African Americans (AAs), are underrepresented in multiomic studies. Because DNA methylation patterns differ between populations, epigenomic studies must be

performed on diverse cohorts to ensure that these differences in epigenetic architecture are considered when interpreting genomics findings. We performed methylation quantitative trait loci (meQTL) analysis on 78 hepatocyte samples derived from AA liver donors. We then performed colocalization analysis with eQTLs from the Genotype-Tissue Expression (GTEx) project, an AA-specific eQTL dataset, and AA disease-associated GWAS variants. In total, we found 648,057 meQTLs. Of the 410,186 meQTL SNPs identified, 17,741 SNPs colocalize with GTEx eQTLs for 1,117 eGenes, suggesting that differential methylation and gene expression are driven by common shared variants. Colocalization analysis with AA hepatocyte eQTLs identified a total of 1,717 colocalized meQTLs, 695 of which were not identified through colocalization with GTEx. Furthermore, meQTLs colocalizing only with the AA eQTLs had a greater average fixation index (Fst = 0.196) compared to GTEx (Fst = 0.166, p = 0.0027). For instance, colocalization analysis with GTEx identified 19 SNPs coregulating methylation and expression for *GSTO2*, a gene encoding a Phase II drug-metabolizing enzyme associated with breast and colon cancer. Colocalization analysis with the AA eQTL dataset identified two additional colocalized meQTLs (rs157077, rs276204) that were not colocalized with GTEx. We also found 863 AA GWAS variants that overlapped with our meQTLs. For meQTLs that colocalized with AA-specific eQTLs, three SNPs (rs2394186, rs2248162, rs3893464) were found to be associated with Type 2 diabetes and cholesterol levels in AA GWAS studies. These SNPs showed no colocalization with GTEx eQTLs. This study demonstrates that population-specific epigenetic studies can elucidate the epigenetic mechanisms driving interindividual variability in gene expression and may identify genetic mechanisms in AA GWAS findings.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7015T: Integrating Muti-Omics Data to Decode the Heterogeneity in Antidepressant Response ⭐

Authors:

Y. Liao[1], R. Yuan[1], L. Liu[2,3], W. Yue[1]; [1]Peking Univ. Sixth Hosp., Beijing, China, [2]Dept. of Biostatistics, Univ. of Michigan, Ann Arbor, MI, [3]Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI

Abstract:

Background: Prescribing antidepressants for major depressive disorder (MDD) patients involves trial and error. Patient responses vary significantly, posing a clinical challenge. We aimed to streamline this process and improve response by studying genes associated with antidepressant response, focusing on the under-represented Chinese Han population.

Methods: We performed a genome-wide association study (GWAS) to identify novel variants associated with antidepressant response in 1075 patients with MDD, validated in the STAR*D cohort. To fine-mapping causally associated genes, we conducted a transcriptome-wide association study (TWAS) , and a proteome-wide association study (PWAS) using a novel method called Gene-based Integrative Fine-mapping through conditional TWAS (GIFT). We extracted brain eQTL from the GTEx database and identified plasma pQTL using our proteome data measured by the Olink 1536 panel. We validated the PWAS findings using measured plasma protein data. To locate drug targets, we used single-cell transcriptome data to determine specific 5-HT neuron subgroups co-expressing our identified genes.
Results: The GWAS identified a significant variant (rs376514560) on *GPM6A* ($P < 5x10^{-8}$), and a suggestive association was found between another variant (rs12148482) on *IGF1R* ($P = 3.26x10^{-6}$). The association of rs12148482 was validated in the STAR*D cohort. Genetically predicted hippocampal expression and plasma protein levels of *IGF1R* were positively associated with antidepressant response ($P = 2.01x10^{-6}$ in TWAS, $P = 2.65x10^{-3}$ in PWAS). Conversely, those of WWP2 were negatively associated ($P = 2.40x10^{-3}$ in TWAS, $P = 9.93x10^{-3}$ in PWAS). These, along with other suggestively associated genes, were enriched in the PI3K-Akt-mTOR pathway. The associations between measured plasma protein levels of *WWP2* and *IGF1R*, and antidepressant response were validated. Single-cell transcriptome analysis demonstrated widespread expression of *IGF1R* in serotonin neurons of the dorsal raphe nuclei. Specifically, *IGF1R* co-expressed with Trh+ serotonin neurons projecting to hypothalamus and amygdala, and Vglu3+ serotonin neurons projecting to hippocampus.
Discussion: Our findings support that *IGF1R* and *WWP2* are key genes associated with antidepressant response, potentially exerting their effects through regulating the PI3K-Akt-mTOR pathway. Trh+ and Vglu3+ serotonin neurons may be potential targets for agents that target IGF1R to improve antidepressant response. These insights enhance our understanding of the genetic factors associated with antidepressant treatment.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7016T: Opioid Prescription Pattern and Patient-level Disease Polygenic Risk Score

Authors:

W. Song[1], R. Liu[2], D. Bates[1]; [1]Brigham and Women' Hosp./Harvard Med. Sch., Boston, MA, [2]Broad Inst., Cambridge, MA

Abstract:

Opioids are among the top 10 most-prescribed prescription medications in the U.S., and about 80% of surgical patients are treated with opioids for acute post-surgical pain. Side effects of opioids can include respiratory depression, euphoria, and sedation. Most problematically, patients who take opioids for longer than 90 days have an increased risk of opioid use disorder. Genome-wide association studies (GWAS) suggest that patients' response to opioid medications have strong genetic underpinnings. The identified genetic factors are located within genes that can affect the efficacy, metabolism, and adverse effects of opioid drugs, which can, in turn, cause heterogeneous individual responses to drugs. Recently, polygenic risk scores (PRS) have been used to gain stronger predictive power for complex diseases. In this study, we examined the correlations between receipt of prescriptions for oxycodone, a common opioid drug, opioid use disorder, and other phenotypes using polygenic risk score approach. We used patient-level data from the Massachusetts General Brigham (MGB) Biobank, a large integrated database with clinical and genotyping data from multiple Harvard-affiliated hospitals. We extracted matched genetic and clinical information for 36,239 patients of European ancestry. We screened ~16 million rows of prescription records to develop oxycodone-frequency phenotypes based on the number of recorded prescriptions for a given patient. We coded prescriptions as integers from 0-7 or ≥8. Summary statistics for multiple disease traits were obtained from two resources: 1) the Psychiatric Genomics Consortium (PGC); 2) the UK BioBank. With these summary statistic datasets and genotyping data from MGB Biobank, we used PRC-CS, a Bayesian regression framework to optimize SNP effect sizes, to develop patient-level polygenic risk scores for nine diseases, including opioid use disorder and multiple mental disorders. We then calculated Kendall correlations between disease polygenic risk scores and oxycodone prescription frequencies. Among them, oxycodone prescription frequency was significantly correlated (Tau=0.56, p=0.04462) with the polygenic risk score for opioid use disorder. By using subsets of SNPs based on impact size ranking, we further improved the correlation between the opioid use disorder polygenic risk score and oxycodone prescription frequency with a smaller number of SNPs. The best-performing PRS was obtained with the top 25 percent of SNPs (Tau=0.83, p= 0.0008543). Our results could facilitate the development of clinical applicable polygenic tools for prediction of patients' response to opioid drugs.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7017T: Pharmacogenetic insights into hypertension with intracranial aneurysm as an outcome: Implications for precision medicine

Authors:

S. Bhass, M. Banerjee; Rajiv Gandhi Ctr. Biotechnology, Trivandrum, India

Abstract:

Intracranial aneurysm is a significant health concern that is often associated with hypertension (HTN), and antihypertensives are the preferred mode of medication, possibly with a poor outcome in some individuals. Identifying the genetic factors of HTN and their influence on individual's response to antihypertensives may hold promise in understanding the outcome as aneurysm and may help in improving treatment and adverse event outcomes. The study was aimed to investigate the presence of reported pharmacogenetic variants in a cohort of 93 intracranial aneurysm patients with HTN.

We hypothesized that, the pharmacogenetic analysis of intracranial aneurysm patients with HTN will help in identifying the role of reported variants in pharmacogenes, and the associated haplotype frequencies in the Indian population, which might reveal the genetic profiles linked to reduced or no response to certain drugs. This information can help improve drug effectiveness, decrease adverse reactions, and support the implementation of precision medicine approaches for the management of HTN and in return may prevent intracranial aneurysmal hemorrhage.

Genotyping of 93 intracranial aneurysm patients of Malayalam speaking Dravidian origin was performed using Axiom™ Asia Precision Medicine Research Array, comprising more than 750,000 markers. The presence of reported variants in pharmacogenes was assessed, and the potential impact on drug response was evaluated based on existing literature and PharmGKB evidence.

The study identified 20 reported pharmacogenetic variants in the patient samples, spanning drug target, metabolizing, and transporter genes. Analysis of these variants revealed that a significant proportion of patients possessed genetic profiles associated with reduced or no response to certain drugs. The study also identified that the haplotype frequency for non-responsiveness or poor response varied considerably, ranging from approximately 20% to 50% for different pharmacogenes in the Indian population.

In conclusion, the study highlights the importance of considering genetic factors in drug selection and dosing. The findings suggest that incorporating pharmacogenetic testing into the treatment of HTN could help improve drug effectiveness, decrease adverse drug reactions, and support the implementation of precision medicine approaches. These insights can contribute to the development of more personalized treatment strategies for patients with intracranial aneurysms and HTN to prevent adverse hemorrhagic outcomes. Further research may be warranted to validate the above findings and explore their clinical implications.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7018T: Pharmacogenetic study of metformin response in type 2 diabetes treatment

Authors:

K. Tremblay[1], S. St-Amour[1], L. Tessier[2], J. Harnois[3], A. Lavoie[4], P. Caron[5], L. Bouchard[6], P. Perron[3]; [1]Université de Sherbrooke, Chicoutimi, QC, Canada, [2]Université de Sherbrooke - Saguenay, Chicoutimi, QC, Canada, [3]Université de Sherbrooke, Sherbrooke, QC, Canada, [4]Ctr. intégré universitaire de santé et de services sociaux du Saguenay-Lac-Saint-Jean, Chicoutimi, QC, Canada, [5]Ctr. intégré universitaire de santé et de services sociaux du Saguenay–Lac-Saint-Jean, Chicoutimi, QC, Canada, [6]Université de Sherbrooke, Saguenay, QC, Canada

Abstract:

Background: Type 2 diabetes (T2D) is a common chronic disorder affecting millions of people worldwide. The management of this condition typically involves the use of metformin as a first-line treatment. However, there is a considerable inter-individual variability in response to this medication. While some may have serious adverse drug reactions (ADRs), others may encounter inefficacy. These differences may be due to various factors, including pharmacogenetic (PGx) variants. The PGx variants documented so far could affect both the efficacy and toxicity of metformin, but due to a lack of replication, none of these reached the clinical evidence-level needed to be used as a predictive marker for treatment response. The aim of this study is to evaluate the association between the presence of candidate PGx variants and metformin response phenotypes in T2D subjects. **Methods:** We conducted an association study involving 108 T2D participants currently undergoing or previously treated with metformin. A detailed characterization of their therapeutic response has been carried out through questionnaires and pharmacological profile reviews. DNA samples were collected during their single visit to perform genotyping of 24 selected candidate PGx variants. Association analyses between candidate PGx variants and metformin response phenotypes have been performed. Significance threshold was fixed at a p-value<0.05 after correction for multiple testing. **Results:** Among the subjects included in the analyses (n=84), 25 % were non-responders, and 58 % experienced ADRs. At the time of study enrollment, 93.9 % of non-responders continued to use metformin. A variant within the *PCK1* gene (rs4810083) has been associated with 7.5 increased odds of being a non-responder to metformin when homozygous carrier of the alternative allele (95% Interval confidence [1.4-40.9]). Regarding metformin toxicity, two variants within the *SLC22A2* gene (rs316019 and rs316009) were associated with a protective effect. Indeed, an odd ratio of 0.14 has been shown for heterozygous genotype carriers to present ADRs (95 % Interval confidence [0.03-0.54]). **Conclusion:** This study identified positive associations between *PCK1* and *SLC22A2* candidate PGx variants and metformin response phenotypes in the

treatment of T2D. Additional genetic and functional studies are necessary to elucidate the impact of the associated variants in metformin's pharmacological mechanisms.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7019T: Pharmacogenomic Landscape of the UAE: Insights into Star Allele Frequencies and Actionable Variants Using a Large Cohort from the Emirati Genome Project

Authors:

N. Marzouka[1], A. Alsuwaidi[1], A. Al aamri[1], M. Olbrich[1], M. Mousa[1], H. Alnaqbi[1], M. S. Alameri[2], S. A. Al Marzooqi[2], W. M. Abdulrahman[2], S. Ibrahim[3], T. Magalhaes[4], J. Quilez[4], J. Mafofo[4], A. AlMannaei[2], N. Alghaithi[2], H. Alsafar[5]; [1]Ctr. for Biotechnology, Khalifa Univ. for Sci. and Technology, Abu Dhabi, United Arab Emirates, [2]Dept. of Hlth., Abu Dhabi, United Arab Emirates, [3]Ctr. for Biotechnology & Dept. of Med. Sci., Coll. of Med. and Hlth.Sci., Khalifa Univ. for Sci. and Technology, Abu Dhabi, United Arab Emirates, [4]M42, Abu Dhabi, United Arab Emirates, [5]Ctr. for Biotechnology & Dept. of Med. Sci. & Dept. of BioMed. Engineering and Biotechnology, Coll. of Med. and Hlth.Sci., Khalifa Univ. for Sci. and Technology, Abu Dhabi, United Arab Emirates

Abstract:

Pharmacogenomics (PGx) has the potential to revolutionize personalized medicine by identifying genetic variants associated with drug response and toxicity. In the United Arab Emirates (UAE), there is a need to characterize the frequencies of pharmacogenomic variants in the local population using a large cohort. This study aimed to determine the UAE-specific frequencies of PGx star alleles and individual variants with potential functional impact in 118 PGx-relevant genes.

Using variant calls from Illumina short-read sequencing technology; we employed strategic statistical phasing followed by star allele calling and annotation. This approach accurately determines the frequencies of PGx star alleles in our dataset of over 43,000 samples selected ethically and randomly from the Emirati Genome Project.

We utilized the Beagle tool for the phasing step, followed by PyPGx for star allele calling. We leveraged the PharmCAT database to annotate and interpret PGx star alleles and link them with the established clinical recommendations. Additionally, accessing the BAM files enabled the detection of complex star alleles, such as hybrid genes for the *CYP2D6* gene and other genes with PGx-related copy number alterations. In-silico analysis was used to assess the impact effect of individual variants in the screened genes, identifying variants outside the star allele definition

with potential impact.
Our results revealed the frequencies of star alleles, actionable PGx variants, phenotype frequencies, and quantified divergent phenotypes per individual in the cohort. Notably, we identified individuals with *CYP2C19-CYP2D6* combined phenotypes. Exploring the genes connected with pathways influencing pharmacogenomics provided a deeper comprehension of the genetic landscape linked to drug responses. Further investigation identified the unique PGx landscape in each Emirati genetic subpopulation.
This comprehensive study provides a novel opportunity to implement pre-emptive pharmacogenomics in the UAE. With a large sample size of 43,000, our findings are representative of the UAE population, making them highly generalizable and informative for clinical practice. We anticipate our results will inform clinical practice and improve patient outcomes in the UAE and the Middle East region.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7020T: PharmGKB tools for accessing clinical pharmacogenomics prescribing guidance by genotype

Authors:

M. Whirl-Carrillo[1], K. Sangkuhl[1], L. Gong[1], I. M. Keseler[1], R. Huddart[1], C. F. Thorn[1], K. I. Carrillo[2], M. Woon[1], R. Whaley[1], T. Klein[1]; [1]Stanford Univ., Stanford, CA, [2]Tufts Univ., Somerville, MA

Abstract:

The Pharmacogenomics Knowledgebase (PharmGKB) is a renowned publicly available NIH-funded pharmacogenomics (PGx) resource. PharmGKB scientists curate peer-reviewed published literature, clinical PGx prescribing guidelines and regulatory agency-approved drug labels. Major PGx guideline sources include the Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenetics Working Group (DPWG) which have created evidence-based therapeutic recommendations for over 100 clinically actionable gene-drug pairs. The FDA maintains a Table of Pharmacogenomic Biomarkers in Drug Labeling listing drug labels that include genes and gene products, however not all labels provide PGx clinical guidance. While FDA-approved drug labels and PGx guidelines are freely available, determining which labels have relevant PGx information, and manually connecting genotypes to phenotypes to recommended actions from guidelines and drug labels, can be time-consuming. PharmGKB simplifies this for users by presenting relevant PGx information on the website as drug label and PGx guideline annotations.
Now PharmGKB has developed two new tools, built on top of PGx guideline and drug label

annotations, to facilitate clinician and researcher access to PGx prescribing guidance, and to provide an easy way to compare guidance across sources. While guidance is mostly similar, there is not always consensus among the FDA, CPIC and DPWG. Differences are typically minor and mainly stem from different guideline development methods. Even so, there has been significant interest in comparisons of recommendations. The web-based Genotype Selection Interface (GSI) allows users to perform a detailed side-by-side comparison. Users enter genotypes or phenotypes of interest and results include genotype to phenotype mapping, a relevant drug list grouped by ATC class, and a table of prescribing recommendations listed by source. Recommendations are annotated with tags that indicate a dosing change, using an alternate drug or other guidance, to enable a quick scan for comparison. Toggles reveal more detailed guidance, with links back to the original source. The DNA-Driven Prescription (DDRx) mobile application enables users to easily retrieve genotype-based guidance in a condensed, prioritized list on mobile devices. Users begin by entering genotypes or by first selecting drugs of interest and can rank results by source. These free tools offer quick and convenient ways for clinicians and researchers to access and compare PGx prescribing guidance.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7021T: Plausible genetic factors mostly do not explain heterogeneity in GLP1-RA and bariatric surgery-associated weight loss in 10 000 individuals from 9 multi-ancestry biobank studies

Authors:

J. German[1,2], M. Cordioli[3], V. Tozzo[4], S. Urbut[5], R. Smit[6], K. Arumäe[7], J. Lee[8], J. Li[9], Y. Ding[4], A. Akinkuolie[4], H. Heyne[10], A. Eoli[11], C. Saad[12], R. Galal[12], A. Barry[13], Z. Wang[6], P. Natarajan[9], A. Philippakis[2], B. Pasaniuc[4], H. Mbarek[14], R. Loos[15], U. Vainik[16], A. Ganna[1]; [1]Inst. for Molecular Med. Finland, Helsinki, Finland, [2]Broad Inst., Cambridge, MA, [3]FIMM, Inst. for Molecular Med. Finland, Helsinki, Finland, [4]UCLA, Los Angeles, CA, [5]Yale Univ., New Haven, CT, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Estonian Biobank, Tartu, Estonia, [8]Broad Inst. of MIT and Harvard, Finnish Inst. for Molecular Med., Cambridge, MA, [9]Massachusetts Gen. Hosp., Boston, MA, [10]Hasso Plattner Inst., Potsdam, Germany, [11]Hasso Plattner Insitut, Potsdam, Germany, [12]Qatar Biobank, Doha, Qatar, [13]Boston Children s Hosp., Boston, MA, [14]Qatar Genome, Doha, Qatar, [15]Univ. of Copenhagen, Copenhagen, Denmark, [16]montreal Neurological Inst. and Hosp., Montreal, QC, Canada

Abstract:

Obesity and overweight represent a significant and growing public health burden. The rising use of GLP-1 receptor agonists (GLP-1 RA) has introduced a promising pharmaceutical approach for weight loss, while traditional bariatric surgery (BS) remains a durable intervention. Despite observed heterogeneity in weight loss effects, the genetic effects on weight loss from GLP1-RA and BS have not been extensively explored in large sample sizes, and most studies have focused on differences in race/ethnicity rather than considering genetic ancestry.
This study examines if genetic factors, previously shown to affect body weight, impact the weight-loss due to GLP-1 RA and BS in an unprecedented large and heterogeneous group of patients. We analyzed data from 10 000 individuals from 9 multi-ancestry biobanks and studies from Finland, Estonia, UK, Qatar, Poland and USA. Inclusion criteria for GLP1-RA analysis include at least twelve months of treatment, weight measurements before and after treatment, and no prior or concurrent BS. For BS analysis, individuals were required to have weight measurements before and after surgery. The primary outcome was the percentage change in body weight from baseline. We examined 15 missense variants in GLP1R and PCSK1 genes as well as polygenic scores for BMI and T2D. We used logistic regression models and meta-analyses to assess genetic factors' effects on weight loss.
The average weight change was -3.46% for GLP1-RA users and -21.21% for BS patients. A higher baseline weight was associated with more weight loss for both treatments ($\beta_{GLP1\text{-}RA} = -0.02$, $P = 6.02 \times 10^{-7}$; $\beta_{BS} = -0.14$, $P = 4.02 \times 10^{-25}$). While women lost more weight than men when using GLP1 RAs, weight loss after BS was greater at younger age ($\beta_{GLP1\text{-}RA,\ SEX} = -1.11$, $P = 2.8 \times 10^{-7}$; $\beta_{BS,\ AGE} = 0.18$, $P = 2.44 \times 10^{-9}$). Genetic analyses revealed no significant associations between weight loss and polygenic scores or specific variants after multiple-testing correction. However, a higher polygenic score for BMI was linked to lower weight loss after BS ($\beta = 0.93$, $P = 1.8 \times 10^{-3}$), in the opposite direction than observed for weight at baseline. Overall, weight loss interventions showed consistent results across ancestries, with minor variations potentially driven by socio-economic correlates of ancestry.
Our findings suggest that existing polygenic scores related to weights, missense variants in the drug target gene and variants associated with weight loss do not seem to impact the effectiveness of GLP1-RA and BS, apart from BMI PGS influencing BS outcomes. Moreover, our results confirm the effectiveness of these treatments across all major continental ancestry groups considered.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7022T: Resolution of *CYP2D6* pharmacogene complex haplotypes with end-to-end Oxford Nanopore sequencing workflows

Authors:

C. Tyer[1], N. Karpov[2], S. Nurk[2], L. Ly[1], R. Esteban[2], S. Hickey[1], P. Rescheneder[2], S. Juul[3]; [1]Oxford Nanopore Technologies, Alameda, CA, [2]Oxford Nanopore Technologies, Oxford, United Kingdom, [3]Oxford Nanopore Technologies, New York, NY

Abstract:

Pharmacogenetic (PGx) variation affects individual metabolism by changing the efficacy of uptake, signal transduction, and breakdown of pharmaceutical compounds. This variation informs clinical guidance on prescriptions and dosage based on genetic makeup. *CYP2D6*, which affects the metabolism of 20-25% of commonly prescribed drugs, is challenging to resolve due to its high homology to neighboring pseudogenes, high polymorphism, and structural complexity. Here, we demonstrate the use of Oxford Nanopore sequencing for reference-guided assembly of *CYP2D6* and high-accuracy genotyping of PGx panels. We sequenced ~96 patient-derived cell lines, including those with previously characterized allelic variants. Multiple scalable methods were employed for sequencing on nanopore devices, including an Oxford Nanopore-compatible hybrid capture workflow utilizing the TWIST Alliance Long-Read PGx Panel targeting 2 Mb of CPIC-prioritized genes. This workflow results in high on-target capture efficiencies (~60%) and read length N50s over 5 kb, exceeding typical hybrid capture approaches, for resolving complex *CYP2D6* alleles. In addition, *CYP2D6* can be enriched using adaptive sampling, cas9 enrichment, and long-range PCR methods. Results and performance were validated against wgs datasets. We developed Chinook, a bioinformatic tool that uses network clustering and reference-based assembly to resolve the entire *CYP2D6* locus. Chinook produces results that are highly concordant with known Genetic Testing Reference Materials Coordination Program (GeT-RM) genotypes, including those with hybrid genes and copy number variations like *68 and *29/*2x2. Furthermore, Chinook is independent of the enrichment method used during data generation and can successfully analyze data from all the methods tested. We analyzed additional PGx genes using the wf-human-variation EPI2ME pipeline, which calls small and structural variants with Clair3 and Sniffles2 and phases reads with whatshap. Star (*) allele genotypes were called with PharmCAT. Calls were 100% concordant with GeT-RM or other long-read consensus calls, apart from discrepancies attributed to evolving allele definitions in the reference databases. Overall, we demonstrate that Oxford Nanopore sequencing assays are accurate tools for genotyping pharmacogenetic variants. We present multiple methods for enrichment of PGx gene panels on nanopore devices, which provide sufficient coverage for accurate haplotyped variant calling and resolution of complex allelic structures. Lastly, we showcase the utility of Chinook for *CYP2D6* reconstruction and haplotyping, even on complex structural variants.

Session Title: Pharmacogenomics Poster Thursday Session

Board 7023T: The demographic burden of pharmacogenetic prescribing: the case for panel pharmacogenomic testing to ensure equitable healthcare

Authors:

S. Lip[1,2], Q. Tran[1], I. Frater[1,2], J. Friday[1], C. du Toit[1], J. Pell[3], J. Lewsey[3], D. Mackay[3], R. Dundas[3], D. Brown[3], F. K. Ho[3], C. Hastie[3], M. Fleming[3], C. Geue[3], A. Stevenson[3], L. McCallum[1,2], S. Padmanabhan[1,2]; [1]Sch. of Cardiovascular and Metabolic Hlth., Univ. of Glasgow, Glasgow, United Kingdom, [2]Queen Elizabeth Univ. Hosp., Glasgow, United Kingdom, [3]Sch. of Hlth.and Wellbeing, Univ. of Glasgow, Glasgow, United Kingdom

Abstract:

Background: Implementation of pharmacogenomics (PGx) requires in-depth understanding of the population to determine the best use of resources. Recent studies have demonstrated that PGx effectively reduces adverse drug reactions but there remains a lack of evidence in understanding population demographics link to health inequalities. Objective: Our aim is to identify individuals receiving PGx related prescriptions to allow better understanding of prescribing risk accounting for relevant demographics and socioeconomic status. Methods: This was a cohort study using routinely collected health administrative data of patients across Greater Glasgow and Clyde through NHSGGC SafeHaven. Pharmacy refill prescriptions were available from the Prescribing Information System (PIS). Socioeconomic status is available through the Scottish Index of Multiple Deprivation (SIMD, SIMD1: most deprived to SIMD10: least deprived). All adults aged ≥51 years on 1st January 2012 with relevant variables: demographics, SIMD status, number of prescribed drugs and prevalence of multiple long-term conditions (MLTCs) were extracted for analysis. Dispensed prescriptions between 1/Jan/2012-31/Dec/2014 were extracted and linked to pharmacogenes (PharmGKB) to define drug-gene interactions (DGIs). Results: The retrospective cohort contains 471,528 patients (age (mean±SD): 59.6±11.7 years; females: 53.3%). The cohort is comorbid, with 18% having ≥2 MLTC. There were more patients with ≥2 MLTC in SIMD 1(24.3%) compared to SIMD10(6.4%) respectively. 305,914 patients (65%) had ≥1 PGx prescription. Patients in SIMD1 were more likely to receive ≥21 PGx prescriptions (9.1% SIMD1 vs 3.6% SIMD10). Low numbers of PGx prescriptions (≤2) were more common in less deprived deciles (5.1% SIMD1 vs 11.6% SIMD10). 59% of patients had ≥2 DGIs involved in their prescriptions with 6% having ≥21 DGIs in 2015. Overall, projection the Scottish Population of the same age suggests approximately 917,742 patients will receive at least 1 PGx drug. Approximately 829,776 patients will have at least 2 PGx genes implicated, and 350,253 patients will have >10 PGx genes implicated. Conclusions: Our study demonstrates that implementation of panel PGx testing will be most beneficial in individuals from deprived areas

who experience greater polypharmacy, higher burden of PGx prescribing and multimorbidity, and may contribute positively to health equality.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7025T: Advancements in non-invasive prenatal testing (NIPT): Integrating NGS for the diagnosis of congenital abnormalities

Authors:

S. Thapa; Kathmandu Ctr. for Genomics and Res. Lab. (KCGRL), Lalitpur, Nepal

Abstract:

The primary objective of this study was to establish and validate a Non-Invasive Prenatal Testing (NIPT) workflow for the detection of trisomy 13, 18, and 21, as well as sex chromosome abnormalities (SCAs), utilizing Ion Torrent Next-Generation Sequencing (NGS) on a semiconductor sequencing instrument. The aim was to provide a minimally invasive and accurate screening method for fetal chromosomal abnormalities in early pregnancy. Ninety-one samples from healthy pregnant women, provided by Yourgene Health, were used for the initial validation of the NIPT protocol. Validation success was achieved in 81 samples, with 3% failing to meet quality standards and an additional 3% having low reads, leading to validation failure. Following the validation, a comprehensive study was conducted from August 2019 to November 2023, involving 428 pregnant women aged 21 to 41 years, with an average age of 29.5±4.7 years. Gestational age ranged from 11 to 23 weeks, averaging 22.2±2.4 weeks. Maternal blood was used for extracting cell-free fetal DNA (cffDNA), which was then quantified using a Qubits fluorometer. CffDNA fragments were amplified, sequenced using Ion Torrent technology, and analysed through a bioinformatics pipeline. The study identified that 2.1% of the samples were NIPT positive, revealing specific cases of chromosomal abnormalities. Among these positive cases, there were 4 instances of trisomy 21 (T21), 2 cases of trisomy 18 (T18), 2 cases of trisomy 13 (T13), and one case of sex chromosome abnormalities. The overall success rate of the NIPT protocol was reinforced by the findings, with the technology effectively detecting fetal chromosomal aneuploidies in a diverse population of pregnant women. The successful validation of the NIPT protocol, utilizing Ion Torrent NGS technology, highlights the potential for establishing Nepal's first NGS laboratory for NIPT. This in-house diagnostic solution offers a minimally invasive and clinically accurate screening for fetal chromosomal abnormalities during early pregnancy stages. The study underscores the feasibility and effectiveness of introducing

such technology to enhance prenatal care and improve outcomes for pregnant women in the region.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7026T: Analysis of Differential Gene Expression in Umbilical Cord Blood and Peripheral Blood Mononuclear Cells from Healthy Children

Authors:

T. Oliver, T. L. Gonzalez, C. A. Jefferies, M. D. Pisarska; Cedars-Sinai Med. Ctr., Los Angeles, CA

Abstract:

Background: Early-life immune responses are crucial for long-term health, particularly for chronic inflammatory conditions like asthma and autoimmune disease. This study examines the transcriptional profiles of umbilical cord blood mononuclear cells (CBMCs) and peripheral blood mononuclear cells (PBMCs) from 5-year-olds in the Childhood Asthma Study (CAS). We aim to understand developmental and sex-specific differences in immune regulation. This analysis uses data from the NCBI GEO database (GSE184383).
Methods: Samples were obtained from 50 CBMCs and 27 PBMCs, divided into four control cohorts based on sex: Male Cord (n=22), Female Cord (n=16), Male PBMC (n=10), and Female PBMC (n=10). Samples treated for asthma or wheezing at age 5 were excluded. Outliers were also removed. Differential expression analysis was performed using DESeq2, with UMAP visualization for clustering and pathway enrichment analyses conducted using Gene Ontology.
Results: Significant differentially expressed genes (DEGs) at FDR<0.05 were observed between cord blood and PBMCs, and between sexes. Overall, there were 2465 DEGs between birth and childhood, 2490 sex-dependent DEGs between male and female cord blood, and 651 sex-dependent DEGs between male and female PBMCs from childhood. The immunoglobulin heavy-chain locus (*IGH*) showed notable differences, with genes associated with mucosal immunity (IgA) upregulated in PBMCs. X-linked genes showed greater variability in cord blood compared to PBMCs, with notable genes including *XIST* and *KDM5C*. Pathway enrichment analysis highlighted cytokine-cytokine receptor interaction and Jak-STAT signaling as differentially regulated. Genes related to T cell receptor signaling and natural killer cell-mediated cytotoxicity were enriched in PBMCs, indicating immune maturation. Female cord blood had higher expression of genes involved in miRNA catabolic activity and protein demethylation (*KDM6B*), while male cord blood exhibited higher expression of genes related to chronic inflammatory responses (*IL36G*) and regulation of lymphocyte apoptosis (*BCL2*).
Conclusion: This research provides critical insights into transcriptional changes in immune cells

from birth to early childhood, emphasizing the importance of age and sex in immune development. The findings have implications for understanding early-life immune development and its impact on long-term health outcomes, potentially guiding prenatal and early childhood care strategies.

Funding: Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (R01HD091773).

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7027T: Deregulated MicroRNAs (miRNAs) as Promising Biomarkers for Monitoring Preeclampsia (PE) in Pregnancies

Authors:

A. Argi[1], S. Shah[2], P. Usha[1], R. Rawal[3], U. Radhakrishna[4]; [1]Andhra Med. Coll., Visakhapatnam, India, [2]Smt NHL Municipal Med. Coll., Ahmedabad, India, [3]Gujarat Univ., Ahmedabad, India, [4]Green Cross Pathology Lab., Paldi, Ahmedabad, India

Abstract:

Background: Preeclampsia (PreE), characterized by high blood pressure and organ damage, is a major cause of maternal and fetal deaths worldwide Furthermore, it can affect the liver, kidneys, brain, and clotting system in both the mother and the infant. The estimated prevalence of preeclampsia is 2-10% of pregnancies globally, but in India, it affects 13-15% of pregnant women. Recent research indicates that microRNAs (miRNAs) are pivotal in the development of various developmental disorders, including their potential involvement in pregnancy-related complications like preeclampsia. This study aims to investigate the role of microRNAs (miRNAs) in preeclampsia and explore their potential use as biomarkers for this pregnancy-related complication. **Methods.** Genome-wide DNA methylation analysis was conducted on 20 women with severe preeclampsia and an equal number of normotensive control pregnancies using the Illumina Epic Array platform. A comprehensive analysis, integrating statistical and bioinformatic methods, was then performed to identify differentially methylated microRNAs associated with preeclampsia. **Results.** A total of 525 out of 2410 interrogated microRNA CpG sites showed significantly different methylation, with 550 of these sites associated with miRNA genes linked to preE. Among these, 45 miRNA CpG sites showed significant methylation in preE and had been previously reported and presently identified to be associated with the condition. These include miRNAs such as miR-675, miR-210, miR-378a, miR-411, miR-744, miR-519E, miR-525, miR-100, miR-328, miR-127, miR-379, miR-377, miR-195, miR-376c, and miR-378A.These altered miRNAs are expected to target several coding genes, potentially contributing to the development of

preEby regulating various cellular processes such as trophoblast proliferation, migration, invasion, apoptosis, differentiation, metabolism, and angiogenesis. **Conclusion:** The study found significant methylation changes in specific miRNAs in pregnant women who developed preE. This suggests these miRNAs are possible promising biomarkers that may help to predict preE in the future. These findings open a pathway for personalized treatment, but additional studies are crucial to confirm and explore their utility in diagnosis and therapy.Understanding their target genes can offer insights into the molecular mechanisms underlying preE.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7028T: Disruption and Amplification of *PRKN* gene (505 kb) in a case of fetal demise with a normal karyotype

Authors:

S. Ali[1], P. Marinescu[2], L. Glick[1], M. Maali[1], T. Knebel[1], A. Cox[1], B. Zhang[1], M. Iqbal[1]; [1]Univ. of Rochester Med. Ctr., West Henrietta, NY, [2]Univ. of Rochester, Rochester, NY

Abstract:

Background: Gene amplifications (> 4 copies of gene) in humans have widely been studied in somatic alterations such as cancers, however the incidence in germline cells is a rare event. *PRKN* gene encodes ubiquitin-E3- ligase, and is a transcriptional repressor of p53. Multiple cases with birth defects including neurodevelopmental disorders, facial dysmorphism, and skeletal anomalies have been reported with the copy number variations involving *PRKN* gene. To the best of our knowledge, the case we present is the first known of *PRKN* gene amplification with ~7 copies observed in amniotic fluid in the setting of fetal demise. **Case Report:** A 31-year-old gravida 3, para 2 presented at 21w0dGA with a complaint of decreased fetal movement without any bleeding, cramping, fever, or chills; ultrasound assessment demonstrated a single fetus without cardiac activity and measuring 16 weeks gestation. A prior ultrasound at 16w4d GA weeks gestational age demonstrated an anterior placenta previa and choroid plexus cysts; no other fetal structural or growth abnormalities were identified Amniocentesis was performed for cytogenetic workup after the demise was confirmed. Of note, the patient's first child was born with a single multicystic dysplastic kidney. The patient, herself, also has a history of heterozygous Factor V Leiden mutation which was diagnosed the year after her 2nd pregnancy; she has no personal history of clots. **Results:** Chromosomal microarray analysis (CMA) on Agilent 4x180k CGH+SNP array revealed 505 kilobase amplification ~7 copies at 6q26, arr[GRCh37/hg19] 6q26(162584300_163089403)x~7, causing *PRKN* gene disruption between exon 3 and 4. Follow

up targeted metaphase BAC FISH (Empire Genomics, RP11-148P13) confirmed amplification as tandem on one homolog. Parental CMA is in progress to determine the origin and nature of the amplification. **Conclusion:** 1. The *PRKN* gene is located in FRA6E, the 3rd most frequently observed common fragile site in the human genome and possesses the mutation hot spot of the *PRKN* gene between exons 3 to 8. Amplification in our case corresponds to the breakpoint within this mutation hot spot and may have resulted in the differential expression. 2. Additional cases with *PRKN* gene amplification in fetal demise or fetal abnormalities will shed more light on the role of *PRKN* in fetal development. 3. Prenatal CMA is essential for the detection of copy number variants in cases with spontaneous abortions and should be utilized as a Tier-1 test.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7029T: Enhancing Fetal Fraction in Non-Invasive Prenatal Screening (NIPS) Through Differential Methylation Status ⭐

Authors:

S. Daryabari, S. Giroux, A. caron, F. Rousseau; Laval Univ., Quebec City, QC, Canada

Abstract:

Non-invasive prenatal screening (NIPS) for trisomy 21, 18, and 13 relies on cell-free DNA (cfDNA) sequencing but faces challenges due to low fetal fraction (<4%), leading to screen failures or no calls in 0.5-3% of samples, affecting maternal mental health and increasing costs. Fetal-derived cfDNA fragments are typically shorter, differentially methylated, and enriched at specific genomic locations compared to maternal cfDNA. Adequate fetal fraction is crucial for high sensitivity and specificity. Differences in methylation status between fetal and maternal DNA suggest that using methylation-sensitive restriction enzymes could enhance fetal fraction by selectively digesting maternal DNA. **Method:** Peripheral blood samples from 84 pregnant women (40 normal, 44 trisomy cases) were analyzed, including 73 male and 11 female fetuses. Cell-free DNA was extracted from 4 mL of maternal plasma, and two libraries were prepared per sample: one before digestion and one after digestion with MspJI and FspEI enzymes. Libraries underwent size selection, targeting DNA fragments of 229 to 296 bp, and were sequenced in four groups: 1) before digestion and size selection, 2) after digestion but before size selection, 3) before digestion but after size selection, and 4) after digestion and size selection. Y ratio in male samples and chromosomal ratio changes in trisomic samples were analyzed using paired t-test and one-way ANOVA. **Result:** Group 1 (no digestion/size selection) showed the smallest change in chromosomal ratios compared to euploid samples. Group 4 (digestion and size selection) exhibited the most significant changes, with mean deviations of 5.1%, 6.6%, 11.3%, and 13.3%

across the groups. One-way ANOVA revealed significant differences between Group 1 and Groups 2, 3, and 4, as well as between Groups 3 and 4. In 73 male fetal samples, the Y chromosome ratio mirrored chromosomal ratio trends, with Group 1 having the lowest Y ratio (mean 4.1) and Group 4 the highest (mean 9.0). **Conclusion:** Methylated-CpG digestion preferentially cuts maternal over fetal DNA. Digestion followed by size selection increases the fetal fraction, as indicated by higher Y and chromosomal ratios in trisomic samples. This combined approach enhances fetal fraction in NIPS, improving test accuracy and potentially reducing failures. Further development of methylation-sensitive enzymes could enhance this method's efficacy.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7030T: Exome-Based Genetic Screening for Preconception in the Mexican Jewish Community

Authors:

T. Wegman-Ostrosky[1,2,3], S. Harari Arakindji[4], T. Metta-Harari[5], L. Mariscal[2,3], P. Mendelrsberg[6], S. Gershon-Naamat[7], R. Birnbaum[8], H. Mor-shaked[9,10]; [1]Lab of Precision Med., Inst. Natl. de Cancerologia, Cdmx, Mexico, [2]Centro Med. ABC, Mexico City, Mexico, [3]ZUG, Mexico City, Mexico, [4]Univ. Anáhuac México Norte, Mexico City, Mexico, [5]Univ. Anahuac, Mexico City, Mexico, [6]ZUG, Cdmx, Mexico, [7]Dept. of Genetics, Hadassah Med. Ctr., Jerushalem, Israel, [8]Dept. of Genetics, Hadassah Med. Ctr., Jerusalem, Israel, [9]Hebrew Univ. Hadassah Med. Sch., Jerusalem, Israel, [10]Dept. of Genetics, Hadassah Med, Jerushalem, Israel

Abstract:

The Mexican Jewish Community is a unique community with a population of approximately 60,000 individuals, established by Jewish immigrants from the Middle East. The community's preference towards endogamous marriages increases the risk of autosomal recessive (AR) diseases. Exome sequencing (ES) for preconception carrier screening is a comprehensive method to identify couples with increased risk of offspring with AR genetic disorders. Following informed consent, ES was performed. Only Pathogenic/likely pathogenic (P/LP) variants in AR disease-causing genes were reported. X-linked variants of P/LP classification, and ACMG secondary findings were provided upon request. Severity of the variants was determined based on the classification suggested by Lazarin et al., into one of four groups: profound, severe, moderate, or mild. 547 couples participated between the years 2020-2023. 14 (2.6%) and 5 (0.9%) couples were found to be shared carriers of pathogenic variants in AR disease causing genes of moderate-severe, and mild severity, respectively. Notably, in 8/14 (57%) of these

couples, both partners carried the same variant. One couple shared a novel variant. In addition, ACMG secondary findings were identified in 3.29% (36) of the individuals, with the most common variant being a well-established Ashkenazi founder mutation in the BRCA2 gene (BRCA2: c.5946del). 2.6% (14/547) of the couples were found to be carriers of pathogenic variants in AR moderate-severe disease-causing genes. The high incidence of shared variants among couples highlights the crucial role of preconception carrier screening in isolated populations. When comparing with routinely used carrier screening panels, 3 out of these 14 couples would not have been identified. This emphasizes the advantage of extracting such data from ES and shows that ES is an effective approach for identifying at-risk couples in isolated communities.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7031T: Fetal carriage of variants near *GCLC* are associated with altered levels of multiple placentally-derived proteins in maternal serum across gestation ⭐

Authors:

J. A. Mack[1,2], U. Sovio[1], F. Day[1], F. Gaccioli[1], E. Cook[1], N. Bayzid[3], M. Cotic[3], N. Dunton[3], G. Madhan[3], A. Motsinger-Reif[2], J. Perry[1], S. Charnock-Jones[1], G. C. Smith[1]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]NIH, Durham, NC, [3]Univ. Coll. London, London, United Kingdom

Abstract:

Placental function plays a key role in determining pregnancy outcomes. Proteins derived from the placenta, such as placental growth factor (PlGF), beta-human chorionic gonadotropin (hCGBeta), and pregnancy-associated plasma protein A (PAPP-A), regulate maternal homeostasis, endothelial function, and fetal growth. Placental dysfunction is implicated in complications including fetal growth restriction and preeclampsia, and in cardiometabolic disorders for mother and offspring in later life. We investigated the genetic basis of differences in circulating serum levels of placental proteins across gestation in the Pregnancy Outcome and Prediction study, a prospective cohort of nulliparous women. Analyses included 3,685 maternal samples and 3,484 fetal samples that passed genotype quality control. Maternal serum placental protein levels (PlGF, hCGBeta, PAPP-A, alpha fetoprotein and soluble fms-like tyrosine kinase-1) were measured at around 12-, 20-, 28-, and 36 weeks of gestational age (wkGA) using the Roche Cobas e411. In linear regression modeling, the dependent variable was z-score of log-transformed serum protein measurement at four timepoints, adjusted for exact wkGA, maternal weight, and sample storage time. Covariates included fetal sex, age, maternal race/ethnicity,

and the top 10 genetic principal components. We identified 1,500 fetal and 357 maternal genome-wide significant protein quantitative trait loci ($P < 5 \times 10^{-8}$). The strongest genetic variant, rs2013039, near the luteinizing hormone subunit beta (*LHB)* gene on chromosome 19, was associated with lower levels of hCGBeta at 28 wkGA in the fetal genome (Effect: -0.24, 95% CI: -0.29, -0.19 , $P = 4.41 \times 10^{-24}$). The same allele in the maternal genome had a weaker effect (Effect: -0.14, 95% CI: -0.18, -0.09 , $P = 9.71 \times 10^{-10}$). Fetal variants close to the glutamate-cysteine ligase catalytic subunit gene (*GCLC)* and the antisense RNA gene (*GCLC-AS1)* on chromosome 6 were associated with lower circulating levels of hCGBeta, PAPP-A, and PlGF across all gestational ages. This SNP set includes loci within an enhancer region downstream of *GCLC*. Associations between these variants and maternal serum protein levels were invariably stronger in the fetal than the maternal genome. We speculate that genetic determinants of placental protein levels could elucidate mechanisms underlying pregnancy-related complications, potentially aiding in improved diagnostics and therapeutic strategies. This study also demonstrates that genetic variants inherited from the father could alter maternal physiology through effects on maternal serum levels of placentally-released proteins.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7032T: Genes encoding inhibitory KIR contribute to pre-eclampsia risk in Uganda

Authors:

K. Kichula[1], N. Font Porterias[1], J. Jayaraman[2], L. Guethlein[3], N. Nemat-Gorgani[4], J. Traherne[2], P. Parham[3], A. Nakimuli[5], A. Moffett[2], P. Norman[1]; [1]Univ. of Colorado, Aurora, CO, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Stanford Univ Sch Med, Stanford, CA, [4]Stanford Univ., Stanford, Ca, CA, [5]Makerere Univ., Kampala, Uganda

Abstract:

Worldwide rates of hypertensive disorders of pregnancy (HDP) are increasing annually, being a significant cause of mortality especially affecting women of African ancestry. Indeed, pre-eclampsia (PE), which is an HDP that develops from insufficient placentation, is one of the main causes of mortality of mothers as well as infants in sub-Saharan Africa. Facilitating proper placentation, uterine natural killer (uNK) cells interact with invasive fetal trophoblast cells at the maternal-fetal interface during early gestation. The effector activities of uNK cells are modulated by killer-cell immunoglobulin like receptors (KIRs) that interact specifically with HLA-C expressed by the trophoblast cells. KIR vary by gene presence/absence, allelic polymorphism, and may be inhibitory or activating in function. Across cohorts of diverse ancestries, inhibitory

KIR have consistently been associated with greater risk of developing PE. Despite this compelling evidence, high-resolution analyses of the highly polymorphic *KIR* genes have been limited. In this study, we analyzed a large cohort of pregnant women from the Mulago Referral and Teaching Hospital in Kampala, Uganda (PE = 589, controls = 641), to characterize the contribution of *KIR* polymorphism to PE risk in this understudied region. We used our established high-throughput targeted sequencing methods to generate high-resolution *KIR* and *HLA* genotypes. We observed the inhibitory *KIRAA* genotype to be enriched in cases (p = 0.03, OR = 1.3 [1.02, 1.65]), consistent with previous studies. Inhibitory *KIR2DL3*00101* was identified as contributing to risk of PE (p_c = 0.036, OR = 1.3, [1.04, 1.63]), and the risk increased with copy number of this allele (1 copy, p = NS; Additive model, p = 0.016, OR = 1.2 [1.04, 1.44]; 2 copies, p = 0.005 (OR = 1.6 [1.14, 2.12]). Importantly, analysis of this highly heterozygous population permitted elimination of *KIR* alleles that would be seen in linkage disequilibrium with *KIR2DL3*001* in other populations. Additionally, we observed presence of genes expressing activating KIR (*KIR2DS2 and KIR2DS1)* to be associated with protection from PE, supporting the hypothesis that NK cell activation is necessary for proper placentation. This study demonstrates the need for large diverse cohorts to further our understanding of pregnancy progression and disorder.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7033T: Genetic burden of glycaemic dysregulation is associated with early-life cognitive development and neurology-related protein expression

Authors:

J. Huang[1], M. Z. Kee[1], J. Y. Huang[2]; [1]Agency for Sci., Technology and Res. (A*STAR), Singapore, Singapore, [2]Univ. of Hawai`i at Mānoa, Honolulu, HI

Abstract:

Insulin resistance and glucose metabolism have been associated with neurodevelopmental disorders. However, in the metabolically more susceptible Asian populations, it is not clear whether the genetic burden of glycaemic dysregulation influences early-life neurodevelopment. In a multi-ethnic Asian prospective cohort study in Singapore (Growing Up in Singapore Towards healthy Outcomes (GUSTO)), we constructed child and parental polygenic risk scores (PRS) for glycaemic dysregulation based on the largest genome-wide association studies of type 2 diabetes (T2D) and fasting glucose among Asians. We found that child PRS for HOMA-IR was associated with a lower perceptual reasoning score at 7 years (β=-0. 141, p-value=0.024, 95%CI -

0. 264 to -0. 018) and a lower standardised academic achievement mean score (WIAT-III) at 9 years (β=-0.222, p-value=0.001, 95%CI -0.357 to -0.087). These associations were consistent in direction among boys and girls, and were not influenced by parental PRS. In addition to these results that have been published in Translational Psychiatry (Jan 2024, PMID: 38177108), we further constructed cluster-specific PRS by accounting for T2D genetic clusters and found that PRS related to beta-cell function, insulin synthesis, and obesity-mediated insulin resistance were inversely associated with perceptual reasoning score and WIAT-III mean score (P<0.005). A higher PRS related to beta-cell function was also positively associated with the expression levels of several neurology-related proteins at age 8 years, including Ephrin A4 (EFNA4, β=0.084, P= $3.5x10^{-5}$, 95%CI 0.045 to 0.124) and brain-specific chordin-like protein (Brorin, also known as VWC2, β=0.104, P=0.001, 95%CI 0.042 to 0.165). In addition, EFNA4 and VWC2 were inversely associated with WIAT-III mean score (EFNA4: β=-0.908, p-value=0.001, 95%CI -1.438 to -0.378; VWC2: β=-0.448, p-value=0.007, 95%CI -0.770 to -0.126). However, mediation analysis suggested an interaction between higher PRS related to beta-cell function and expression of EFNA4, rather than a mediating effect via EFNA4 expression. In summary, our findings suggest that higher child PRS for HOMA-IR was associated with lower cognitive scores. Analyses of cluster-specific PRS suggested genetic determinants of beta-cell function may influence cognitive development, potentially via mechanisms related to glucose homeostasis and insulin processing.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7034T: Genetic Link Between Androgen Receptor Repeats and Sex Hormones Levels in Premature Ejaculation.

Authors:

S. Abbas, S. Bhatti, H. L. Khan; Lahore Inst. of Fertility and Endocrinology, Hameed Latif Hosp., Lahore, Pakistan

Abstract:

Despite its high global prevalence, the characterization, epidemiology, and management of premature ejaculation (PE) remain elusive and challenging. Premature ejaculatory dysfunction in males is recognized as an increasing issue, manifesting through spontaneous ejaculation across various populations. This study aimed to elucidate the relationships between trinucleotide repeats in the androgen receptor (AR) gene, sex steroids, and pituitary hormones with sexual function in men with type 2 diabetes mellitus (DM) who report acquired PE. The study enrolled 250 control subjects and 550 subjects with both PE and DM. The Premature

Ejaculation Diagnostic Tool (PEDT) and Beck Depression Inventory-II (BDI-II) were utilized for these assessments. Pearson's correlation analysis was employed to examine the relationships between clinical, hormonal, and genetic variables. Dependence between genetic, clinical, and demographic parameters was evaluated using Ward's minimum variance cluster analysis and principal component analysis. The study found that patients with the lowest number of (≤21) (CAG)n repeats exhibited higher serum oxytocin levels (114.2 pg/ml; n = 54, 43.2%) compared to controls (69.18 pg/ml; n = 22, 17.6%) and patients with the highest number (≥26) of (CAG)n repeats (62.9 pg/ml; n = 108, 43.2%). Conversely, patients with the highest number of (≥26) (CAG)n repeats demonstrated elevated serum testosterone levels (6.1 ng/ml; n = 108, 43.2%) and lower prolactin levels (3.01 ng/ml; n = 108, 43.2%) compared to controls and patients with the lowest (≤21) (CAG)n repeats. Additionally, their TSH levels (1.53 mIU/L, $P < 0.05$) were significantly lower than those of the controls. In the Pearson correlation model, self-estimated IELT exhibited a significantly negative correlation with both (CAG)n and (GCC)n repeats ($r = -0.16$, $p = 0.0001$; $r = -0.19$, $p = 0.0001$, respectively). These repeats also showed a positive correlation with PEDT scores ($r = 0.28$, $p = 0.0001$; $r = 0.24$, $p = 0.0001$, respectively) and an inverse correlation with BDI-II scores ($r = -0.25$, $p = 0.0001$). The findings of this study indicate that polymorphisms in the androgen receptor gene significantly influence the endocrine regulation of the ejaculatory reflex, contingent upon various cofactors. Our results underscore a strong association between extended trinucleotide repeats in the androgen receptor and the interplay of sex steroids, pituitary hormones, and thyroid hormones in the context of acquired PE in diabetic patients. However, the endocrine regulation of PE is a multifaceted phenomenon that warrants further investigation to fully elucidate its complexities.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7035T: Genome-wide association study implicates genes related to development of cardiovascular system and immunology in predisposition to preeclampsia

Authors:

J. Tyrmi[1], A. Lokki[2], T. Jääskeläinen[3], T. Laisk[4], K. Reis[5], P. Palta[6], Estonian Biobank Research Team, FinnGen, H. Laivuori[7]; [1]Univ. of Oulu and Tampere Univ., Oulu and Tampere, Finland, [2]Dept. of Bacteriology and Immunology and Translational Immunology Res. Program, Helsinki, Finland, [3]Med. and Clinical Genetics, Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland, [4]Inst. of Genomics, Univ. of Tartu, Estonia, Estonia, [5]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [6]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [7]Tampere Univ. Hosp. and Tampere Univ., Tampere, Tampere, Finland

Abstract:

Introduction: Pre-eclampsia (PE) is a vascular pregnancy disorder that affects 3-5% of all pregnancies. Genetic contribution to PE susceptibility is well established, but understanding of the actual risk loci remains incomplete. The previous genome-wide association studies (GWAS) so far have identified loci mostly connected to hypertension. To make further discoveries of the underlying genetic architecture, we performed a new GWAS for maternal PE. **Methods:** We combined the genetic resources of Finnish FinnGen project, Estonian Biobank and UK Biobank.The controls for each phenotype comprised all parous women not diagnosed with PE. GWAS analyses were first performed separately in each cohort, and the results were then meta-analysed. In total the meta-analysis included 13,626 PE cases. **Results:** We detected 16 loci, of which four have not been previously associated with preeclampsia but harbour genes with putatively important role in maintaining a healthy pregnancy. One of the loci is located near PLEK, a gene whose dysfunction has earlier been suggested to contribute to PE via immunological pathway in *in silico* modeling. Another interesting gene, SSR3, has been shown to be required for proper development and vascularization of placenta in mice. Two other loci harbor genes previously associated with hypertensive pregnancies. **Discussion:** The associations discovered in this study close to genes affecting blood pressure support the idea of pregnancy as a window to future cardiovascular health: the increased genetic susceptibility to cardiovascular disease may become evident for the first time during pregnancy. The novel association of PLEK and SSR3 shed more light into molecular mechanisms contributing to PE.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7036T: Human placental gene expression at different collection sites

Authors:

T. Gonzalez, A. M. Schaub, A. G. Novoa, A. Swarna, E. Hernandez, R. Hussaini, B. J. Shabani, A. E. Flowers, J. Williams, III, M. D. Pisarska; Cedars-Sinai Med. Ctr., Los Angeles, CA

Abstract:

Placenta develops during pregnancy and mediates communication between the mother and fetus, acting as a mediator for nutrient and gas exchange, as well as a regulator of fetal growth and maternal immune tolerance. Placenta morphology varies across species and within species, including diversity of shape, organ thickness, tissue density, and cell distribution. Furthermore, placenta morphology varies near and distal from the umbilical cord. Placental sampling location has been suggested to differ based on distance from the umbilical cord, but gene expression

changes in human placenta have not been examined. Differences by collection site have implications for placental experimental design, interpretation of biomedical research, and ability to compare results between studies. Term placental biopsies on the fetal surface from healthy singleton pregnancies were collected within one inch of the umbilical cord insertion ("inner placenta") and two inches from the outermost edge ("outer placenta"). We performed bulk messenger RNA-sequencing, principal components analysis, and differential gene expression analysis (16 inner versus 18 outer placenta) adjusted for fetal sex. We identified 22,457 genes expressed at term, with highest expression from *chorionic somatomammotropin hormone 2* and *1* (*CSH2* and *CSH1*). We found 198 genes with nominally significant differences ($P<0.05$) but none remain significant after adjustment for multiple comparisons. In conclusion, there do not appear to be transcriptomic differences in human placenta collected at variable proximity to the cord insertion site, providing more flexibility in sample collection and processing. However, variations between the maternal and fetal surfaces of placenta remain to be determined and future studies are needed to understand these differences. Research reported in this abstract was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number R01HD091773. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7037T: Incidental detection of maternal cancer in prenatal cell-free DNA screening: Clinical laboratory experience and qualitative approach to Multiple Structural Variants ⭐

Authors:

T. Boomer, S. Hancock, F. Liao, K. Brown, I. Singale, S. Comer; Illumina, Inc., Foster City, CA

Abstract:

Prenatal cell-free DNA screening (cfDNA) via whole genome sequencing (WGS) can identify Multiple Structural Variants (MSV); characterized by multiple regions of genomic gain or loss across chromosomes. Discovered as an incidental finding, MSV potentially represents a significant risk for maternal cancer. Currently, understanding of MSV on prenatal cfDNA is limited and standardized clinical management protocol is lacking. In this study, we present a cohort of MSV cases, a subcohort with confirmed maternal cancer, and an analysis of the chromosomal aberrations identified.
Patient blood samples from over 400,000 pregnancies were submitted to the CLIA-certified

Illumina Laboratory and processed via Verifi™ and Verifi™ Plus Prenatal Screen. Samples with MSVs were routed for expert review. Test reports indicated cancellation due to MSV and the associated risk for maternal neoplasia. Clients were also informed of a relevant NIH clinical study (IDENTIFY). Reported samples with this cancellation comment were selected for study and clinical outcomes were requested. Normalized chromosome values (NCV), comparable to Z score statistics, were analyzed, plotted, and categorized via R statistical Software (v4.2.2). Previously described criteria for chromosomal aberrations seen in malignancy suspicious-cfDNA samples were applied (Heesterbeek et al, 2022) with anomalies characterized by NCV into mild (NCV≥3, <-3) and strong (NCV≥10, <-10) aberrations.
A total of 26 samples (0.006%) met criteria. Clinical outcomes were received for 9 cases, and maternal cancer was reported in 5 (56%). The median maternal age of the overall cohort is 35.6 years and median gestational age (GA) is 13 weeks. The cancer subcohort had a median age of 32.0 years and similar GA. In the MSV cohort, an average of 6 strong aberrations and 7.8 mild aberrations were noted per case, while the cancer subcohort noted an average of 5 strong aberrations and 11.2 mild aberrations per case. Duplications in chromosomes 2 and 8 and depletions in chromosomes 4, 6, and 14 were most common in the MSV cohort, while duplications in chromosomes 2 and 3 and depletions in chromosomes 6 and 18 were most common in the cancer cohort. The majority of cases (23) presented with complex CNV/monosomy, with 3 cases presenting with multiple trisomies.
These analyses align with published data that there is a significant overrepresentation of maternal malignancy in cases with MSV, especially those harboring multiple, strong, and complex chromosomal changes. Such detailed data may eventually lead to focused follow-up recommendations for patients based on the characteristics of MSV findings in prenatal cfDNA screening.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7038T: Investigation of clinical and genetic heterogeneity in human fetuses with arthrogryposis multiplex congenita ⭐

Authors:

S. Ms; Kasturba Med. Coll., Manipal, Manipal, India

Abstract:

Background/Objectives: Arthrogryposis is a condition with contractures in two or more joints associated with approximately 400 conditions. The study aims to elucidate the clinical and genomic profile of human fetuses with arthrogryposis Methods: We performed perinatal

evaluation of fetuses referred for autopsy following pregnancy loss and clinical evaluation in neonates after taking informed consent. Fetal tissues and blood samples were used to obtain genomic DNA. Chromosomal microarray (CMA), whole exome, and genome sequencing were performed as required. Spinal muscular atrophy (SMA) and common aneuploidies were ruled out whenever necessary. Results: We have recruited 29 families (35 subjects; males -20 ; females- 15) with arthrogryposis in the study. Fourteen consanguineous families (48%) were noted. Gestation ranged from 14-32 weeks and there were three neonatal losses (3-28 days). The study identified rare phenotypes associated with *ADGRG6*, *ANKS6*, *AGRN, CHRNA1*, *CEP55, GUSB*, *COASY, CEP290*, *HS2ST1 and ASCC1.* Two families were diagnosed with extra copy of chromosome 18 (trisomy 18). Gene of uncertain significance, *KIF21A* and *UNC50* were identified. Biallelic loss of function variants in *KIF21A* caused severe arthrogryposis and brain anomalies in a family. A homozygous variant c.644-13_644-9delCTTTC in *UNC50 lead to* aberrant splicing with an insertion of 17 nucleotides following reverse trancriptase-polymerase chain reaction and identified as cause of recurrent fetal akinesia deformation sequence in a family. Conclusion: We achieved a molecular diagnosis in fourteen families. Most of the families had autosomal recessive (10 families) inheritance and were lethal. Detailed perinatal evaluation with integration of genomic tests in fetuses with arthrogryposis improves diagnosis.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7039T: Phenotypic characterization of Global Developmental Delay and Intellectual Disability of genetic etiology.

Authors:

K. Singh, A. S. Parameswaran, L. Saini, T. Gupta; All India Inst. of Med. Sci. Jodhpur India, Jodhpur, India

Abstract:

Background: Global developmental delay (GDD) and Intellectual Disability (ID) of genetic origin among children with unexplained neurodevelopmental disorders constitutes a diagnostic dilemma to clinical geneticist, compounded by failure to exclude common acquired disorders like perinatal insult & metabolic disorders. **Objectives**: The study aimed to investigate the clinical presentations and genotype among patients with unexplained neurodevelopmental disorders and establishing correlation of genetics with clinical patterns. **Materials and Methods**: This prospective cross-sectional observational study included 78 children (1-18 years) diagnosed with GDD/ID attending the clinical genetics OPD at our centre after ruling out the commonly

known conditions that causes intellectual disability. All children were subjected to a comprehensive clinical evaluation along with IQ/SQ/DQ, neuroradiological assessment and genetic testing. **Results**: The mean age of presentation being 5.2 years, with male to female ratio of 2:1, 38% of the children were diagnosed with chromosomal disorders, 33% reported to have disease causing single gene variants. Mean IQ (50.84±6.23) and DQ (62.55±7.55) were significantly lower in children diagnosed to have chromosomal disorders. Missense mutations were found in 29.3% children, deletion in 22.7%, and 8% children were having duplication. There were 17.3% children with frameshift mutation and 12% had nonsense mutation. Simultaneous Deletion/Duplication were found in 9.3%. Although mis-sense mutations formed the single major type of genetic alterations, only 22.7% (5/22) were established to have phenotypic effects. Among children with chromosomal disorders, 62.7 % had copy number variants larger than 5Mb. *De-novo* mutations were found in 40% of the study participants. Coarse facies and microcephaly were more commonly found in children with pathogenic CNV as compared to disease causing SNV (p value < 0.05). The whole exome sequencing was able to pick up even copy number variants with detection rate of 71%. **Conclusion**: Chromosomal disorders had more likely to have coarse facies, microcephaly, and behavioral disorders. The CNV larger than 5Mb suggests importance of karyotype in settings of financial constraints. Whole exome sequencing is the best genetic test in a child with unexplained development disorder.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7040T: Placental dysfunction is associated with rare damaging variants in the stem cell differentiation pathway

Authors:

H. Chavez, N. Guo, M. Huang, C. Barba, E. Zhang, C. Mason, K. Nelson, M. Meads, O. Aisagbonhi, M. Parast, M. Horii, K. Fisch; Univ. of California, San Diego, La Jolla, CA

Abstract:

Placental dysfunction occurs in ~8% of all pregnancies, defined by the placenta failing to meet the nutrient and oxygen demands of the fetus for normal development and growth, and can lead to preeclampsia (PE). PE is a hypertensive disorder of pregnancy that affects 3-8% of pregnancies and is responsible for 10-15% of maternal mortality. Widespread placental mosaicism has been observed in healthy placentas but has not yet been investigated in placental dysfunction. We investigated rare SNVs and INDELs in RNAseq data from N=112 human placentas from normotensive (N=53) and preeclamptic (N=59) pregnancies with detailed

clinical adjudication and histopathological review using the GATK Best Practices Workflow for RNAseq short variant discovery implemented in bcbio-nextgen. We implemented stringent variant quality filters and prioritized rare variants resulting in 2,383 genes harboring rare variants and 949 genes harboring rare damaging variants. We calculated expressed mutational burden (eMB) as the number of expressed rare variants per capture region (50X coverage) per 10^6. Gene set enrichment analysis and a Fisher's Exact test was performed to identify gene sets that were enriched in specific clinical phenotypes. Enrichment analysis of rare damaging variants revealed that significantly dysregulated pathways (adj. $p < 0.05$) by gene expression were also enriched for rare damaging variants (adj. $p < 0.05$) and included programmed cell death, apoptotic process, innate immune system, negative regulation of stem cell differentiation and response to oxidative stress. eMB was higher in placentas from normotensive pregnancies with a gestational age at delivery <34 weeks compared to those delivering at >=34 weeks (Mann-Whitney U $p = 0.018$), but did not significantly differ between PE and normal ($p > 0.05$). However, in placentas with observed placental histopathology, the eMB of normotensive pregnancies with a gestational age at delivery <34 weeks was significantly higher than both PE ($p=0.01$) and normotensive ($p = 9.3e-5$), delivering >=34 weeks. In addition, placentas with rare mutations in the stem cell differentiation pathway had a significantly increased risk of having placental pathology (OR = 3.84, $p = 0.019$). To functionally evaluate these findings, we are leveraging an established iPSC-TSC model to recapitulate the mutational burden and patterns observed in PE by analyzing RNAseq data from 6 patient-derived UC-MSCs and iPSC-TSC (N=3 PE; N=3 Normal) across three replicates. Functional evaluation leveraging these models is ongoing. The results of this study will lead to an improved understanding of the genetic underpinnings of placental dysfunction.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7041T: Quantifying the proportion of women at risk of an FNAIT pregnancy in diverse populations

Authors:

A. Hayward[1], R. Armstrong[1], J. Cook[2], J. Card-Gowers[2], T. Coker[2]; [1]Rallybio, New Haven, CT, [2]Hlth.Lumen, London, United Kingdom

Abstract:

Fetal and neonatal alloimmune thrombocytopenia (FNAIT) is a rare and potentially serious condition of pregnancy, characterized by destruction of fetal platelets by maternal alloantibodies. The clinical consequences can be severe, including intracranial hemorrhage that

may lead to death, or severe neurological deficits in the baby. Maternal alloantibodies to HPA-1a account for ~75%-80% of FNAIT cases. HPA-1a negative women (rs5918 CC homozygotes) bearing an HPA-1a positive fetus are at risk of FNAIT, and those who also carry the HLA-DRB3*01:01 allele are at higher risk (~25x more likely to alloimmunize). While risk for FNAIT is well characterized in Caucasians, data in non-Caucasian populations are limited. Allele frequencies were obtained from gnomAD v4 for HPA-1a and the US National Marrow Donor Registry (NMDR) for HLA DRB3*01:01. HPA-1a negative rates and HLA-DRB3*01:01 carrier frequency were calculated from the population-specific allele frequencies assuming Hardy-Weinberg equilibrium. The proportion of women at risk of alloimmunizing in each ancestry was taken as the HPA-1a negative proportion, and the proportion of women at higher risk was obtained by multiplying this by the population-specific HLA-DRB3*01:01 carrier frequency. Proportions of women at risk of alloimmunizing were highest in Caucasian populations, with the highest proportions in the Ashkenazi Jewish population (2.36% and 0.65% of women at risk and at higher risk respectively), followed by non-Finnish Europeans (2.34% and 0.64%), Middle Eastern (2.25% and 0.62%), Amish (2.25% and 0.62%), White Hispanic (2.25% and 0.59%) and Finnish (2.03% and 0.56%). Additionally, women in some non-Caucasian population groups were at comparable risk with the highest proportions in the Caribbean Hispanic population (1.48% and 0.33%), followed by African / African American (1.13% and ~0.28%). Women of South Asian, East Asian and Amerindigenous ancestries were found to have lower risk. Our study is the first to report FNAIT risk in diverse ancestries. As FNAIT risk in some non-Caucasian populations is comparable to Caucasians, it demonstrates the need for awareness across populations and the need for both broad prenatal FNAIT risk screening and availability of treatment options for all pregnant women.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7042T: Reconstructing clonal dynamics of the human neural crest by genetic mosaicism

Authors:

K. Vong[1], X. Yang[2], C. Chung[1], J. Gleeson[1]; [1]Univ. of California San Diego, La Jolla, CA, [2]Univ. of Utah, Salt Lake City, UT

Abstract:

The neural crest generates a broad range of cell types, but conflicting results cast controversies on in vivo multipotency and developmental trajectories, particularly in human neurodevelopment. Here using somatic mosaic variants as inherent genetic lineage barcodes,

we perform tracing of the human trunk neural crest. From three independent adult neurotypical human donors, we isolate 95 sensory ganglia and 97 sympathetic ganglia, representing two major trunk neural crest derivatives. Using deep whole genome sequencing, we identify 1,278 mosaic variants and deconvolved their clonal, anatomical and lineage relationship. In contrast to current models in which multipotent neural crest progenitors migrate exclusively dorsoventrally to colonize both the sensory and sympathetic ganglia, we observe that mosaic variants are instead predominantly shared along the rostrocaudal axis, but rarely among different ganglia within the same level. Moreover, single cell multiomic analysis suggests both neurons and glia recapitulate this rostrocaudal clonal structure. Furthermore, unlike in the human neocortex where left-right axis is established prior to the rostrocaudal axis, we identify frequent ganglia-specific bilateral clones. Mathematical modeling suggests neural crest cell fate determination precedes left-right separation or migration from the neural tube, supporting an alternative model to the currently accepted multipotent neural crest theory. Thus, leveraging genetic mosaicism, our findings reveal blueprints of human neural crest development that could inform models of human disease.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7043T: Sex differences in placental gene expression and intraplacental variation in normotensive versus hypertensive pregnancies

Authors:

S. Plaisier[1], M. Smith[2], A. Abdulahad[3], T. Powell[3], L. Nordstrom[3], D. Coonrod[3], C. Roberts[2], M. Wilson[4]; [1]Arizona State Univ., Chandler, AZ, [2]Flinders Univ., Bedford Park, Australia, [3]Valleywise Hlth.Med. Ctr., Phoenix, AZ, [4]Arizona State Univ., Tempe, AZ

Abstract:

Preeclampsia is a pregnancy disorder characterized by hypertension and proteinuria and thought to be rooted in placental dysfunction. Many studies have reported fetal sex biases in the incidence of preeclampsia, but the impact of fetal sex on pregnancy complications is poorly understood. Having previously demonstrated sex differences in placental gene expression in uncomplicated pregnancies, we sought to determine if preeclampsia affects placental gene expression differently based on fetal sex. We performed sex-informed differential gene expression analysis of total RNA from placental tissue from full-term deliveries of ten hypertensive and ten normotensive pregnancies, each with half carrying a fetus assigned female at birth (46,XX karyotype) and half assigned male at birth (46,XY karyotype). For each placenta,

we sampled 4 evenly distributed locations across the placental disc to assay for intraplacental variation in gene expression. We observed high intraplacental variation in genes involved in extracellular matrix organization, hormone secretion, and the immune response. Accounting for multiple samples per placenta, we found a distinct profile of placental genes differentially expressed between normotensive and hypertensive pregnancies for each fetal sex. Differentially expressed genes from placenta in male-bearing pregnancies were functionally enriched for metabolism and cellular response to stress, while those from female-bearing pregnancies were enriched for genes involved in invasion, immune response, and angiogenesis. Placental gene expression in hypertensive pregnancies showed better correlation between published preeclampsia signatures after segregating by fetal sex. Known biomarkers of preeclampsia (e.g. LEP, FLT1, FSTL3) were only differentially expressed in placenta from male-bearing pregnancies. We found sex-specific differences in the imputed proportion of several cell types in hypertensive pregnancies, suggesting potential mechanisms by which sex differences in placenta biology could lead to different health outcomes. In conclusion, our study demonstrates the importance of sex as a biological variable, because there are large differences in the placental gene expression pattern of gestational hypertension based on the fetal sex assigned at birth which may be due in part to different proportions of specific placental cell types. Our study highlights the importance of increased sampling of the placenta when studying pregnancy complications as many genes' expression differs between male-bearing and female-bearing pregnancies and varies across the placental disc.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7044T: spatiotemporal transcriptomic map of non-human primates at early development from the dGTEx project

Authors:

A. Guillaumet[1], A. Russell[1], A. Sule[1], T. Coorens[1], C. Vanderburg[1], S. Silva de Castro[2], Q. Zhang[2], G. Feng[2], F. Chen[1], K. Ardlie[1], dGTEX Consortium; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]MIT, Cambridge, MA

Abstract:

Many human diseases are the result of defects in developmental processes that occur during gestation or in early life. To date, large gene expression and cell atlas studies have been primarily focused on adult tissues, and thus lack information on human development and critical cell states that only manifest during specific developmental windows. In parallel, little is known about the conservation of developmental programs across non-human primate (NHP)

species, with implications for human evolution, drug development and clinical testing. The developmental Genotype Tissue Expression (dGTEx) project aims to expand upon the adult GTEx dataset by profiling tissue-specific gene expression patterns throughout developmental stages in humans and two NHPs: rhesus macaques (Macaca mullata) and common marmosets (Callithrix jacchus). For the earliest developmental sampling time points for NHPs, which are prenatal, it is challenging to isolate individual tissues to interrogate expression. Here, we have applied Slide-tags, a novel, affordable and high resolution spatial capture method, on sections of whole bodies of marmosets up to gestation day 100 (G100). Slide-tags allowed us to simultaneously interrogate both the spatially resolved expression patterns and chromatin accessibility at a single nucleus resolution. This data allows us to identify cell types, expression states and epigenetic dynamics in early development, even in extremely small tissues, and serves as a basis for the analysis of gene regulatory changes across the lifespan. These efforts will establish a robust reference set of transcription and regulation across early ages and cross-species comparison, as well as building unique tissue repositories for future investigations.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7045T: Study of the mutational landscape of premature ovarian failure by whole exome sequencing in Bangladeshi women

Authors:

R. Mim[1], D. Pervin[2], M. Rahaman[1], T. Binte Eshaque[1], K. Uddin[1,3], A. Islam[1,4], M. Hasan[1,4], H. Akter[1,5], M. Uddin[6,4]; [1]Genetics and Genomics Med. Ctr., NeuroGen Hlth.care, Dhaka, Bangladesh, [2]Dept. of Obstetrics & Gynaecology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, Dhaka, Bangladesh, [3]Dept. of Biochemistry, Holy Family Red Crescent Med. Coll., Dhaka, Bangladesh, Dhaka, Bangladesh, [4]Cellular Intelligence Lab, GenomeArc Inc., Toronto, ON, Canada, [5]Dept. of Biochemistry and Molecular Biology, Univ. of Dhaka, Dhaka, Bangladesh, Dhaka, Bangladesh, [6]Mohammed Bin Rashid Univ. of Med., Dubai, United Arab Emirates

Abstract:

Background Premature Ovarian Failure (POF) is a condition affecting 1-5% of women worldwide, causing infertility and other health issues. It is diagnosed by menstrual disturbances lasting four or more months and elevated serum FSH levels (> 25 IU/l) at least 4 weeks apart. Genetic mutations, autoimmune diseases, and environmental factors are implicated, while the specific etiology of the condition is still unknown. About 14-31% of POF sufferers are shown to have at least one affected relative with POF, indicating a substantial hereditary propensity to this disease. However, there is limited information on the genetic basis of POF among women in

Bangladesh.
MethodsWe investigate a cohort of 30 clinically suspected cases of POF with 10 controls within Bangladeshi population. The cohort of women diagnosed with POF based on clinical criteria, hormonal assays, and ultrasonographic findings was recruited from the department of Obstetrics & Gynecology, Bangabandhu Sheikh Mujib Medical University (BSMMU), Dhaka, Bangladesh. We performed whole exome sequencing to look for disease causing variants and the pathogenic variants were further validated by sanger sequencing. We have used human genome build GRCH38/UCSC hg38 as reference. Variant classification analysis was conducted based on the American College of Medical Genetics (ACMG) guidelines.
Results Whole exome sequencing revealed 87% (26/30) clinically relevant variants (both pathogenic and variants of uncertain significance (VUS)), with 23.33% (7/30) of samples carrying pathogenic variants. We have identified and validated three novel likely pathogenic variants in this cohort, comprising of two frameshift deletions (compound heterozygous) [c.626delA(p.H209Pfs*11)] and [c.628_631del (p. S210Tfs*9)] in *TG* gene, and one missense variant [c.1904T>C (p.I635T)] in *TSHR* gene. The study revealed a significant genetic predisposition in the case study due to the presence of familial cases, where we got a pathogenic stopgain variant in exon 3 [c.463C>T (p.Q155X)] of *HROB* gene. All of the controls were negative for POF of the variants found in the case studies.
ConclusionsThis genetic study provides valuable insights into the etiology of POF in Bangladeshi women, revealing both known and novel genetic variants associated with the condition. These findings underscore the importance of genetic screening for early diagnosis and personalized treatment strategies.
Keywords: Premature Ovarian Failure (POF), Whole Exome Sequencing (WES), Sanger Sequencing (SS), Follicle Stimulating Hormone (FSH).

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7046T: Use of single cell genomic strategies to prioritize candidate genes in *Sox10*^{Dom} aganglionosis genetic modifier intervals

Authors:

J. Benthal, J. A. Avila, E. Southard-Smith; Vanderbilt Univ., Nashville, TN

Abstract:

Mammalian enteric nervous system (ENS) development requires migration of neural crest cells into the foregut and then along the developing intestine to the end of the gastrointestinal tract. During this process, enteric neural crest cells differentiate into enteric neurons and glial. *Sox10*,

a transcription factor that has been shown to interact with factors that influence chromatin accessibility, plays a key role in the development of the ENS, as exemplified by the *Sox10*^{Dom} mouse model of Hirschsprung's disease. *Sox10*^{Dom} mutants exhibit deficits of bowel colonization that results in aganglionosis mimicking human Hirschsprung disease (HSCR). The *Sox10*^{Dom} mouse model of HSCR also recapitulates the variable penetrance of the aganglionosis phenotype that is present in human patients. Previously our group used a standard F1 intercross strategy to map genetic modifiers of *Sox10*^{Dom} aganglionosis. To gain further resolution and potentially refine modifier intervals we sought to apply genome-wide linkage approaches in an extended pedigree of *Sox10*^{Dom} mice on a mixed C57BL/6J x C3HeBFeJLe-a/a F1 background. From this, we uncovered sex-specific and non-sex-specific associations which modify the aganglionosis phenotype of *Sox10*^{Dom} mice. We next sought to apply single cell multiomics tools (snATAC-seq, bulk and single cell RNA-seq, set of conserved Sox10 binding sites) to identify and prioritize the most relevant candidate genes within our genetic modifier intervals for aganglionosis. Using these data modalities, we found a set of genes including know genes important for ENS development such as *Phox2b* and novel genes that previously have not been attributed to ENS development. Several of these genes also contain intronic Sox10 transcription factor (TF) binding motifs, implicating potential direct Sox10 interactions. In addition, we find that E-box binding TF motif enrichment in male-specific modifier intervals within differentially accessible chromatin from snATAC-seq. Our approach identifies candidate genes and regulatory elements that are likely to impact the severity of the *Sox10*^{Dom} aganglionosis phenotype and points to novel putative mechanisms in which SOX10 interacts with genetic elements to produce phenotypes relevant for human disease.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Thursday Session

Board 7047T: Validation of fetal and maternal Spinal Muscular Atrophy (SMA) and Hemoglobin (Hb) Barts screening with FirstGene, a combined non-invasive prenatal cfDNA assay for fetal aneuploidy, recessive diseases, and serological screening

Authors:

R. Patel, J. Wang, K. Trettin, C. Battey, A. Srinivasan, J. Saikumar, D. Kushnoor, N. Xu, S. Pierson, K. Johansen Taber, G. Gould, D. Muzzey; Myriad Genetics, Inc., Salt Lake City, UT

Abstract:

Objectives: Spinal muscular atrophy (SMA) and Hemoglobin (Hb) Barts are serious genetic conditions often caused by *SMN1* copy number loss and *HBA1* and *HBA2* double-cis deletion, respectively. Guidelines support routine screening for both conditions, yet the necessity of screening both reproductive partners to identify the risk of an affected fetus is logistically challenging. The FirstGene assay addresses this challenge by predicting fetal recessive disease status without prior knowledge of maternal or paternal carrier screening results. However, determining fetal copy number for both the *SMN1* and the *HBA1* and *HBA2* gene regions is complicated by their high homology to other parts of the genome. We describe the analytical validation of the fetal and maternal SMA and Hb Barts screening components of FirstGene.
Methods: Genomic DNA (gDNA) mixtures from 5 trios with SMA carrier mothers and one trio with an Hb Barts carrier mother were included in this validation study. One proband from each set of trios was affected by the corresponding diseases. Differentially fragmented gDNA from the proband and the mother were mixed at 3%, 5%, 10%, 15% and 20% ratios to represent a range of fetal fractions (samples below 3% fetal fraction are routinely failed in FirstGene). These mixed samples were run on FirstGene and assessed for concordance with results of the corresponding single source gDNA samples run on a validated orthogonal assay. In addition, two simulation approaches were developed to estimate the analytical sensitivity and specificity for fetal calls at diverse fetal fraction levels representative of a general population. The first utilized observed plasma data to simulate affected and carrier fetuses. The second approach used subsampled data from *RHD*, a gene known to have many copy number changes, to mimic SMA and Hb Bart's data.
Results: FirstGene correctly identified 25/25 SMA carrier mothers, 4/5 SMA affected fetuses, 9/10 SMA carrier fetuses, 9/10 SMA non-carrier fetuses, 5/5 Hb Barts carrier mothers and 4/5 Hb Barts affected fetuses. The three SMA and one Hb Barts fetal samples that were not called positive were all no-called and had a 3% mixture ratio. Simulations demonstrated that the expected analytical sensitivity and specificity are >96% for fetal SMA status and >98% for fetal Hb Barts status.
Conclusions: FirstGene accurately identified maternal and fetal SMA and Hb Barts disease status. This screening approach streamlines the identification of at-risk pregnancies, especially in cases where carrier status is unknown for the reproductive partner.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4002T: (less un)Realistic Simulation of Complex Trait Architectures for Open Benchmarking of Statistical Genetic Estimators

Authors:

R. Border[1], S. Sankararaman[2], D. J. Benjamin[1], A. Young[3], A. W. Dahl[4], N. Zaitlen[2]; [1]Univ. of California Los Angeles, Los Angeles, CA, [2]UCLA, Los Angeles, CA, [3]Univ. of California, Los Angeles, Los Angeles, CA, [4]UChicago, Chicago, IL

Abstract:

Statistical genetic models are often simplistic compared to social and biological reality. There is growing recognition that many common assumptions made for the sake of tractability (e.g. additive effects, random mating, and/or independence of parent/offspring environment) are unlikely to apply to many phenotypes and that methods reliant on such assumptions can yield misleading results. However, investigations of the consequences of violating these assumptions have typically focused on a single perturbation in isolation (e.g. bivariate assortative mating with other assumptions maintained). Moreover, methodologists repeatedly create bespoke simulations exploring a small set of scenarios to evaluate each new proposed method, making it challenging to compare methods across plausible trait architectures in a comprehensive manner.

Given these challenges, we introduce xftsim, a forward-time simulation library capable of modeling a wide range of genetic architectures, mating regimes, and transmission dynamics, to facilitate the systematic comparison of existing approaches and the development of robust methods. Using xftsim, we probe the robustness of widely-used estimators (e.g. LD score regression, GWAS, colocalization) to an array of novel perturbations to standard assumptions, including high-dimensional assortative mating, multivariate parent-offspring transmission, and non-additive genetic effects. We find that widely-used polygenic estimators for unrelated individuals are unreliable across many scenarios and that these perturbations operate synergistically to distort conclusions. For example, high-dimensional assortment and multivariate transmission together more than quadruple GWAS false-positive rates. The severity of this bias increases with sample size, underscoring an emerging challenge in the era of multi-million-participant GWAS.

Finally, we introduce statgenbench, an open library of simulated multivariate phenotype / genotype datasets under a broad array of trait architectures and dynamics including high-dimensional assortative mating and parent-offspring effects, gene-gene and gene-environment interactions, spatial and ancestral structure, and admixture, in isolation and in combination.

statgenbench seeks to recreate the success of the standardized open benchmarks initiatives underlying the rapid pace of development in machine learning (e.g. ImageNet). Together, xftsim and statgenbench will enable researchers to compare the performance of existing methods and novel methods across shared datasets using well-defined performance metrics.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4003T: A Bayesian Approach to Correcting the Attenuation Bias of Regression Using Polygenic Risk Score

Authors:

G. Zhou[1], X. Qie[1], H. Zhao[2]; [1]Yale Univ., New Haven, CT, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract:

Polygenic risk score (PRS) has become increasingly popular for predicting the value of complex traits. In many settings, PRS is used as a covariate in regression analysis to study the association between different phenotypes. However, measurement error in PRS causes attenuation bias in the estimation of regression coefficients. In this paper, we employ a Bayesian approach to accounting for the measurement error of PRS and correcting the attenuation bias in linear and logistic regression. Through simulation, we show that our approach is able to obtain approximately unbiased estimation of coefficients and credible intervals with correct coverage probability. We also empirically compare our Bayesian measurement error model to the conventional regression model by analyzing real traits in the UK Biobank. The results demonstrate the effectiveness of our approach as it significantly reduces the error in coefficient estimates.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4004T: A community of networks approach for multi-omics integration.

Authors:

A. Leshchyk[1], S. Monti[2], P. Sebastiani[3]; [1]Boston Univ., Boston, MA, [2]Boston Univ. Sch. of Med., Boston, MA, [3]Tufts Med. Ctr., Boston, MA

Abstract:

Recent studies reveal that long-lived individuals experience a significant delay in age-related diseases such as Alzheimer's, dementia, and heart disease. Investigations into centenarians' genetics indicate that individuals carrying the APOE e2 allele are more likely to achieve longevity. This allele is associated with serum proteins and metabolites that provide insights into its influence on health outcomes. However, modeling the mechanisms linking genetic, molecular, and phenotypic data is challenging. We developed "Community of Networks", a method using Bayesian networks to integrate data from genetics, proteomics, metabolomics, and phenotypes. This method handles missing data through multiple imputation, creating a community of networks for downstream analyses. To reduce the dimension of multiple omics features, we summarize their effects into "omics-scores," providing a comprehensive summary of an individual's omics profiles. The method uses probabilistic reasoning with Bayesian networks to understand paths from genotypes to phenotypes. We applied this method to a multi-omics dataset from the New England Centenarian Study, including 362 centenarians, 672 centenarian offspring, and 435 controls. This approach identified paths connecting APOE alleles with molecular signatures of genetic variants and aging traits. For example, individuals with the APOE e2 allele have lower metabolomics risk scores for dementia compared to those with e4 and e3 alleles. Proteomic risk scores for dementia show a 40% change in the odds for a high score comparing e4 versus e2 carriers. Our research offers a framework for understanding the molecular underpinnings of longevity and age-related diseases, paving the way for targeted interventions to promote extended health span and quality of life in aging populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4005T: A computationally efficient pipeline to perform rare variant burden tests with imputed data in the 23andMe database

Authors:

A. Kwong[1], Z. Fuller[1], K-H. Lin[1], 23andMe Research Team, S. Pitts[2], A. Auton[1], X. Wang[1]; [1]23andMe, Sunnyvale, CA, [2]23andMe, South San Francisco, CA

Abstract:

A recent area of focus in human genetics has been on investigating associations between rare damaging mutations and disease with gene-based burden tests. Most studies have relied on whole exome sequence data to call variants with extremely low minor allele counts. We have previously demonstrated the use of imputed data in millions of samples from 23andMe research participants to perform rare variant burden tests. Here, we demonstrate a computationally efficient pipeline to carry out burden tests for over 1,100 self-reported

phenotypes. Burden tests were performed across various allele frequency thresholds and variant effect annotations, including the use of deleteriousness scores from protein language models, with correlated p-values summarized using combination tests. Our pipeline also conditions on independently associated variants to adjust for nearby common signals. Performing a total of more than 200M association tests, our pipeline can generate results for >7M samples for <$5k USD and <100k CPU hours. We show our rare variant burden associations replicate previous results from smaller whole exome sequencing datasets, particularly for common diseases. Additionally, we detect a number of known protective loss-of-function associations of therapeutic interest. Finally, we demonstrate how signals uncovered from rare variant burden testing can reveal causal genes for known Mendelian diseases, not available from single variant GWAS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4006T: A deep generative network for genotype-phenotype prediction using high-content zebrafish imaging

Authors:

H. Hu, R. Valdarrago, J. Uribe-Salazar, M. Dennis, G. Quon; Univ. of California, Davis, Davis, CA

Abstract:

Model organisms, such as zebrafish, are often used to explore the effects of disease-associated mutations by enabling high-content imaging at early developmental stages, which is not typically feasible in human studies. The primary goal of this work is to compare images of mutant and control zebrafish to determine if a genetic mutation causes significant morphological changes. However, experimental limitations often introduce batch effects that can potentially overshadow differences due to mutation.To tackle this issue, we present a deep learning framework designed to eliminate batch effects in high-content zebrafish phenotypic imaging data, thereby enhancing genotype prediction accuracy. Our method employs a compositional mechanism combined with the adversarial optimization strategy to separate the effects of mutation from batch variations in the image data. This allows for a batch-effect-reduced feature extractor that preserves the mutant effects, leading to improved predictive performance. By integrating compositional mechanisms and adversarial components within a generative model framework, such as a variational autoencoder (VAE), our approach further facilitates direct comparison between subjects with different genotypes through in silico perturbation. Our compositional VAE outperforms traditional models, including canonical and conditional VAEs, by an average of 13.24%. Additionally, by standardizing images to the same

batch, we can create difference-images that visually highlight morphological changes due to mutations, eliminating the need for complex feature attribution techniques. In summary, our method provides a novel strategy for modeling genotype-phenotype relationships, offering a robust tool for life sciences to bridge the gap between genotype and phenotype with greater precision and clarity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4007T: A deep learning algorithm to predict risk of coronary artery disease using genetics and electronic Health record trajectory

Authors:

A. Arero[1], V. Prabhuleti[2], Y. Veturi[1,3]; [1]Dept. of Biobehavioral Hlth., Coll. of Hlth.and Human Dev., Pennsylvania State Univ., University Park, PA, [2]Cybersecurity Analytics and Operations, Coll. of Information Sci. and Technology, Pennsylvania State Univ., University Park, PA, [3]Dept. of Statistics, Eberly Coll. of Sci., Pennsylvania State Univ., University Park, PA

Abstract:

Background: Time-dependent longitudinal electronic health records (EHR) enable understanding of how the lifetime accumulation of morbidities affects the onset of complex multifactorial diseases like coronary artery disease (CAD). The integration of genetics with longitudinal EHRs for predictive modeling remains largely unexplored. **Objective:** This study aims to evaluate a transformer-based deep learning model that integrates genetic data with longitudinal EHR data to predict the onset of CAD. **Method:** In this study, we used the "Coherent Data Set," a synthetic dataset derived from Synthea™, which generates longitudinal patient-level EHR data to emulate realistic patient populations, focusing on cardiovascular disease (CVD). The dataset also includes simulated genomic data. We created a genetic risk score for CAD from GWAS-significant SNPs associated with CAD. The score was computed based on a PRS formula that pertained to CAD risk alleles since only genetic risk alleles associated with CVD were reported in the synthetic dataset. The dataset comprises clinical data from 3,500 patients; out of them, 889 had synthetic genomic sequence data. The observation window consisted of several previous visit records, with at least five visits before the first diagnosis of CAD. The prediction window included visit records in the subsequent year to identify high-risk instances. To allow the model to capture temporal dependency, we first converted diagnosis codes and visit timestamps into numerical sequences using a tokenizer and embedding layer. The model outputs were concatenated with genetic risk score inputs and passed into dense layers for the final prediction of CAD onset risk. For comparison, we trained a model without genetic risk

score integration, using only EHR-derived sequences. Model performance was evaluated using 5-fold cross-validation and the Area Under the Curve (AUC). **Result:** The performance of the proposed deep learning model showed no significant improvement when the genetic risk score was integrated with the EHR data. The AUC score for predicting CAD was 0.6384 with the genetic risk score, compared to 0.6330 without genetic risk score integration. **Conclusion**: Our study found that integrating the genetic risk score into the EHR data did not significantly improve the model performance in predicting CAD. Several limitations may have impacted these results, including the use of synthetic data, limited genomic information, and a small sample size. Our future work will address these limitations by utilizing real EHR data from TriNetX and linking genomic data from biobanks such as eMERGE and All of Us to further validate and enhance the model

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4008T: A framework to maximise genetic diversity in genome wide association study meta-analyses

Authors:

C. Yap, A. P. Morris; Univ. of Manchester, Manchester, United Kingdom

Abstract:

There have been recent efforts by the human genetics research community to increase the genetic diversity of participants contributing to genome-wide association studies (GWAS). The standard approach is to first assign participants to an “ancestry group” and then aggregate results across ancestry-specific GWAS through multi-ancestry meta-analysis. However, participants are typically assigned “continental” labels that may not reflect personal views of ethnicity/race or fully represent genetic diversity. Furthermore, some admixed participants are excluded because they cannot be assigned to a single ancestry group. Here, we present a novel pipeline for multi-ancestry meta-analysis that employs a continuous and multi-dimensional representation of ancestry by first projecting all participants onto genetic principal components (PCs) derived from the Human Genetic Diversity Project. Association analyses are then conducted for all participants in a mixed model with adjustment for PCs as covariates. Testing for interaction with PCs enables assessment of ancestry-correlated heterogeneity. We applied the new pipeline to meta-analysis of GWAS of type 2 diabetes in up to 55,090 cases and 503,503 controls of diverse ancestry. Despite increasing the genetic diversity of participants included in the meta-analysis, we demonstrate that association summary statistics were not more inflated for the new pipeline than the standard approach (genomic control lambda 1.103

for both). The new pipeline identifies 187 loci at genome-wide significance (P<5x10-8), compared with 178 for the standard approach. The new pipeline increases sample size and the diversity of participants included in multi-ancestry GWAS meta-analysis, offering enhanced power for discovery of loci that are relevant across global populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4009T: A knockoff-based framework to identifydisease-causing genes in conditionaltranscriptome-wide association studies

Authors:

X. Zhang[1], L. Wang[1], H. Zhao[2]; [1]Yale Univ., New Haven, CT, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract:

Transcriptome-wide association studies (TWASs) have been developed to nominate candidate genes associated with complex traits by integrating genome-wide association studies (GWASs) with expression quantitative trait loci (eQTL) data. However, most existing TWAS methods evaluate the marginal association between a single gene and the trait of interest without accounting for other genes within the same genomic region or the same gene from different tissues. Additionally, false-positive gene-trait pairs can arise due to correlations with the direct effects of genetic variants. In this study, we introduce TWASknock, a new knockoff-based framework for detecting causal gene-tissue pairs using GWAS summary statistics and gene expression-mapping data. Unlike marginal testing in traditional TWAS methods, TWASknock examines the conditional independence for each gene-trait pair, considering both correlations in cis-predicted expression across genes and correlations between gene expression levels and genetic variants. First, we develop per-gene expression prediction models using data from an expression-mapping study, employing basic statistical methods such as lasso regression and elastic net. These models are then used to predict individual-level gene expression in the GWAS cohort. Methodologically, we derive the correlation matrix for all genetic elements, including cis-predicted expression across genes and genotypes for genetic variants. TWASknock approximates this theoretical correlation matrix by averaging estimations from parametric bootstrap samples and then performs knockoff-based inference to detect causal gene-trait pairs while controlling the false discovery rate (FDR). Using empirical simulations based on genotype data from the UK Biobank samples, we compared the performance of TWASknock with GhostKnockoff, a knockoff inference method using summary statistics, and GIFT, a conditional TWAS method jointly modeling the expression of all genes residing in the local region. For a

fixed FDR range from 0.05 to 0.5, TWASknock consistently achieved the highest power in detecting causal gene-trait pairs within the selected risk genetic region. Compared to both GhostKnockoff and GIFT, TWASknock improved the average power of detecting genes causal to complex traits from 0.73 for GhostKnockoff and 0.66 for GIFT to 0.84. We will also illustrate the applications of TWASknock to several real data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4010T: A meta-analysis of electronic health record-based cohorts identifies rare burden signals associated with amyotrophic lateral sclerosis

Authors:

A. Guvenek[1,2], R. Wang[1], V. Pounraja[1], A. Moscati[1], J. Bras[1], J. Sul[1], E. Chen[1], N. Parikshak[1], O. Levy[2], GHS-RGC DiscovEHR Collaboration, MAYO-RGC Project Generation, Colorado Center for Personalized Medicine-RGC Collaboration, UCLA-RGC ATLAS Collaboration, Regeneron Genetics Center, E. Stahl[1], G. Coppola[1], S. Gelfman[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron Pharmaceuticals, Tarrytown, NY

Abstract:

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder impacting upper and lower motor neurons in the brain and spinal cord. While genetic studies have identified several loci associated with ALS and enhanced our understanding of both familial and sporadic ALS, the etiology of over 80% of ALS cases remains elusive. To further our understanding of the genetics of ALS, we harnessed the power of large health system cohorts to conduct exome-wide association studies using 1,252 cases and 736,369 controls of diversified ancestries. We identified genome-wide significant gene burden associations for three previously reported ALS causal genes: NEK1 (OR=6.4, p=5E-10), TBK1 (OR=45.7, p=2E-11), and SOD1 (OR=22.9, p=3.5E-9). Despite having a significantly smaller case cohort than previous studies, the signals for these known genes were stronger than previous studies owing to the large number of highly curated controls, which help in identifying rare signals substantially depleted in healthy populations. For TBK1, in particular, we found a distinctive loss-of-function signal, the effect of which is an order of magnitude higher than previous reports that included both missense and loss-of-function. These findings underscore the value of studying large health system cohorts with highly curated population controls to detect rare genetic signals associated with ALS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4011T: A new multi-trait fine-mapping method using a non-local prior, with applications in circulating metabolic biomarker level analysis

Authors:

V. Karhunen[1], J. Kettunen[2], S. Burgess[1], M. Evangelou[3], M. J. Sillanpää[2]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Univ. of Oulu, Oulu, Finland, [3]Imperial Coll. London, London, United Kingdom

Abstract:

Fine-mapping aims to identify independent causal variants within a particular genomic locus, adjusting for linkage disequilibrium (LD) patterns. Despite a number of existing fine-mapping methods for summary-level data, there is still room for improvement in the joint analysis of multiple traits, dealing with a misspecified LD reference, and incorporating functional annotation of the variants.

We introduce Multi-FiniMOM (multi-trait fine-mapping using inverse-moment priors), a new Bayesian method for simultaneous fine-mapping of multiple traits based on summary-level data. The method builds on the previously published fine-mapping approach FiniMOM, which uses a non-local inverse-moment prior for the effect sizes. In Multi-FiniMOM, the model formulation includes a hyperparameter which allows controlling for LD misspecifications, and includes the possibility to include variant-specific prior information. Rapid Markov Chain Monte Carlo sampling from the posterior distribution is achieved by using a recently proposed version of Laplace's method for integral approximation, with running times comparable to variational inference approaches.

We evaluate the performance of the method in simulations, and compare the results to another multi-trait fine-mapping method mv-SuSiE. We demonstrate that a particular strength of our method is its robustness to misspecifications in the LD reference panel. Finally, through an applied example of genetic associations for circulating metabolic biomarker levels within the APOB gene locus, we show that the method can distinguish both shared and distinct causal variants for several phenotypes. The method is implemented in a freely available R package: https://vkarhune.github.io/finimom.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4012T: A new pipeline for Killer-cell Immunoglobulin-like Receptor (KIR) genotyping from Illumina short-reads

Authors:

E. Castelli[1], R. N. Pereira[1], G. S. Paes[1], M. R. Ferreira[1], Í. S. F. Santos[1], N. Pollock[2], P. Norman[2], C. T. Mendes-Junior[3], D. Meyer[4]; [1]Unesp, Botucatu, Brazil, [2]Univ. of Colorado, Aurora, CO, [3]USP, Ribeirao Preto, Brazil, [4]USP, São Paulo, Brazil

Abstract:

The killer cell immunoglobulin-like receptors (KIRs) regulate Natural Killer (NK) cell responses by activating or inhibiting their functions. In humans, *KIR* major ligands are the *HLA* class I molecules. The ligand specificity and mode of function of each KIR are determined by its genetic sequence. Due to the high polymorphism and homology of the *KIR* genes, it is challenging to genotype them from short-read second-generation sequencing data. Alignment issues arise due to gene similarities and extreme polymorphism. Several bioinformatics pipelines have been proposed and utilized to examine *KIR* allelic diversity. Still, none of these pipelines report variants in the context of the hg38 reference genome, and they often result in conflicting outcomes. We have developed a new pipeline named **kir-mapper** for analyzing *KIR* genes from short-reads, focusing on detecting *KIR* alleles, copy numbers, and variants (SNPs and InDels) in the context of the hg38 reference genome. In brief, the pipeline aligns the reads against *KIR* genes considering known sequences (multi-referenced alignment) and forces the final alignments to the hg38 reference genome. After, it detects the copy number variation based on the observed coverage of KIR genes and some references. Then, it calls SNPs and InDels using freebayes, phasing them with whatshap. After, it reports the allele combination that best fits the observed genotypes and haplotypes. The outcomes are copy numbers, SNPs, and InDels in the context of the hg38 reference genome and allele calls based on the IPD-IMGT/KIR database. We compared the allele calls from the kir-mapper with the ones reported for 34 long-read phased assemblies from the Human Pangenome Reference Consortium (HPRC), in which Illumina short-read data is also available from the 1000 Genomes project. Kir-mapper reported the correct genotype (allele and copy number) in about 95% of the comparisons, with a 100% match for nine of the *KIR* genes, including the most polymorphic ones, *KIR3DL2* and *KIR3DL3*. We also genotyped five population samples from the 1000 Genomes dataset, comparing the results among different *KIR* genotyping tools. Kir-mapper was much faster than PING but slower than T1K. PING and Kir-mapper reported more overlapped calls than T1K, and Kir-mapper had the best performance for highly polymorphic genes such as *KIR3DL3* and others. Despite eventual conflicting outcomes, this comparison highlights the importance of applying multiple tools to obtain accurate *KIR* genotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4013T: A novel method to build context-specific polygenic scores improves prediction in complex human traits ⭐

Authors:

R. Fonseca[1,2], A. Dahl[2]; [1]Dept. of Human Genetics, Univ. of Chicago, Chicago, IL, [2]Section of Genetic Med., Univ. of Chicago, Chicago, IL

Abstract:

Polygenic scores (PGS) are poised to impact clinical care as predictors of disease risk. However, PGS can vary in predictive accuracy across individuals due to differences in their contexts, such as sex and age. This context specificity is especially important because it likely magnifies known biases in PGS portability across ancestries. To begin understanding and redressing these biases, we developed a novel method called PGSC. Our method extends the ordinary PGS by incorporating gene-by-context (GxC) interaction effects. PGSC is computationally efficient because it exploits existing tools for constructing PGS and performing GxC GWAS. PGSC is motivated by sex as a context, as sex explains large amounts of variance in many complex traits, but PGSC can be applied to any binarized context. In simulations, PGSC underperforms PGS when GxC signals are completely absent, while PGSC outperforms PGS at realistic levels of GxC heritability.

We evaluated PGSC in the UK Biobank across 30 phenotypes and 3 measures of context: sex, age, and smoking status. We performed a GxC GWAS, identifying >2,800 context-specific hits across all contexts and phenotypes. Applying PGSC to sex as a context improved prediction R2 over the standard PGS for 10/30 phenotypes at Bonferroni significance ($p < 0.05/30$). On average, across all 30 traits, PGSC outperformed the PGS R2 by 7.9%, with the largest improvements seen in traits with known sex relevance (e.g., testosterone by 18.5%, creatinine by 29.2%). We found that PGSC made comparable gains over the PGS R2 using binarized age or smoking status as contexts—6.5% and 11.2%, resp. We next compared PGSC to ampPGS, a simple model that allows the ordinary PGS to have context-specific effect sizes. Applied to age and smoking status, PGSC outperforms ampPGS for all 30 traits (mean R2 gain of 5.9% and 7.2%, resp), while ampPGS never significantly outperforms PGS. By contrast, ampPGS works better for sex as a context when applied to certain phenotypes, particularly those tied to sex hormones, consistent with prior studies—for example, ampPGS outperforms PGS and PGSC for testosterone (128.9% and 41.8%, resp). Nonetheless, on average, PGSC improves over ampPGS using sex as a context (mean R2 gain of 3.1%). We will present replication across ancestries in

UKB and in several US EHR databases, where we hypothesize PGSC will improve portability by adapting to dataset-specific distributions of contexts. Our results demonstrate that non-additive models can significantly improve genetic prediction. More broadly, we illustrate an underappreciated structural bias in current PGS, which likely magnifies their inequities and reduces their clinical utility.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4014T: A novel multivariable Mendelian randomization framework to disentangle highly correlated exposures with application to metabolomics

Authors:

L. Chan, M. M. Malakhov, W. Pan; Univ. of Minnesota, Minneapolis, MN

Abstract:

Mendelian Randomization (MR) utilizes genome-wide association study (GWAS) summary data to infer causal relationships between exposures and outcomes, offering a valuable tool for identifying disease risk factors. Multivariable MR (MVMR) estimates the direct effects of multiple exposures on an outcome. This study addresses the challenge of highly correlated exposures common in metabolomics data, where existing MVMR methods often become unscalable or suffer reduced statistical power due to multicollinearity. We propose a robust extension of the MVMR framework, named MVMR-cML-SuSiE, which employs constrained maximum likelihood (cML) and incorporates the Sum of Single Effects (SuSiE) model originally developed for fine mapping, to identify independent clusters of exposure signals. Applying MVMR-cML-SuSiE to the UK Biobank's metabolomics data for the largest Alzheimer's disease (AD) cohort through a two-sample MR approach, we identified two independent signal clusters for AD: glutamine and lipids, with posterior inclusion probabilities (PIPs) of 95.0% and 81.5%, respectively. Our findings corroborate the hypothesized roles of glutamine and lipids in AD, providing quantitative support for their potential involvement.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4015T: A Novel Varying Coefficient Model for Identification of Latent Effect Modification in Genomic Studies

Authors:

R. Wang[1], J. Jin[2]; [1]Univ. of Pennsylvania, philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Identifying associations between -omic data and disease outcomes is a common task in genomic studies. But in some scenarios, the association can be modified by some latent features of the -omic data, which may capture certain environmental factors that could affect the disease risks. the possibility of having such varying-by-latent-feature associations have not been previously investigated. In microbiome studies, for example, some latent features in the microbiome data are hypothesized to reflect diet/lifestyle that may modify the effect of microbiomes on individuals' health status. We develop a varying coefficient model, where the effects of some risk factors (e.g., microbiomes) vary as a function f of some latent features (e.g., diet/lifestyle) of the risk factors. If f is linear, then it basically adds interaction terms between risk factors and their latent features. The latent features can be linear combinations of the risk factors generated by linear factor models or nonlinear features generated by deep-learning autoencoders. Applying the varying-coefficient model on a dataset from the American Gut Program (AGP) to model metabolites by microbiomes led to a 37.5% increase in median prediction R^2 compared to main effects model on main effects of microbiomes and identified a series of latent effect modifications. Another application to a dataset from a lung cancer dataset from The Cancer Genome Atlas (TCGA) Program showed that, for modeling lung cancer stages by gene expression data, our approach led to a 49.1% - 123.5% increase in (AUC-0.5) for distinguishing between different pairs of cancer stages compared to main effects model and identified some interesting latent effect modifications associated with certain gene pathways.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4016T: A per-domain study of rare variant burden in Parkinson's Disease risk genes

Authors:

S. Parlar, C. Chantereault, K. Senkevich, Z. Gan-Or; McGill Univ., Montreal, QC, Canada

Abstract:

Objective: To study Parkinson's Disease (PD) rare variant burden per the functional domain regions of genes associated with or nominated for PD risk. **Background:** Numerous genes have been associated with PD, and many more whose associations remain unclear have been nominated through a combination of genome-wide association studies and machine learning (ML) approaches. The study of functional domain-level associations of genes with PD could allow us to identify novel associations due to the more granular nature of the approach, further revealing functional targets for study. **Methods:** Across 150 PD risk-associated/ML-nominated genes, rare variant (minor allele frequency ≤ 0.01) burden analysis was conducted via optimized sequence kernel association test (SKAT-O) in UK Biobank and Accelerating Medicines Partnership program for Parkinson's disease (AMP-PD) cohorts and MetaSKAT for meta-analysis in a total of 5,452 PD cases and 66,191 controls. Domain intervals were demarcated based on widest range across available domain annotation resources: SMART, PANTHER, Pfam, PRINTS, PROSITE, superfamily, Gene3D, CDD, and HAMAP. After variants were grouped by their domains, for further granularity of analysis, they were further stratified into groups as exonic, non-synonymous, loss of function, and potentially deleterious (high CADD-score). Lastly, Benjamini-Hochberg corrections were applied to reduce the likelihood of false discoveries. **Results:** Post-corrections, we only observed statistically significant rare variant burden in the *LRRK2* kinase domain in exonic, non-synonymous, and high CADD-score variants, with significance in UK Biobank ($P = 0.003$; $P = 0.003$; $P = 0.003$) and in the meta-analysis ($P = 0.002$; $P = 0.002$; $P = 0.002$). Our per-domain study yielded no significant results in other gene domains; however, pre-correction, nominally significant association was observed in a total of 39 domains in 34 genes. Amongst these genes, we identified burden driven by potentially protective variants present only in controls. These were p.D6E, p.I33M, p.S31N in *MPP2* L27-1 domain in UK Biobank and p.R59Q in UK Biobank and AMP-PD in *CCSER1* coiled-coil domain. **Conclusion:** The association of *LRRK2* kinase domain is well-established, as affirmed by the results of this study. We found several more associations which did not survive corrections in other genes of interest and further nominated potentially protective variants. However, these findings should be viewed with caution as they did not preserve statistical significance post-correction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4017T: A scalable approach to uncovering non-additive genetic associations in large biobanks. ⭐

Authors:

I. Molotkov[1], M. Kurki[2], FinnGen, A. Palotie[3], M. Daly[4], M. Artomov[5]; [1]The Ohio State Univ., Columbus, OH, [2]Broad Inst., Boston, MA, [3]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Nationwide Children s Hosp., Columbus, OH

Abstract:

Despite successful applications of non-additive models in genome-wide association studies (GWAS), their theoretical properties were not extensively explored in the context of large biobank data. Currently, dominant, recessive, and two-degrees-of-freedom models are used alongside the additive model to test all variants for the association with diseases, which is computationally intensive. Since additive and non-additive models yield correlated results, we hypothesized that readily available additive summary statistics can be used to pre-filter the variants, leaving only those likely to be significant in the non-additive GWAS. We show in simulations that common variants (MAF > 0.05) with additive model p-values greater than $5x10^{-5}$ or $1x10^{-3}$ almost never reach genome-wide significance in dominant and recessive GWAS, respectively, and thus can be omitted from the analysis. When applied to the data from 400,000 FinnGen participants, this approach reduced the number of common variants analyzed in non-additive GWAS by two to three orders of magnitude, depending on disease polygenicity, without losing the power to discover novel genetic associations. This enables the efficient scaling of the recessive and dominant analyses of common variants to thousands of phenotypes typically available in large biobanks.

Using population genetics, we modeled distributions of allele frequencies and effect sizes of non-additive variants. Then, by taking the expected statistical power with respect to these distributions, we compared the ability of models to detect variants under specific selection pressures, an aspect that was previously missing in analyses of non-additive models. We showed that the dominant model is expected to identify a higher number of evolutionary neutral non-additive variants than the recessive one. On the other hand, recessive and two-degrees-of-freedom models are more suited for identifying deleterious rare variants.

Aside from providing a framework for efficient non-additive GWAS that can be applied to any large biobank, we also used it in FinnGen to highlight the ability of non-additive models to produce meaningful biological findings. Studying 18 aging-related traits in FinnGen, we discovered 21 novel non-additive risk variants mapped to 12 genes lacking previous genetic association but with a known functional relationship with relevant diseases. We linked *CASC15* with type 1 diabetes, which was previously only associated with renal complications of diabetes. Additionally, we connected *LDB2* to hypertension, whereas it was previously linked only to cardiovascular diseases, with no known connection to hypertension.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4018T: A unified genealogical measure of genetic similarity accounts for relatedness and admixture in GWAS ⭐

Authors:

D. Veliz-Otani, B. Guo, T. D. O'Connor; Univ. of Maryland Baltimore, Baltimore, MD

Abstract:

INTRODUCTION
Any two haplotypes of a species, long enough back in time, coalesce to a common ancestor. By modeling the covariance of shared evolutionary history we get a similarity measure between any two individuals, regardless of population structure and relatedness level. We call this statistic GeSi (Genealogical Similarity) and demonstrate that it can be calculated directly from genotype data. We applied GeSi to population structure detection and to adjust for confounding due to population structure and relatedness in GWAS. To our knowledge, GeSi is the first approach to represent genetic similarity as a true continuum under any level of population structure and familial relatedness.
Methods
We explored whether running a PCA on the GeSi matrix could distinguish the discretely sampled, yet closely related, populations in 1KGP. Next, we simulated WGS data under a demographic model of Latin American populations. We simulated four different sampling scenarios: (1) 5000 unrelated (UNR) samples from Spain (IBS); (2) 5000 UNR samples from multiple Latin American sources (LAT); or 1400 UNR+ 3600 relatives (600 pairs each of full sibs, half sibs and full cousins) from IBS (3) and LAT(4). We simulated a polygenic trait with 0.5 h^2. All simulations were run in 10 replicates. We run a GWAS of the simulated data using either a PC-Relate GRM + 10 PCs or only the GeSi matrix (without PCs).
Results
GeSi-PCA depicted a greater distinction of closely related Latin American populations, as well as African populations, and European populations in the 1000 Genomes Project than with traditional PCA. In the GWAS benchmarking, estimating the effect sizes showed a similar power, accuracy, and bias for PC-Relate and GeSi across all scenarios. The type-II error rate of GeSi ranged between 36% and 60% of that of PC-Relate. The mean GC Lambda ranged between 0.998 and 1.01 across the four scenarios for GeSi, and between 1.04 and 1.11 for PC-Relate. To assess the bias, we regressed the estimated effect sizes on the true effect sizes, with a slope closer to one indicating less bias. The slope ranged between 0.572 and 0.609 for GeSi, and between 0.557 and 0.587 for PC-Relate. The accuracy (Pearson's correlation), ranged between 0.702 and 0.756 for GeSi, and between 0.691 and 0.751 for PC-Relate. The relative power of GeSi ranged between 85.4 and 97.2% of that of PC-Relate.

Conclusions:
We demonstrated that GeSi captures population structure and relatedness and offers an enhanced control of type II error rate in GWAS compared to a GRM+PCA while maintaining a similar power, and slightly better accuracy and bias levels. Benchmarking of PRS estimation in admixed populations is undergoing.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4019T: Analysis of maturity-onset diabetes of the young pathogenic gene variants indicate potential oversight in diagnosis

Authors:

B. Forstrom[1], J. Denny[2], J. Grose[1], M. Davis[1]; [1]Brigham Young Univ., Provo, UT, [2]NIH, Bethesda, MD

Abstract:

Maturity-onset diabetes of the young (MODY) is a rare form of diabetes mellitus that is classified by autosomal dominate alleles responsible for disruption in insulin production processes, leading to elevated blood glucose levels and other diabetes-related conditions. Because of overlap in symptoms with other forms of diabetes mellitus and the additional genetic testing required for diagnosis, up to 80% of MODY patients are misdiagnosed and potentially administered ineffective treatments. To investigate this phenomenon, we utilized the All of Us Research Program which combines over 206,173 participants with Electronic Health Record (EHR) and genomic data. To identify individuals with a MODY genotype, we filtered genomic data for variants classified as pathogenic or likely pathogenic for a MODY condition in ClinVar. We observed 189 carriers of a pathogenic or likely pathogenic variant in GCK, HNF1A, HNF4A, or HNF1B. We then identified controls by matching 10 individuals for each carrier based on sex, race, age, and body mass index. Our analysis revealed a significant increase in blood glucose measurements and A1C, as well as increases in the rates of diabetes conditions for individuals with a MODY variant. The mean hemoglobin A1C for carriers was 6.94%, while that of the matched controls was 6.13%. Half (50.26%) of carriers had a condition occurrence of diabetes mellitus present in the EHR, compared with only 17.88% in the matched controls. In the All of Us database, less than 20 individuals have a condition code specific to MODY. Among carriers, not one was present in this group, suggesting potential oversight in diabetes diagnosis. Individuals MODY-related gene variants have a significant increase in diabetes mellitus phenotypes which necessitates proper diagnosing protocols to ensure treatments leading to quality patient outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4020T: Analysis of parent-of-origin effects in over 100,000 individuals in the UK Biobank ⭐

Authors:

B. Hill, F. Lassen, D. Palmer, C. Lindgren; Big Data Inst., Univ. of Oxford, Oxford, United Kingdom

Abstract:

The phenotypic effects of genetic variation can be influenced by the parent-of-origin yet previous methods for estimating the contribution of such effects in biobanks have been limited by their reliance on recent relatedness, reducing sample sizes and the power of association studies. We introduce a method that leverages pairwise identity-by-descent (IBD) segments to link the maternally inherited X chromosome to autosomal haplotypes in males. We then apply the method to 170 billion IBD segments (>2cM) in UK Biobank, increasing the number of individuals with parent-of-origin annotations available from 26,393 to 118,836.
This four-fold increase in samples allows us to conduct the largest study of the role of parent-of-origin in complex traits to date. Alongside recapitulating of 6 out of 8 associations ($p<5x10^{-8}$) previously identified by Hofmeister et al, we find 18 novel parent-of-origin associations at genotyped sites. These include an allele associated with increased glycated haemoglobin (HbA1c) levels when inherited maternally, but decreased levels when paternally inherited (differential $p=7.0x10^{-12}$, rs2237886). Confirming these findings, this allele had previously been established as protective for type 2 diabetes (for which HbA1c is a key indicator) when inherited maternally and increasing risk if paternally inherited. In addition we identify a strong differential ($p=2.9x10^{-8}$) parent-of-origin effect on height at rs74668776. This association is among several previously unidentified in additive testing and highlights the importance of considering epigenetic information in developing a complete understanding of the role of genetic variation in otherwise saturated human traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4021T: Analysis of rare exome sequenced variants in UK Biobank to discover causal genes and fine-map causal variants for lung function

Authors:

N. Shrine[1], J. Chen[1], A. G. Izquierdo[1], A. Williams[1], A. L. Guyatt[1], C. John[1,2], R. Packer[1], L. V. Wain[1,2], I. P. Hall[3], M. D. Tobin[1,2]; [1]Univ. of Leicester, Leicester, United Kingdom, [2]Natl. Inst. for Hlth.Res., Leicester Respiratory BioMed. Res. Ctr., Leicester, United Kingdom, [3]Univ. of Nottingham, Nottingham, United Kingdom

Abstract:

Genome-wide association studies (GWAS) have discovered 1,020 mostly common genetic variants associated with lung function, implicating over 500 putative causal genes. Whole-exome sequencing (WES) is better suited to study rarer protein-coding variants that may not be well imputed. We analysed 12 million UK Biobank WES coding variants, 99% of minor allele frequency (MAF) <1%, to fine map causal variants and identify putative causal genes for lung function, not yet detected by GWAS. We performed single variant tests and gene-based collapsing tests to enhance power to detect rare variant effects in aggregate.
We included 343,104 European UK Biobank samples with WES data and four quantitative lung function phenotypes: forced expiratory volume in 1 second (FEV_1), forced vital capacity (FVC), FEV_1/FVC and peak expiratory flow (PEF). For gene-based tests, 16,141,130 variants across 24,463 genes were annotated to create 2 collapsing masks: (i) predicted loss-of-function (pLoF, 418,465 variants); (ii) pLoF + deleterious missense (additional 1,874,894 variants). Qualifying variants were tested in aggregate for each gene by regenie in two ways: (i) burden tests, and (ii) tests that account for different directions of effect. Tests were run for each mask, with MAF filters of <1%, <0.1%, <0.01%, <0.001% and singletons for burden tests and <1% for directional tests. 18,468 genes containing a qualifying variant were tested. Regenie returns an omnibus P value for each gene (GENE_P) by combining tests with the aggregated Cauchy association test (ACAT).
After single-variant testing of 6.7 million variants with MAF<1% and minor allele count >=4, 14 genes had a variant associated with lung function ($P<5x10^{-9}$), 4 of which were not previously implicated by GWAS. In gene-based tests, using a 5% threshold adjusted for 18,468 genes (GENE_P $<2.69x19^{-6}$), we found 17 associated genes, 4 of which have not been reported in GWAS, giving 8 novel genes in total. The associations at known genes indicate that rare as well as common variants have a role in lung function at these genes. We dissected the 17 gene-based associations by leave-one-variant-out analysis, which showed the gene-based association was driven by a single variant at 6 genes, a few variants (1-5) at 6 genes and by a burden of many deleterious variants (70-900) at 5 genes. The different architecture of rare-variant associations at these genes has implications for how we prioritise these genes for further mechanistic investigation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4023T: Benchmarking DNA foundation models for genomic sequence classification

Authors:

H. Feng[1], P. Wei[2], C. Wu[3]; [1]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [2]Univ. of Texas MD Anderson Cancer Ctr., Sugar Land, TX, [3]UT MD Anderson Cancer Ctr., Houston, TX

Abstract:

The rapid advancement of DNA foundation language models has revolutionized the field of genomics, enabling the decoding of DNA sequences to help understand the epigenetic patterns, transcriptional regulations, and disease associations. Pre-trained on large genomic datasets, these DNA foundation models have shown promising results in DNA sequence classification tasks after fine-tuning. However, the current evaluation of these models often depends on fine-tuning and limited datasets, which introduces biases and limits the assessment of their true potential. Here, we present a benchmarking study of three recent DNA foundation language models, including DNABERT-2, Nucleotide Transformer version-2 (NT-v2), and HyenaDNA, focusing on the quality of their zero-shot embeddings across a diverse range of genomic tasks and species through analyses of 56 real datasets. We found that DNABERT-2 exhibits the most consistent performance across human genome-related tasks, while NT-v2 excels in epigenetic modification detection. HyenaDNA stands out for its exceptional runtime scalability and ability to handle long input sequences. Furthermore, we found that using mean token embedding consistently improves the performance of all three models compared to the default setting of sentence-level summary token embedding. Notably, the performance differences between these models are significantly reduced when using mean token embedding. In summary, our evaluation identifies the factors affecting the power of DNA foundation models, and highlights the potentials and pitfalls of current models on different meaningful genomic tasks.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4024T: Benchmarking of Long Read vs Short Read Low-Coverage Human Whole Genome (lcWGS) Sequencing for Genotype Imputation and Genetic Relationship Analysis

Authors:

M. Rahardja; seqWell, Beverly, MA

Abstract:

Long read sequencing offers a powerful toolset for genomic research by providing deeper insights beyond the reach of traditional short read technologies. With recent advancements in the throughput and accuracy of this technology, such as the PacBio Revio, large-scale applications involving high levels of multiplexing are now increasingly feasible. Low coverage whole-genome sequencing (lcWGS) is an example application; lcWGS involves sequencing a genome at a low coverage (typically <5X) to enable imputation of common single-nucleotide polymorphisms (SNPs) with accuracy comparable to the widely used commercial microarrays. While short read lcWGS has been firmly established as a practical alternative to array-based genotype imputation, the relative performance and potential benefits of long read lcWGS are relatively less known. Integrating long read sequencing into lcWGS brings potential additional information: more accurate haplotype phasing data or the ability to detect structural variants. Since lcWGS requires less data than standard (~30X) WGS, cost-effective implementation of this method requires higher levels of sample multplexing. This presents an added challenge in long read sequencing, where typical mechanical DNA fragmentation and library preparation methods are more difficult, time consuming and costly. The efficient use of long read sequencing in higher multiplexing applications such as lcWGS depends on the ability to prepare long read sequencing libraries at scale. Here, we benchmark the use of long read vs short read lcWGS in reference cohorts of human genomic DNA samples (CEPH/Utah, Ashkenazi, and Han Chinese). We utilize a multiplexed library construction using Tn5 transposase to fragment and barcode samples to enable pooled PacBio SMRTbell library preparation in a PCR-free method for genotype imputation using lcWGS. The hgDNA samples were run in triplicate, fragmented and barcoded simultaneously to an average size ≥10kb using 500ng per sample. The 24 replicates were then pooled into a single PacBio SMRTbell library construction and sequenced on the Revio to mean coverage of ~5X per sample. Using the open-source GLIMPSE pipeline for variant calling and imputation and PLINK software for kinship analysis, the results of the study demonstrate consistent, reproducible library and sequencing metrics equivalent to mechanical shearing methods, precise and accurate genotyping from 5X coverage of well-characterized human genome and successful imputation of relatedness between each set. The results of our study support that long read sequencing can help unlock broad-based use of lcWGS as an alternative to microarray based technology.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4025T: Calibration of polygenic risk scores in African Americans using estimated genetic ancestry

Authors:

L. Vargas[1], M. Meyer[1], I. R. Konigsberg[1], M. R. Irvin[2], J. A. Shortt[1], L. A. Lange[1], E. M. Lange[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Alabama at Birmingham, Birmingham, AL

Abstract:

Polygenic risk score (PRS) distributions and performance vary across population groups largely due to differences in allele frequencies and linkage disequilibrium (LD) patterns. PRS distributions also vary within populations and can challenge PRS interpretation, especially in genetically admixed individuals. We developed an admixture-based PRS calibration approach, and evaluated it using PGS Catalog PRS for two traits: incident CHD (PGS003725, PGS000899, PGS000010) and height (PGS002802). PRS were calculated for self-identified non-Hispanic black (n = 8,733) and white (n = 1,692) participants from the REGARDS study. For each participant, we inferred the proportion of African ancestry (EPAA) using similarity to the YRI and CEU reference panels (ADMIXTURE v. 1.3). EPAA was strongly associated with the original PRS scores (oPRS) but was not associated with CHD incidence nor height. We calculated calibrated PRS scores (cPRS) for the black participants by regressing out from their corresponding oPRS: each individual's expected PRS mean as a function of EPAA, and dividing by the estimated standard deviation specific to the individual's EPAA decile. As expected, all PRSs performed significantly better in white as compared to black participants for both incident CHD and height. Interestingly, cPRSs performed better in black participants with lower EPAA. For CHD, using the upper decile of the corresponding PRS distribution to define participants that are at "high risk" for incident CHD events resulted in distinctly different groups of participants being in the high-risk group when using oPRS compared to cPRS. For PGS000010, <u>all</u> black participants with <60% EPAA were excluded from being classified as "high risk" for CHD using the oPRS, whereas ~10% of black participants with <60% EPAA were in the high-risk group using cPRS. For height in black participants, the PGS002802 cPRS had higher correlation with residual height (after regressing out age and sex) than the oPRS ($\rho = 0.32$ vs. 0.24). However, in multivariable regression models that include age, sex and EPAA, the cPRS and the oPRS models for height explain the same total proportion of variation in these same participants (multiple $R^2 = 0.55$ for both). Of note, in the oPRS model, EPAA was strongly associated with height, whereas EPAA was not associated with height in the cPRS model. Overall, our results demonstrate improved performance and interpretability of using cPRS over oPRS in black participants. The height example shows that failure to regress out the impact of ancestry on oPRS can diminish and complicate the interpretation of PRS as a stand-alone prediction tool.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4026T: CaseControlAF: An R package for reconstructing case and control allele frequencies from GWAS summary statistics

Authors:

H. Wolff, A. Price, C. Gignoux, A. Hendricks; Univ. of Colorado - Denver, Aurora, CO

Abstract:

Genetic summary data, such as from the Genome Wide Association Study (GWAS) catalog is highly useful including for genetic risk prediction with polygenic risk scores (PRS), genetic correlation, and causal inference using Mendelian randomization. However, robust use can be limited by the data and metadata publicly provided. An analysis of submissions to the GWAS catalog by Hayhurst et al. in 2023 found that >50% of studies were missing at least one mandatory field, with the most common missing fields being effect allele frequency (AF, >40%) and standard error (SE, >25%). Many post-hoc analyses using GWAS case-control statistics require case and control AFs, which are almost always not published-even when the whole sample AF with cases and controls aggregated is provided. Here, we present CaseControlAF, a software that derives case and control AFs from GWAS summary statistics using the whole sample AF or SE along with the odds ratio (OR), and case and control sample sizes. We compare our methods to a previously published method from the ReACt software, that uses the SE, OR, and case and control sample sizes to recapitulate case and control AFs. In simulations and real data, we find estimating case and control AFs using the whole sample AF is highly accurate across all settings. When applying this method to the Pan-UK Biobank diabetes GWAS dataset in Europeans (16550 cases, 403923 controls), the Lin s Concordance Correlation Coefficient (CCC) was >0.999. Conversely, we find that the previously published method in ReACt underestimates the MAF for higher MAF bins (mean CCC=0.975 (95% CI [0.844-0.984]) for MAF bin [0.0, 0.01); mean CCC=0.023 (95% CI [0.013-0.028]) for MAF bin [0.4, 0.5]). Through simulations, we find this bias is present when the GWAS included covariates, which subsequently changes the SE. To enable estimation of case and control AFs using SE, we present a bias correction by estimating and correcting for a function of bias fit using gnomAD v3.1.2 AFs as a proxy for true AFs. This correction improved the mean CCC across the bins from 0.454 to 0.714. While estimating the case control AF using the whole sample AF is preferred due to its high accuracy, estimating from the SE can be used more broadly since the SE can be calculated from the p-value and beta estimate, which are more commonly provided. CaseControlAF expands the utility of publicly available genetic summary statistics by reconstructing case and control sample-specific AFs which can be used to estimate substructure within summary data and for secondary analyses such as group PRS and case-case GWAS. Both methods have been implemented in an R package CaseControlAF, which is freely available on GitHub.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4027T: Cell Type-Specific contributions of PRS of 34 complex traits

Authors:

M. Li, Y. Yang, X. Zhu; Case Western Reserve Univ., Cleveland, OH

Abstract:

Introduction: Polygenic risk score (PRS) has been developed to predict a disease trait using genome wide genetic variants, irrespective of the causal relationship between these variants and the trait. Most of the genetic variants identified from genome wide association studies (GWAS) are noncoding variants that may influence a trait by affecting transcriptional regulatory elements in specific cell types. Zhang et al. (2021, Cell) provided a single-cell atlas of chromatin accessibility in the human genome, revealing the open chromatin status for about 1.2 million candidate cis-regulatory elements (cCREs) across 111 adult cell types. This resource provides an opportunity to partition a PRS to cell-type specific contributions and identify most relevant cell-types for a trait.
Methods: We downloaded GWAS summary statistics of 34 traits, including metabolic, cardiovascular, blood cell, mental, and neurological related traits, and the 111 cell-types specific CRE annotations. For each trait, we constructed its PRS using SBayesRC and obtained the corresponding variant effect sizes. We next made a regression of these variant effect sizes on the 111 cell-type specific CRE annotations to calculate cell-type specific contributions. Due to the overlap among the cell-type specific CRE annotations, we used the Pratt index (PI) (defined as the marginal correlation of a dependent variable and a predictor multiplied by the regression coefficient) to evaluate cell type importance. We applied the weighted lasso to select relevant cell type annotations and block-wise bootstrap for calculating standard errors (SE).
Result: We observed relevant cell types ranging from 1 to 8 in the 34 traits we studied. As some examples, the most relevant cell types for Attention-Deficit/Hyperactivity Disorder are Glutamatergic Neuron 1 (PI=0.09, SE=0.02), and Glutamatergic Neuron 2 (PI=0.08, SE=0.02). For Cystatin, the most relevant cell types are Hepatocyte (PI =0.11, SE=0.03), Glutamatergic Neuron 1 (PI =0.08, SE=0.03), and Pancreatic Beta Cell 1 (PI=0.07, SE=0.02). For Low-Density Lipoprotein, the most relevant cell types are: Colon Epithelial cell 3 (PI = 0.11, SE=0.04) and Thyroid Follicular Cell (PI = 0.07, SE=0.03). In summary, 66 cell types are identified as relevant cell types at least once with Glutamatergic Neuron 1, Follicular Cell, and Hepatocyte most frequent.
Conclusion: Our study systematically identified key cell types for common complex traits, providing crucial data for future transcriptome studies for understanding GWAS loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4028T: Challenges associated with the use of biobank data as external controls in genome-wide association studies

Authors:

J. Pattee[1], L. Fishbein[1], C. Gignoux[1], T. Else[2], K. Koc[1], A. Hendricks[3]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Michigan, Ann Arbor, MI, [3]Univ. of Colorado Denver, Denver, CO

Abstract:

Genome-wide association studies can often benefit from the use of external controls. Such controls can be used to augment an internal control set or to provide controls when none are available (e.g., for a rare disease registry or for hereditary cancers). Biobank data is a natural choice for external controls given that it is often accessible, inexpensive, and has already been cleaned and processed. However, there are challenges associated with achieving unbiased inference using biobank data as external controls. These challenges include batch effects arising from different genotyping platforms, differences in processing, and ascertainment bias. We investigate the challenges associated with using biobank data as external controls via a genome-wide association study (GWAS) of pheochromocytoma and paraganglioma, a rare type of neuroendocrine tumor, as an example. The primary analysis involves germline DNA from 1,217 rare cancer cases and 27,410 external controls drawn from the Colorado Center of Personalized Medicine Biobank. We describe our harmonization pipeline including quality control, imputation, and association. We detail the usefulness of re-genotyping a set of biobank controls alongside the cases for performing checks throughout the process including initial concordance and a control-control GWAS. We describe the importance of balancing case and control samples across imputation batches to avoid confounding and bias. Lastly, we show how a control-control GWAS of the 27,410 Biobank controls against the 453 internal controls helped to identify a peak of spurious association signals on chromosome 2 and a broader bug in the imputation method implemented. These processes facilitate the investigation of potentially spurious associations and provide an understanding of where biases may be arising from. We present general guidelines for the selection of a disease-appropriate control sample from biobank data and describe how this process was performed in our study. The effective use of biobank data as external controls facilitates GWAS analysis where it would otherwise not be possible, such as for rare disease. Leveraging cost effective, easily accessible controls lowers the barrier to entry for GWAS and can enable powerful analysis for rare or uncommon conditions by increasing the

sample size or providing a set of controls where none exist. We gratefully acknowledge the support of NIH/NCI R01CA246586 and of R35HG011293.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4029T: Characteristics of rare variants relevant for dilated cardiomyopathy depend on local ancestry: The DCM Precision Medicine Study

Authors:

D. D. Kinnamon, J. O. Mead, J. R. Cowan, E. Jordan, R. E. Hershberger; The Ohio State Univ., Columbus, OH

Abstract:

The multi-site DCM Precision Medicine (PM) Study found that probands of predominantly African ancestry (AA) were less likely to have clinically actionable rare (alternate allele frequency ≤0.1%) variants in 36 established dilated cardiomyopathy (DCM) genes relative to probands of predominantly European ancestry (EA). Major contributors to this difference were a paucity of predicted loss of function (pLOF) variants and less available evidence to support the pathogenicity of more numerous non-pLOF variants in DCM probands of AA compared to EA. We hypothesized that these differences might reflect dependence of DCM-relevant rare variant characteristics on local ancestry. To explore this hypothesis, we used array-based genotypes to estimate genome-wide local ancestry in 1107 DCM PM Study probands with ≥87.5% combined global AA and EA using 1000 Genomes high-coverage whole genome sequencing data as a reference. For each DCM-relevant rare variant identified in these probands, local ancestry dosages at the closest 5' and 3' markers within 1 MB were linearly interpolated to estimate the percentages of local AA and non-AA in each individual harboring the variant. If multiple individuals harbored a rare variant, these percentages were averaged across individuals to obtain a variant-level average local ancestry percentage. Missense variant effects were predicted with AlphaMissense, ESM1b, and CADD v1.7, all of which incorporate novel deep protein language models and were not trained on existing variant curation databases. For 68 pLOF variants found among 522 admixed probands with ≥12.5% global AA, the variant's average local AA proportion was a mean of 0.078 lower (95% CI: -0.155 to -0.002; p=0.05) than the average global AA proportion of its carriers, suggesting that pLOF variants were less likely to occur on local AA than local non-AA. While local and global AA proportions were comparable for the 309 non-pLOF variants in these admixed probands, an analysis of 589 missense variants found among all probands showed that the 279 with ≥25% average local AA had lower

pathogenicity scores for AlphaMissense (p=0.004), ESM1b (p=0.09), and CADD (p=0.09) compared to the remainder when considering variants with scores from a given technique. The proportions classified as pathogenic using recommended thresholds were also lower (AlphaMissense 21% vs. 32%, p=0.004; ESM1b 48% vs. 56%, p=0.08). We concluded that DCM-relevant rare variant characteristics depend on local ancestry; ongoing work will more comprehensively characterize this dependence, including whether these variants' effects in first-degree relatives who share them depend on local ancestry.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4030T: Characterizing Genetic Diversity in the All of Us Research Program in the Context of Genetic Association Studies

Authors:

S. Gunn[1], P. Lazcano[1], F. Shiau[1], D. Wang[1], G. Merolli[1], S. Leibovitz[1], F. Gulamali[2], M. Kim[1], A. B. Rovsing[1], T. Singh[1]; [1]New York Genome Ctr., New York, NY, [2]Icahn Sch. of Med., New York, NY

Abstract:

The All of Us Research Program (AoURP), a major US-based precision medicine initiative, aims to represent populations historically underrepresented in genome-wide association studies (GWAS). Until now, large biobanks studied within the genetics community have mostly originated from more genetically homogeneous populations like Finland and the UK. Studying the diverse populace of the US requires methods that both consider and leverage population structure and diversity in genetic discovery and analysis.

The v7 release of AoURP provides genetic ancestry (GA) definitions from a multi-class random forest classifier trained on the Human Genome Diversity and 1000 Genomes (HGDP-1KG) Projects reference data and only 16 global principal components (PCs). To further characterize the genetic diversity within AoURP, we performed a two-stage decomposition of the GA of the 245K individuals in the v7 release. First, we annotated groups in relation to continental populations. Second, we decomposed GA within each group with tier-2 PCA and built a population-specific multi-class classifier, relating individuals to mixtures of known historical migrations into the US.

Globally, we find broad correspondence between our subpopulations and the GA groups reported by AoURP. However, global PCs cannot capture structure within subpopulations. Reflecting recent immigration from Latin America to the US, the American tier-2 PCs resolve population structure well, delineating Mexican, Puerto Rican, and Peruvian populations. In contrast, the European (EUR) tier-2 PCs suggest evidence of intra-continental admixture,

corresponding to multiple waves of EUR immigration. We also see significant inter-continental admixture with the African tier-2 PCs.
The level of inter- and intra-continental admixture in the AoURP is unprecedented in large genomic biobanks. We evaluated the significance of controlling for intra-continental population structure by performing GWAS on 10 traits of different architectures. We assessed various GWAS metrics, including genomic control, LDSC heritability and intercept estimates, and the number of genome-wide loci discovered. For example, in height and BMI GWAS in EUR, we find reductions in lambda, LDSC-intercepts, and heritability, suggesting that more refined controls are needed to account for underlying population substructures, reduce spurious associations, and generate more robust population-level estimates. Our findings show that fine-grained population control beyond the provided global PCs is needed to maximize the value of AoURP's diversity and suggests a path forward for future multi-ancestry meta-analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4031T: Chasing Shadows: Identifying genes for undiagnosed rare diseases through statistical analysis of a large hospital cohort ⭐

Authors:

K. Kim[1,2,3], R. Koesterer[2], H. Zhang[1], C. French[1], R. Doan[1,3], B. Raby[1,3], A. Beggs[1,3], A. Poduri[1,3], S. Rockowitz[1], J. Flannick[1,2,3]; [1]Boston Children's Hosp., Boston, MA, [2]Broad Inst., Cambridge, MA, [3]Harvard Med. Sch., Boston, MA

Abstract:

The process of a genetic diagnosis for a rare disease is often referred to as a "Diagnostic Odyssey" due to complex phenotypes and oligogenic effects that can result in misdiagnosis and delays. Traditional genetic analysis is only able to identify causal genes in approximately 40% of patients. We conducted a study to use statistical analyses of large hospital cohorts with various phenotypes to identify patients who remain undiagnosed by traditional methods. We analyzed 17,242 patients from 46 cohorts within the Children's Hospital Rare Disease Cohorts (CRDC). These patients were defined by contributing investigators and IRB protocols. Phenotypes were assigned to patients by mapping their Electronic health record information to the Human Phenotype Ontology (HPO). To identify potential novel genetic diagnoses, we applied a four-step statistical approach: (a) using "Training cohorts" of patients to statistically learn a profile of phenotypes that characterize a disorder; (b) scoring each patient within the CRDC for similarity to these profiles; (c) applying ACMG criteria to screen for variants in patients with high similarity scores; and (d) conducting statistical "Burden" analysis to identify novel genes associated with

similarity to the profile. We applied this approach to cohorts collected for four diseases: Epilepsy, Nemaline myopathies, Pulmonary and respiratory disorders, and ADHD. We identified 19, 101, 72, and 37 samples, respectively, that demonstrated high phenotypic similarity to the training cohorts despite being ascertained on different criteria. These patients carried variants in several genes previously implicated in the training diseases: *SCN1A* (epilepsy), *MYL22* (myopathies), *SFTPB* (respiratory), and *SPTBN1* (ADHD). Gene-burden tests identified additional candidate genes, including *GPR61* and *CBX1* for epilepsy and myopathies. The findings illustrate the significance of large hospital cohorts in guiding genetic discovery for rare diseases. Our approach is general and scalable, and its effectiveness will improve as larger collections of rare disease patients are integrated.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4032T: Clinical Utility, Feasibility and Cost-Effectiveness of Polygenic Risk Scores for Breast Cancer Screening in Europe

Authors:

D. O'Mahony[1], I. S. Brorson[2], T. Hovda[3], K. Kollstad[4], S. Bos[5], K. Sahlberg, Kleivi[3], G. G. Alnaes[1], K. Michailidou[6], A. Y. Sundaram[1], K. Kruuv-Käo[7], E. Grindedal[1], E. Jørgensen[4], E. Hovig[1], S. Sõber[7], P. Padrik[7], V. Kristensen[8]; [1]Oslo Univ. Hosp., Oslo, Norway, [2]Univ. of Oslo, Oslo, Norway, [3]Vestre Viken Hosp. Trust, Drammen, Oslo, Norway, [4]Oceona, Ås, Oslo, Norway, [5]Univ. of Oslo, Oslo, Oslo, Norway, [6]The Cyprus Inst. of Neurology and Genetics, Nicosia, Cyprus, [7]OÜ Antegenes, Tartu, Estonia, [8]Oslo Univ., Oslo, Norway

Abstract:

Background. Polygenic risk score (PRS) is an established risk assessment tool in breast cancer. Despite evidence on guiding early screening and diagnosis in high-risk individuals, several considerations halt PRS implementation in the clinic. We of the AnteNOR consortium, to facilitate risk-based mammographic screening, would like to assess the potential of PRS in the clinic, its transferability between countries and its cost-effectiveness compared to current practices. **Methods**. PRSs were computed for 2,680 breast cancer cases and 267 controls from Norway and 31,868 cases and 680 controls from Estonia. PRS was associated to clinical features and transformed to 10-year absolute risks for breast cancer. A population-based prospective clinical trial was conducted where 80 women from Norway aged 40-50, with no history of breast cancer, premalignant breast disease or genetic testing were given risk-specific recommendations based on AnteBC PRS. A cost-utility analysis was conducted using a Markov model for biennial screening program at 50-69 (standard screening in Estonia & Norway), compared to PRS-tailored

screening for 1,000 women aged 35 as they moved between the states healthy/at risk, stages 0-IV and dead, over a lifetime. The incremental cost-effectiveness ratio was calculated by dividing cost difference costs by the difference in quality-adjusted life-years (QALY) between the two strategies. **Results**. PRSs showed comparable high performance between Estonia and Norway, in which 318 (7.5%) and 265 (8.8%) women respectively of ages 30-49 had a higher 10-year breast cancer risk than individuals at age 50. A higher PRS was associated with ER-, PR-status and tumour grade, and for the first time we report associations to therapy response, metastasis and relapse. The clinical trial identified 40 women (49%) at relative risk ≤1, 31 (39%) >1 and ≤2 and 10 (12%) >2. Based on this, 40 women were recommended standard screening, 40 were recommended biennial screening before 50 and of these were recommended annual mammography at the age when their risk was twice the average risk at age 50. The majority of women were positive to risk assessment and did not experience anxiety for breast cancer. Risk-stratified screening led to redistribution of breast cancer stages, with earlier (0-I) and fewer advanced stages (II-IV) and averted 1.5 breast cancer deaths per 1,000 women screened. Although PRS-based screening resulted in larger net costs, a gain of 3.85 QALYs indicated cost-effectiveness. **Conclusions**. Overall, new evidence is provided for the usefulness of PRS as a personalized risk prediction tool to support its clinical usage and guide risk-based mammographic screening.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4033T: Coding cis pQTLs from proteogenomic data allow evaluation of the performance of missense variant effect predictions and the utility of their application to rare variant association analyses.

Authors:

S. Klasfeld[1], J. Chiou[1], E. Fauman[1], M. Miller[2], R. Moccia[3], H. Kim[1]; [1]Pfizer, Cambridge, MA, [2]Pfizer, Dorchester Center, MA, [3]Pfizer Inc., Cambridge, MA

Abstract:

Large-scale sequencing studies have identified numerous rare protein-altering coding variants, but their functional effects remain largely unknown. The latest variant effect prediction methods, such as AlphaMissense and ESM1b, utilize unsupervised machine learning to predict the impact of missense variants based on sequence conservation, context, and protein structure and show superior performance over existing methods. This study aims to evaluate the performance of these new prediction methods using coding cis pQTLs as a benchmark and assess their utility in rare variant association analyses using proteomics and exome sequence

data from the UK Biobank (n = 42,513 individuals of European ancestry).
We generated a catalogue of coding variant effects on protein levels by testing the association between 2,941 plasma protein levels and cis coding variants in 32,325 individuals. Adjusting for GWAS pQTLs as covariates removed a substantial number of coding cis pQTLs and revealed novel ones, the number of which varied across minor allele frequencies and variant impact classes. A total of 3,484 significant coding cis pQTLs (p-value < 1.2e-7) were identified and used to benchmark prediction scores from dbNSFP, ESM1b, and AlphaMissense. Missense variants with higher prediction scores for being deleterious were found to be enriched for coding cis pQTLs and have stronger effects on protein levels. The correlations between prediction scores and effect sizes ranged from 0.2 to 0.3 across the methods, although these may be underestimating the performance as pQTLs only capture the effects on protein levels and not protein function. We next assessed whether using missense prediction scores to guide variant inclusion improves rare variant association analyses in the remaining 10,188 individuals. Each prediction method identified unique deleterious variants in addition to those commonly found by other methods. Among the 2,900 protein-gene pairs tested, dbNSFP prediction alone found 581 significant associations while incorporating ESM1b and AlphaMissense predictions identified 663 significant associations, including 116 that were not found with dbNSFP alone, indicating improved statistical power.
In conclusion, this study assessed the performance of variant prediction scores using coding cis pQTLs and demonstrated the utility of applying these predictions in rare variant association analyses. Future work will involve evaluating the performance of prediction methods using mutagenesis datasets that capture the variants' effects on protein function and incorporating prediction scores as continuous weights in burden analyses to enhance statistical power.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4034T: Common Variants on Amerindigenous Local Ancestry Share Similar Causal Effects with Those on Other Ancestries in Complex Traits

Authors:

Z. Shi[1], R. Mandla[1], K. Hou[1], Y. Wang[2], K. Tsuo[3], E. Atkinson[4], A. Martin[2], B. Pasaniuc[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Baylor Coll. of Med., Houston, TX

Abstract:

Admixed individuals inherit genetic segments from different ancestries (local ancestry), and the genetic causal effects across these local ancestries in complex traits are not well-studied. Meanwhile, Amerindigenous (AMI) admixed populations are particularly underrepresented in large genomic studies. This study aims to compare the causal effects of common variants between AMI and other local ancestries in complex traits, and we employed the All of Us Research Program, which includes genetic data from 245,394 individuals and 41,778 admixed American (AMR) individuals with proportion of AMI ancestry greater than 0.05. We leveraged the estimation of correlation of causal genetic effect (R_{admix}) across local ancestries from Hou. et al. on 18 complex quantitative traits in the AMR population with an average sample size of 24,554. The result shows notably high correlation between AMI and other local ancestries (meta-analysis R_{admix}=1, 95% credible interval 0.983-1), indicating the causal effects are nearly identical between AMI and other ancestries. The result was replicated using the regression-based marginal local ancestry-aware association method on clumped GWAS significant variants. Based on 251 SNPs across 18 traits, the Deming regression reported a slope of 0.957 (standard error 0.03). The result matches the R_{admix}, given marginal correlation brings inflated heterogeneity. In conclusion, we explored the causal effects on common variants between Amerindigenous and other ancestors across complex traits, and found them highly similar using R_{admix} and regression-based marginal correlation. This work not only promotes the understanding of common variants in Amerindigenous underrepresented populations, but also brings insight into the study of complex traits in local ancestry scale.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4035T: Comparison of genetically regulated prediction models in lipidomics and transcriptomics in a Hispanic cohort

Authors:

W. Zhu[1], H-H. Chen[2], A. Petty[3], L. Petty[3], H. Highland[4], M. Yaser[5], K. Young[6], P. Meikle[7], J. Curran[8], J. McCormick[9], S. Fisher-Hoch[10], K. North[6], E. Gamazon[11], J. Below[12]; [1]Vanderbilt Univ., Nashville, TN, [2]Academia Sinica, Taipei, Taiwan, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [5]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [6]Univ North Carolina, Chapel Hill, NC, [7]Baker Heart and Diabetes Inst., Melbourne, Australia, [8]Univ. of Texas Rio Grande Valley, Brownsville, TX, [9]UTH SPH Brownsville, Brownsville, TX, [10]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [11]VUMC; Clare Hall, Univ. of Cambridge, Nashville, TN, [12]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

Background: Data from different omics such as genetics, transcriptomics, and metabolomics have become more accessible as a result of the rapid development of high-throughput technologies. The integrated analysis of multi-omics data has proven to provide comprehensive insights to complex biological mechanisms. While predicted gene expression levels are widely used in genetic studies, fewer attempts have been reported on modeling in the human lipidome. Moreover, models built in different omics types have limited testing in secondary or independent data resources with paired genomic and molecular profiling. Characterizing the performance of predictive models in new data resources is needed to ensure rigor. **Methods:** Lipidomic profiling on plasma samples was performed on a HPLC and mass spectrometer coupled system. A total of 830 lipid species from 49 lipid classes were detected. The genotyping was processed on Illumina MEGA^{EX} array and imputed to the TOPMed phase 8 reference followed by standard QC. Whole-blood RNAseq was obtained using TruSeq stranded mRNA kit and sequenced on the Illumina NovaSeq platform. Elastic net models were trained using 10-fold cross-validation. Using lipidome traits measured in the Cameron County Hispanic Cohort (CCHC), we trained elastic net prediction models on 830 lipid species through genetic variants selected by GWAS p values on a training set of 1606 samples. A separate test set of 681 samples was used to evaluate the true performance of the models using Pearson's correlation (r). The appropriate p-value threshold to filter SNPs was determined to be 1e-6. The transcriptomics data was split to 1029 for training and 259 for testing. SNPs in the +/- 1MB region of each gene were selected for model building. **Results:** The overall model performance in the lipidomics is high in the training set (mean r=0.67), but low in the test set (mean r=0.11). Models in 40 lipid species have moderate performance (r>0.3). Mean model performance in the transcriptomics was similar in the training set (mean r=0.29) and the test set (mean r=0.18). Out of 25,782 genes, we identified 5,321 models with performances greater than 0.3. **Conclusion:** Our result suggests that while both lipidome traits and transcripts can be predicted through variants using elastic net, lipidopmics models are more likely to suffer from overfitting, and thus have low generalization in the test set. Further evaluation, larger sample sizes with more precise effect estimates, or modified statistical modeling may be needed to rigorously leverage genetically regulated predictions of the lipidome.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4036T: Comparison of methods for accurately estimating hereditary cancer prevalence in diverse U.S. biobanks in the presence of participation bias

Authors:

S. Hanks[1], R. Shemirani[1], C. Caggiano[2], K. Ferar[1], E. Soper[1], E. Kenny[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Icahn Sch. of Med. at Mt Sinai, New York, NY

Abstract:

Electronic health record (EHR)-linked biobanks have facilitated cost-effective genetic discovery and translational efforts, including the development of genomic screening tools for medically actionable diseases. However, the nonrandom selection of study participants into biobanks can bias estimates of disease prevalence and genetic associations derived from these resources. Here, we estimate the impact of participation bias on the prevalence of two CDC-designated Tier 1 medically actionable genetic cancer conditions, hereditary breast and ovarian cancer (HBOC) and Lynch syndrome (LS), in two large, diverse US-based biobanks: the Bio*Me* in New York City and the All of Us (AoU) national program. Bio*Me* is an EHR-linked biorepository in a single large health system in New York City, and participants were enrolled in-person primarily through ambulatory care practices from 2007-2023. Digital enrollment (via website or phone app) to AoU began in 2018 at one of the participating AoU healthcare providers (including regional medical centers, federally qualified health centers, and the Veterans Health Administration) or by direct volunteering. Of the participants who completed surveys, the mean age was 51 in AoU (N=413,360) and 61 in Bio*Me* (N=57,785), 60.1% and 57.9% were female, 83.1% and 65% born in the US, and 53.9% and 34.2% self-report as White. We will compare multiple strategies to derive survey weights for calculating the adjusted prevalence of HBOC and LS in the biobanks. We will first use inverse probability weights from the propensity scores of biobank participation, which we will calculate by leveraging information about demographics, disease, and social determinants of health from external, nationally representative epidemiological data sources. We will also use weights derived from a recent method that estimates a polygenic score of participation using closely related individuals. This method compares allele frequencies between haplotypes that are shared and not shared identical-by descent (IBD) between relative pairs, for which we have called IBD tracts in 859 full sibling pairs and 1,735 parent-offspring pairs in the Bio*Me* biobank. To evaluate these approaches, will use genome- and exome-sequencing data to estimate the prevalence of pathogenic variants in genes underlying for HBOC and LS in Bio*Me* and AoU, apply each adjustment strategy, and compare adjusted HBOC and LS prevalence estimates against published rates. We will also compare adjusted incident rates for the cancer phenotypes to the SEER cancer registries. We expect our findings to help inform genetic screening guidelines, biobank recruitment efforts, and community risk awareness.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4037T: Comprehensive evaluation of pleiotropy methods leveraging GWAS summary statistics

Authors:

R. Salem[1], J. Xi[1], W. Gu[2]; [1]UC San Diego, La Jolla, CA, [2]Univ. of California, San Francisco, San Francisco, CA

Abstract:

Background: Pleiotropy is the phenomenon in which a single locus affects two or more unrelated phenotypic traits and is hypothesized to be ubiquitous in the human genome. Pleiotropy analysis leverages the complexity within the human genetic architecture to improve the power of discovery and uncover the shared etiology of disease. The availability of large-scale GWAS summary statistics has facilitated the emergence of a diverse set of post hoc pleiotropy analysis methods. Nevertheless, a comprehensive evaluation and comparison of these methods under various scenarios is absent.
Methods: We set out to benchmark the performance of 29 pleiotropy statistical methods under a variety of scenarios. We generate a set of R packages. PleioSim allows the simulation of genotypes and phenotypes with configurable parameters such as allele frequencies, sample size, trait heritability, effect sizes, trait genetic correlation (pleiotropy), and environmental correlations. Following data simulation, genomic analyses were conducted to generate summary statistics. The PleioAnalyse package incorporates 29 distinct pleiotropy methods from 12 major libraries, utilizing F1 scores for performance assessment. Finally, we developed PleioVis, a set of tools to help with visualization and comparison of results.
Results: The 29 pleiotropy methods evaluated fell into three broad categories: simple combinations, Fisher combinations, and principal components approaches. The best performance was found for Fisher combination methods, with the MixTippett method generally outperforming others across diverse scenarios, followed closely by MixFisher. Under conditions of lower statistical power, such as with rare variants or high environmental correlations, principal component (PC) -based methods like PCminp and PCLC showed superior performance.
Conclusion: We considered a comprehensive set of simulation scenarios and pleiotropy methods to assess their performance to identify pleiotropic variants in genetic studies. Our findings underscore the utility of Fischer combination methods for wide variety of scenarios while highlighting the utility of PC methods in low-power situations. Further research is needed to develop a universally robust pleiotropy method for all scenarios.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4038T: Considering confounding in rare variant genome wide association studies ⭐

Authors:

A. Hanson, D. Lawson, N. Davies, G. Davey Smith, G. Hemani; Univ. of Bristol, Bristol, United Kingdom

Abstract:

The interrogation of rare genetic variant associations with complex traits is now achievable in large-scale population cohorts of whole-exome or whole-genome sequenced individuals. This class of variant is enriched for mutations with large and predicted deleterious effects on biological function, making them informative in understanding molecular mechanisms of disease, and attractive instruments to proxy biological exposures in mendelian randomisation (MR) studies. The extent to which rare-variant associations are biased by inadequately controlled demographic (e.g. population stratification and assortative mating) and indirect (e.g. parental) genetic effects in population genome-wide association studies (GWAS) is currently unclear. Though rare variants reflect more recently emerging ancestral substructures in human populations, this is not completely captured by rare-variant derived principal components. Using a within-family model, in which such confounders are most successfully controlled, we tested for bias in published rare-variant associations with 166 quantitative traits in the UK Biobank. We saw little evidence of heterogeneity in rare SNP effect estimates from population GWAS and within-family models overall, but evidence for residual bias in some demographically stratified traits, including height and educational attainment, which subsequently biased MR causal estimates. Biased SNP effects captured a mixture of fine-grained population stratification and polygenic confounding, thorough which rare variants appeared to tag common-variant burden for polygenic phenotypes in carriers. Polygenic confounding also appeared to result in some false positive rare-variant associations with height, and may similarly impact associations with other highly polygenic traits. Future use of family studies to detect biased genetic effects will facilitate sound interpretation of rare-variant associations and improve accuracy in causal inference methodologies utilising rare genetic instruments.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4039T: Cox proportional hazards mixed models accurately estimate the heritability of time to event traits

Authors:

K. Taraszka, A. Gusev; Dana-Farber Cancer Inst., Boston, MA

Abstract:

Heritability quantifies the variance in disease susceptibility attributable to genetics and is a fundamental parameter for interpreting disease architecture. Currently, the vast majority of heritability methods are designed for binary and quantitative traits, both of which rely on the same underlying phenotypic generative model (PGM). This PGM assumes a linear relationship between genetics and the trait as well as the assumption that case-control status is invariant. However, some controls may simply have been observed prior to diagnosis, particularly in age-related diseases such as cancer and cardiovascular disease. Fortunately, recent efforts have considered a different PGM reflecting time-to-event (TTE) phenotypes. Under this PGM and with infinite time, all individuals would develop the trait though only a subset of the cases are observed due to censoring (e.g. death). Classically, data of this form would be analyzed with a Cox proportional hazard regression which has been implemented for genome-wide association studies (GWAS), including as a mixed model. While these advancements are promising, the total variance in disease susceptibility attributed to genetics (i.e. heritability) of TTE traits has remained largely unstudied. In fact, though there are numerous heritability methods for quantitative and binary phenotypes, no such method exists for TTE traits. This has limited the characterization of age-of-onset and treatment response phenotypes, the latter particularly relevant for clinical studies.
We address this lack of TTE heritability methods by introducing COXMM, a COX proportional hazard Mixed Model designed to estimate the heritability of traits which follow a TTE PGM. We demonstrate the efficacy of COXMM for TTE heritability estimation, both in simulations and in the UK Biobank. **In particular, we find that COXMM is consistently empirically unbiased under a variety of TTE PGMs while traditional heritability methods showed substantial downward bias.** Using the UK Biobank, we analyzed ten age-related cardiovascular traits and observed the average TTE heritability was 0.54 while the average case-control heritability estimate was 0.28. We validated the higher TTE heritability estimate by showing the TTE PRS was a more significant predictor for 7/10 traits. Ultimately, disease susceptibility is a well studied phenomenon while the heritability of TTE phenotypes such as age-of-onset or progression remain understudied. Furthermore, this new framework enables heritability estimation for interval phenotypes and treatment response traits, neither of which can be properly modeled using traditional PGMs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4040T: Deciphering causal protein biomarkers in Alzheimer's disease: Integrating a novel robust Mendelian randomization method for proteomics data analysis and AlphaFold3 for predicting 3D structural alterations

Authors:

M. Yao[1], G. M. Miller[2], B. Vardarajan[2], A. A. Baccarelli[3], Z. Guo[4], Z. Liu[2]; [1]The Univ. of Hong Kong, Hong Kong, China, [2]Columbia Univ., New York, NY, [3]Harvard Univ., Boston, MA, [4]Rutgers Univ., Piscataway, NJ

Abstract:

Hidden confounding bias is a major threat in identifying causal protein biomarkers for Alzheimer's disease in non-randomized studies. Mendelian randomization (MR) framework holds the promise of removing such hidden confounding bias by leveraging protein quantitative trait loci (pQTL) as instrumental variables (IVs) for establishing causal relationships. However, some pQTLs might violate core IV assumptions, leading to biased causal inference and misleading scientific conclusions. To address this urgent challenge, we propose a novel MR method called MR-SPI that first Selects valid pQTL IVs under the Anna Karenina Principle and then performs valid Post-selection Inference that is robust to possible pQTL selection error. We further develop a computationally efficient pipeline by integrating MR-SPI and AlphaFold3 to automatically identify causal protein biomarkers and predict protein 3D structural alterations. We apply this pipeline to analyze genome-wide summary statistics for 912 plasma proteins in 54,306 participants from UK Biobank and for Alzheimer's disease (AD) in 455,258 samples. We identified seven proteins associated with Alzheimer's disease - TREM2, PILRB, PILRA, EPHA1, CD33, RET, and CD55 - whose 3D structures are altered by missense genetic variations. This discovery offers novel perspectives into their biological roles in AD development and may aid in identifying potential drug targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4042T: Decoding the shared genetic structure of heterogeneous complex diseases

Authors:

S. Banerjee[1], N. Mullins[2], D. Knowles[3]; [1]New York Genome Ctr., New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]New York Genome Ctr. & Columbia Univ., New York, NY

Abstract:

Heterogeneous complex diseases like neuropsychiatric disorders, coronary heart disease and cancer encompass multiple different subtypes with different clinical phenotypes and pharmacology, yet they often culminate in similar clinicopathological outcomes. Despite emerging evidence that genetic heterogeneity and pleiotropy may account for at least some proportion of this phenotypic heterogeneity, its molecular mechanisms remain poorly understood. In this study, we propose a model which assumes that heterogeneous complex traits arise from a combination of both shared and unique genetic factors among the subphenotypes. We introduce novel convex optimization methods to extract a low-rank component and a sparse component from the matrix of summary statistics of genetic associations with the relevant traits. We use the Frank-Wolfe algorithm to solve an optimization problem constrained by a combination of the nuclear norm and the l1 norm. We demonstrate that the method can recover both the low-rank and sparse components even in the presence of corrupted or missing data entries. Through singular value decomposition of the low-rank component, we identify and characterize the trait-specific loadings of the shared latent genetic factors for the subphenotypes. The sparse component corresponds to the unique genetic contributions to individual subphenotypes. The convex optimization guarantees the reproduction of the same results from the same data and robustness against small, functionally insignificant changes in the input data. Extensive simulations reveal that our method outperforms the state-of-the-art approaches (tSVD and FactorGo) in recovering trait-specific loadings of the shared latent genetic factors. We validate our method by recovering the shared traits from GWAS summary data of 2483 phenotypes measured in European-ancestry Pan-UK BioBank individuals. Finally, we apply our methods on the GWAS summary data of 86 phenotypes from the Psychiatric Genomics Consortium. As an illustration of the utility of our approach, we identify shared latent genetic components between bipolar disorder, major depressive disorder and schizophrenia, which are known to be overlapping in their clinical presentation. Our method provides improved understanding of the genetic underpinnings of heterogeneous complex traits, with potential implications for preventive and therapeutic strategies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4043T: Deep learning models of gene regulation and the application in Cross-ancestral PRS in Alzheimer's Disease

Authors:

J-s. Lin[1], C. M. Lakhani[2], A. Das[3], D. Knowles[4], T. Raj[5]; [1]New York Genome Ctr., New York, NY, [2]New York Genome Ctr., Manhattan, NY, [3]Columbia Univ. / New York Genome Ctr., New York, NY, [4]New York Genome Ctr. & Columbia Univ., New York, NY, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Recent Genome-Wide Association Studies (GWAS) have identified several loci related to Alzheimer's Disease (AD), the most prevalent type of dementia. These loci can be used to predict AD via polygenic risk score (PRS). However, most GWAS are European-centric, presenting a challenge in translating the signal to other ethnic backgrounds. We hypothesize that cross-ancestry portability for PRS in AD would improve by identifying of true causal SNPs, as this removes biases from ancestry-based linkage disequilibrium and allele frequencies.
We consolidated omics (Roadmap, cell-type specific signals from microglia and other neuronal cell types) and deep learning (DL) (DeepSEA and Enformer) annotations with higher enrichment and heritability in multiple AD GWAS, resulting in 65 AD-related annotations. We also included 187 baseline (BL) annotations that contain more general information such as MAF and LD structure. We use whole-genome sequencing data from the AD Sequencing Project, consisting of 11.5K samples from European (EUR) populations, 5.8K samples from African American (AFR) populations, and 6.8K samples from Hispanic (AMR) populations after quality control. Notably, the case-control ratios in these cohorts are 54%, 37%, and 27%, respectively. We used the EUR sample to build our reference panel with 16 million SNPs.
We applied functional-informed fine-mapping using PolyFun (Weissbrod et al.) on a recent AD GWAS (Bellenguez 2022) using four different combinations of annotations: (1) none (*NONE*) (2) BL annotations (*BL*) (3) BL+ omics annotations (*BO*), and (4) BL+omics+DL annotations (*ALL*). When using a stringent threshold (PIP > 0.8), we found *NONE* had the most number of fine-mapped variants followed by *ALL* (n_{NONE}=76, n_{BL}=39, n_{BO}=39, n_{ALL}=43). However, the PRS constructed from *NONE* consistently performs the poorest whereas ALL performs the best in every population. In AMR and AFR, we observed a 7% and 4% increase in AUC respectively when comparing *ALL* against *NONE* (AMR: AUC_{ALL} = 0.541, AUC_{NONE}=0.505; AFR: AUC_{ALL}: 0.554, AUC_{NONE}: 0.532). When using a less stringent threshold (PIP>0.5) *NONE* still identified the most fine-mapped variants followed by *ALL*(n_{NONE}=89, n_{BL}=47, n_{BO}=49, n_{ALL}=50). Though *NONE* has slightly better prediction accuracy in EUR compared to *ALL* (AUC_{NONE}=0.576, AUC_{ALL}=0.568), *ALL* has a better prediction accuracy compared to *NONE* in AFR (AUC_{NONE}=0.533, AUC_{ALL}=0.559). Due to these results, we concluded that the addition of annotations as a prior to SUSIE fine-mapping reduces the number of false positive causal variants and can find more causal variants from AD GWAS whose effects generalize across ancestries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4044T: Detection of Large Somatic Copy Number Variation in Normal Tissue Through Single-Cell Multiome Sequencing

Authors:

J. Wang, W. Zhou, R. E. Mills; Univ. of Michigan, Ann Arbor, MI

Abstract:

Somatic copy number variations (CNVs), a predominant category of structural variation, are frequently identified as critical initiators of various cancers. Despite their link to malignancy, somatic CNVs are also present in normal tissues and are thought to play significant roles in the development of other diseases, including multiple birth defects, neurodevelopmental abnormalities, and neuropsychiatric disorders. Recent advancements in single-cell sequencing technologies, like SHARE-seq and 10x Genomics Chromium Single Cell Multiome, enable simultaneous sequencing of open chromatin regions (scATAC) and RNA expression (scRNA) profiles from individual cells, generating numerous datasets across diverse tissues and conditions. While several tools exist for CNV detection using either scATAC-seq or scRNA-seq data independently, there remains a gap in integrating both data modalities which could aid in eliminating biases and enhancing the overall genomic coverage. Here, we present a novel somatic CNV detection tool for single-cell multiome data that utilizes read-depth and employs the Pruned Exact Linear Time (PELT) segmentation method. PELT facilitates genome segmentation on a single-cell basis, offering higher computational efficiency coupled with enhanced accuracy compared to various variations of traditional binary segmentation methods. Our approach first segregates preliminary copy number states using scATAC-seq data. Subsequent verification of putative somatic CNVs is then conducted using the scRNA-seq component to identify regions with commensurate lower expression levels across multiple genes in the region, reducing potential false positives. Our approach has been initially implemented on a publicly available post-mortem healthy brain tissue sample with over 3,000 cells, and we are now extending it to a human cerebral cortex data collection with 45,547 cortical nuclei across six broad developmental time points from fetus to adult. In addition, we are further exploring the integration of multi-platform sequencing technologies to produce a personalized assembly, enabling the construction of long phase blocks for assigning single-cell reads to specific haplotypes and thus identifying allele-specific variations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4045T: Detection of non-additive biallelic effects on trait variation through meta-analysis of 552,042 individuals ⭐

Authors:

F. Lassen[1,2], G. Kalantzis[3], K. Sonehara[4,5,6], S. Namba[4,5,6], B. Hill[7,8], A. Eoli[9,10], S. Hodgson[11], C. Carson[12], W. Zhou[13,14,15], C. Lindgren[16,8,2], A. Verma[17,18,19], H. Heyne[9,10], Y. Okada[20,5,6], d. van heel[21], H. Martin[22], D. Palmer[23,8], Genes & Health Research Team, BRaVa Consortium; [1]Oxford Univ., Oxford, United Kingdom, [2]Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [3]Dept. of Human Genetics, Wellcome Sanger Inst., Hinxton, United Kingdom, [4]Dept. of Genome Informatics, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [5]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Japan, [6]Lab. for Systems Genetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [7]Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom, [8]Nuffield Dept. of Population Hlth., Med. Sci. Div., Univ. of Oxford, Oxford, United Kingdom, [9]Hasso Plattner Inst., Digital Engineering Faculty, Univ. of Potsdam, Potsdam, Germany, [10]Hasso Plattner Inst., Mount Sinai Sch. of Med., New York, NY, [11]Queen Mary Univ. of London, London, United Kingdom, [12]Univ. of Pennsylvania, Philadelphia, PA, [13]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [14]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [15]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [16]Big Data Inst., Oxford Univ., Oxford, United Kingdom, [17]Inst. for BioMed. Informatics, Univ. of Pennsylvania - Perelman Sch. of Med., Philadelphia, PA, [18]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, [19]Dept. of Med., Div. of Translational Med. and Human Genetics, Univ. of Pennsylvania - Perelman Sch. of Med., Philadelphia, PA, [20]Osaka Univ. / The Univ. Tokyo / RIKENDept. of Genome Informatics, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [21]Blizard Inst., Queen Mary Univ. of London, London, United Kingdom, [22]Dept. of Human Genetics, Wellcome Sanger Inst., Cambridge, United Kingdom, [23]Univ. of Oxford, Big Data Inst., Oxford, United Kingdom

Abstract:

In rare genetic association testing, group-based tests are used to increase power by aggregating variants based on functional annotation. However, these tests tend to assume an additive contribution of genetic effects to trait variation. The phenotypic impact of rare recessive effects is largely unexplored, and moreover, when it has been examined, almost always excludes compound heterozygous (CH) variation, the inclusion of which has the potential to increase our power to detect such effects.

To increase power to test for rare recessive effects, we identified CH variant by statistically phasing genetic variants down to a MAF of 0.001% in 5 biobanks: UK Biobank (*N* = 410,030),

Genomics England (*N* = 71,514), Genes & Health (*N* = 39,152), BioMe (*N* = 31,701) and BioBank Japan (*N* = 11,462). We identified CH variation in 465,290 Europeans, 53,321 South/Central Asians, 14,258 Africans, 13,837 East Asians and 5,336 Americans/Admixed.
9,948 genes were disrupted by biallelic putative loss-of-function (LOFTEE HC), damaging missense (REVEL > 0.773 or CADD > 28.1) and splicing variants (SpliceAI > 0.50) at MAF < 5%. Across these genes, 68,676 and 187,206 individuals were CH and homozygous carriers, respectively. Ratio of homozygous to CH variant count in each cohort reflected varying degrees of autozygosity: Genes & Health, a consanguineous cohort, displayed a higher ratio (5.3) than well mixed populations such as the UK Biobank (2.5).
To investigate the phenotypic consequences of biallelic gene inactivation, we performed federated gene-based recessive burden tests of 32 binary and 10 quantitative traits across biobanks. We identified 21 significant (experiment-wise $P < 4.53 \times 10^{-7}$) gene-trait recessive associations across 8 unique genes and 9 traits. Our meta-analysis replicated established recessive associations, such as damaging variants in *FLG* associated with increased risk of asthma ($P = 7.48 \times 10^{-72}$), and identified potentially novel relationships such as increased risk of chronic obstructive pulmonary disease associated with damaging biallelic variation in *ODAD1* ($P = 1.67 \times 10^{-7}$), and biallelic variation in *HBB* associated with total cholesterol ($P = 2.35 \times 10^{-8}$).
This work is part of the **B**iobank **Ra**re **Va**riant analysis consortium recessive working group. We are developing tools for applications, including extracting biallelic genotypes and testing for non-additive associations, facilitating systematic investigation and meta-analysis of recessive associations across biobanks. Our work demonstrates the utility of phasing large-scale cohorts to identify phenome-wide recessive associations, and demonstrates the power gained through global collaboration.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4046T: Differentiating germline and somatic sources of apparent homozygosity

Authors:

J. Valdebenito[1], P. Scheet[2], J. Wong[3], P. Wei[2], Y. Jakubek[4]; [1]The Univ. of Texas MD Anderson Cancer Ctr. UTHlth.Houston Graduate Sch. of BioMed. Sci., Houston, TX, [2]MD Anderson Cancer Ctr., Houston, TX, [3]MD Anderson Cancer Ctr, Houston, TX, [4]Univ. of Kentucky, Lexington, KY

Abstract:

Runs of homozygosity (ROH) are typically short regions inherent to a normal genome in outbred populations, and more frequent in inbred populations. Copy-neutral loss of heterozygosity (cn-LOH) is generally a larger acquired event that may be present in cancer genomes. When cnLOH exists at a high mutant cell fraction (MCF; e.g., >90%), it will be indistinguishable from ROH on the basis of called genotypes, presenting as 'apparent homozygosity'. However, since cn-LOH results in a loss of the heterozygotes, trace signals from the "lost" genotypes may confirm the source of the apparent homozygosity and thus detect this form of mosaic chromosomal alteration (mCA). Typically, mCAs, including cn-LOH, are detected via B-allele frequency (BAF) data from SNP arrays. Standard tools can be applied when the MCF is intermediate (via segmentation methods) or low (using haplotypes). At high MCF, whereas duplications and deletions may be detected by a change in total intensity, cn-LOH becomes indistinguishable from ROH. To differentiate (acquired) high-MCF cn-LOH from (germline) ROH, we assume MCF is less than unity and propose testing the BAF differences between "likely heterozygotes" and "likely homozygotes". We utilize population allele frequencies (linkage equilibrium; LE) and patterns of linkage disequilibrium (LD) to inform group membership for the markers and perform t-tests of the BAFs, accommodating uncertainty in group membership (Bauer et al., Psychometrika, 2021). We applied this hypothesis-testing-based framework to simulated and observed data from the *Mano a Mano* Mexican American Cohort Study. Our simulations consisted of events of different lengths, chromosomes, and MCFs (particularly for cn-LOH). We expect a haplotype-aware method to perform well in differentiating shorter events, as the LD information of nearby markers will better impute the lost information due to cn-LOH and help our test seek significant differences in the BAF information. As events become larger, our haplotype-aware method will behave similarly to when using population allele frequencies to infer our markers' group membership, as LD signals from closely neighbor markers become distant and diluted. Our technique may be useful in detecting modest-sized cn-LOH mCAs in population data, particularly where there exists an excess of ROH, and may help to determine germline vs. acquired homozygosity in genetic testing scenarios.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4047T: Discovery of core genes for type 1 diabetes via genome-wide aggregation of *trans*-effects using proteomics studies ⭐

Authors:

X. Zhou[1], A. Iakovliev[1], S. McGurnaghan[2], P. McKeigue[1], A. Spiliopoulou[1], H. Colhoun[2]; [1]Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract:

Background/Objectives: Type 1 diabetes (T1D) is caused by autoimmune destruction of the insulin-producing beta cells of the pancreas. Our objective was to identify core genes for T1D, using genome-wide aggregated trans-effects (GATE) analysis.Methods: We used pQTL summary statistics from the UK Biobank proteomic study (2923 proteins on the Olink platform measured in 54306 individuals) and the deCODE proteomic study (4719 proteins on the Somalogic platform measured in 35559 individuals) to compute GATE scores in a Scottish dataset comprising 4964 T1D cases from the SDRNT1BIO and 7497 controls from the Generation Scotland study, and 1864 cases and 449572 non-cases in the UK Biobank cohort. We tested GATE scores for each protein for association with T1D. Criteria for a putative core gene were: association of GATE score with T1D at $p < 10^{-6}$, effective number of trans-QTLs > 5, and GATE score not strongly correlated with scores for other genes. Putative core genes were validated as described below. Results: We identified 25 putative core genes in the Scottish dataset and two additional genes in the UKBB cohort. The strongest association was with *PDCD1* (standardized log odds ratio 0.21, $p = 10^{-24}$). Of the 27 putative core genes, four are validated by association of T1D with a cis-pQTL or with SNPs within 200 kb of the transcription site, 11 by association of T1D with circulating protein level, and 8 by experimental perturbation in a mouse model of autoimmune diabetes. Further validation of *PDCD1* as a causal gene is that drugs targeting the encoded receptor, used in cancer immunotherapy, commonly induce T1D. Of these 27 genes, 18 have been characterized as in the immune system. Four - *PDCD1*, *CD5*, *TIGIT*, *LAG3* -- encode immune checkpoint receptors expressed on T cells that are exploited by cancer cells to escape the immune response, four -- *CXCL9*, *CXCL10*, *CCL19*, *CXCL11* -- encode chemokines that control the migration of immune cells, and two - *NCR1*, *KLRB1* -- are expressed on natural killer cells. Conclusion: These results support the “omnigenic” sparse effector hypothesis for the genetic architecture of complex traits, and identify immune checkpoints and natural killer cell receptors as possible targets for drugs to prevent or reverse T1D.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4050T: Dissecting cross-population polygenic heterogeneity of multimorbidity across respiratory & cardiometabolic diseases

Authors:

Y. Yamamoto[1], Y. Shirai[1], K. Sonehara[2], S. Namba[2], K. Yamamoto[1], R. Edahiro[1], K. Suzuki[1], T. Ojima[1], BioBank Japan Project, Y. Takeda[3], K. Matsuda[4], A. Kumanogoh[3], T. Yamauchi[5], T. Kadowaki[5], Y. Okada[2]; [1]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]Dept. of Genome Informatics, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo,

Japan, [3]Dept. of Respiratory Med. and Clinical Immunology, Osaka Univ. Graduate Sch. of Med., Suita, Japan, [4]Lab. of Clinical Genome Sequencing, Dept. of Computational Biology and Med. Sci., Graduate Sch. of Frontier Sci., the Univ. of Tokyo, Tokyo, Japan, [5]Dept. of Diabetes and Metabolic Diseases, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan

Abstract:

Biological mechanisms underlying multimorbidity remain elusive. To dissect the polygenic heterogeneity of multimorbidity in twelve complex traits across populations, we leveraged the biobank resources of genome-wide association studies (GWAS) for 232,987 East Asians (BioBank Japan) and 751,051 Europeans (UK Biobank and FinnGen). Cross-trait analysis for respiratory and cardiometabolic diseases, rheumatoid arthritis, and smoking identified negative genetic correlations between respiratory and cardiometabolic diseases in East Asians, opposite from the positive associations in Europeans. Dissecting the whole genome into local regions with linkage disequilibrium (LD) revealed the directional heterogeneity of local genetic correlations in respiratory and cardiometabolic diseases across populations. Even in multimorbid individuals, Bayesian polygenic risk score (PRS) for asthma and dyslipidemia correlated negatively. Association analysis between Bayesian genome-wide PRS and metabolome data from 51,612 East Asians and 245,349 Europeans revealed the negative association of asthma PRS with circulating lipid and metabolite biomarkers, supporting the genetic causality of asthma to decrease the genetic risk of dyslipidemia in EAS population. Cross-trait pathway enrichment analysis using Bayesian PRS detected the biology responsible for the heterogeneity of multimorbidity between respiratory and cardiometabolic diseases in East Asians, especially among the pair of asthma and dyslipidemia. One of the five pathways identified in the phenotype combination regulated the lipid metabolism via peroxisome proliferator-activated receptor alpha (PPARα). To acquire further insights into the genetics underlying the heterogeneity of multimorbidity about asthma and dyslipidemia, we conducted cell type enrichment analysis using the lung single-cell RNA-sequencing dataset. Cross-population analysis identified the differential functional enrichment of cell types in each phenotype, whose patterns corresponded to the population-specific clinical characteristics of traits. The PPARα pathway showed the heterogeneous specificity of cell types across phenotypes (T cells in asthma and macrophages in dyslipidemia). Our study highlights the heterogeneous contribution of polygenicity to the immune and metabolic dysfunction in multimorbidity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4051T: Distinct Genetic Architecture in the Tails of Complex Traits

Authors:

P. O'Reilly[1], H. Wu[2], C. Hoggart[3], T. Souaiaia[4]; [1]Icahn Sch. of Med., Mount Sinai, NYC, New York, NY, [23]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]SUNY Downstate Med. Ctr., NY, NY

Abstract:

Genetic architecture describes the number, effect sizes and allele frequencies of genetic variants affecting a trait. While genetic architecture has been investigated extensively at the trait-level, little is known about how genetic architecture varies among individuals across the trait continuum. Here we reveal striking departures from polygenicity in the distribution tails of a range of human traits (e.g. in individuals with high LDL, low/high sitting height, early age of menopause) and implicate stabilising selection - pervasive in humans - as a chief cause. Stabilising selection ensures that alleles with large effects are rare, meaning that trait tails may be enriched for rare alleles of large effect and have correspondingly reduced polygenicity. Applying statistical tests to infer trait tail architecture from polygenic risk scores (PRSs) in population data and from sibling trait-similarity in family data, leveraging the large UK Biobank and All of US datasets, we observe consistent patterns of reduced polygenicity in trait tails between tests, cohorts, and ancestries. Moreover, trait tails that exhibit reduced polygenicity are associated with a larger number of rare variants, more illnesses, more stillbirths, and evidence of ongoing selection as inferred by models of reproductive success. Our study provides novel insights into how genetic architecture varies across the trait continuum, demonstrating a pervasive pattern of reduced polygenicity in trait tails that limits PRS accuracy, and highlights the key role of rare variants, specifically within the trait tails, where disease manifests.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4052T: Effect of Polygenic Scores for Serious Mental Illness on Cell Type-Specific Gene Expression in The Brain

Authors:

R. Bercovitch[1], S. Venkatesh[1], D. Mathur[1], M. Koutrouli[2], J. Bendl[3], D. Lee[1], S. Kleopoulos[1], G. Hoffman[1], J. Fullard[1], G. Voloudakis[1], P. Roussos[1], M. Anyfantakis[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Univ. of Copenhagen, København, Denmark, [3]Icahn Sch. of Med. at Mount Sinai, New York City, NY

Abstract:

Background: Alterations in the transcriptional landscape within various brain regions have been observed in serious mental illness (SMI). To enhance our knowledge of the putative causal

processes driving the progression of SMIs, we aimed to determine whether germline genetic variation influences cell type-specific gene expression (CTSGE) changes.
Methods: We investigated the association between polygenic scores (PGS) for SMIs and CTSGE in up to 923 individuals of European ancestry, using post-mortem data from the PsychAD cohort, which comprises three independent brain tissue repositories: Mount Sinai School of Medicine (MSSM), the Human Brain Collection Core at the National Institute of Mental Health (HBCC), and the Rush Alzheimer's Disease Center (RADC). Linear models were used to assess the effect of SMI PGS on gene expression, adjusting for batch effects, genetic ancestry PCs, and clinical variables. Gene set enrichment analysis (GSEA) was conducted to identify pathways enriched in gene sets associated with Schizophrenia (SCZ) PGS. To capture *cis* vs. *trans* regulation, we also compared z-scores (per gene) from our regression analysis to z-scores from a SCZ transcriptome-wide association study (TWAS).
Results: The association between gene expression and SCZ PGS revealed the highest number of significant gene hits. Notably, *DNAJB1* was downregulated in excitatory neurons (EN) in both controls and all individuals (beta = -0.47, FDR q = 3.97×10^{-4}; beta = -0.26, FDR q = 6.80×10^{-6}, respectively). The majority of significant gene expressions associated with SCZ PGS were observed in oligodendrocytes (OGD), followed by inhibitory neuron and EN populations. OGDs exhibited the largest and most robust set of enrichments, demonstrating significance in pyrimidine metabolism, regulation of immune effector processes, and regulation of RAS protein signal transduction. We observed a low correlation (controls: avg. $r = 6.56 \times 10^{-3}$, all: avg. r = 0.013) between our approach and TWAS. Other SMI PGS and gene expression regression results showed fewer, though notable, significant gene hits.
Conclusions: Genetic predisposition for SMI is significantly associated with cell type-specific alterations in gene expression that cannot be captured by *cis*-regulation of gene expression.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4053T: Efficient genealogical compression for biobank-scale datasets

Authors:

L. O'Connor[1], N. Mancuso[2]; [1]Harvard Med. Sch., Brookline, MA, [2]Univ. of Southern California, Los Angeles, CA

Abstract:

Biobank-scale genetic datasets pose a computational challenge due to their massive size. A potential solution is to leverage haplotype sharing between genealogical relatives, in particular

by inferring an ancestral recombination graph (ARG). However, two challenges make genealogical compression unappealing in practice. First, most statistical applications rely on genotype matrix operations that are not easily replicated on the ARG, particularly matrix-by-vector multiplication. Second, existing methods for ARG inference achieve worse than expected compression ratios on real genetic datasets, either because they rely on heuristics or because they have objectives other than parsimony.

We make two contributions. First, we define the linear ARG, a modified ARG in which each node (or "pseudohaplotype") is a linear combination of its ancestors. The weighted adjacency matrix of the linear ARG can be substituted directly into expressions involving the genotype matrix; in particular, the matrix-vector product X'y is computed by solving an extremely sparse, triangular system of equations. Second, we introduce a method to infer the linear ARG directly from genotype data. This method involves a purely algebraic step, leveraging linearity to obtain an initial graph, followed by a clique elimination step that sparsifies the graph by inferring ancestral recombination events. Decompression involves inverting a triangular matrix; in most applications, decompression will be unnecessary.

We applied our genealogical compression method to sequencing data from 1000 Genomes (N=5k phased haplotypes) and to imputed array data from UK Biobank (N=394k diploid samples). We confirmed that decompression recovers the genotype matrix exactly. Compared with a sparse, non-genealogical representation, the average compression ratio was 13x in 1000 Genomes and 72x in UK Biobank (based on chromosome 1). A linear ARG for all of UK Biobank should be under 64GB in size. In a downsampling experiment, this number scales with N^0.7. Moreover, our current compression ratio is probably limited by the use of unphased imputed input data, and with phased WGS data, even better ratios might be achieved.

We envision that linear ARGs will significantly reduce the computational challenge of analyzing biobank-scale genetic datasets. It will be beneficial in memory intensive applications, particularly those involving the GRM, and also in simple applications, like computing a polygenic score, where the rate limiting step is simply reading from a hard drive. We plan to release open-source software and to make precomputed linear ARGs for UK Biobank and AllOfUs available to all approved researchers.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4054T: Efficient saddlepoint approximation in generalized linear model for large-scale genetic association studies

Authors:

Y. Zhao, Z. Zhang, M. Carr, N. Mancuso; Univ. of Southern California, Los Angeles, CA

Abstract:

<Thousands of phenotypes collected in large biobanks when combined with genetics enables identifying genetic variants associated with human diseases and complex traits. Many of these phenotypes are non-continuous including binary traits, count outcomes, and molecular traits such as gene expression counts. The standard hypothesis test after fitting regression models between a sparse outcome (e.g., unbalanced case-control, zeros in count data) and rare genetic variants can inflate type I error due to poor approximation by normal distribution at the tail. Recent work described the saddlepoint approximation (SPA) to address this limitation for binary traits, with extensions to survival data. This approach, while powerful, has taken a case-by-case basis to derive inferential tools and models, despite a singular underlying theory. Here, we propose Giddyup, a saddlepoint approximation framework for generalized linear models (GLM). Giddyup leverages differentiable programming to model and optimize through moment generating functions of exponential families. We implemented Giddyup in JAX, which enables computation across CPUs, GPUs, or TPUs.To benchmark Giddyup, we performed simulations for logistic and Poisson models assuming rare genetic variants with an MAF of 0.01. For the logistic model, we showed the type I error of Giddyup is well calibrated for varying total sample size and case-control ratios, i.e., balanced (prevalence 50%), moderately unbalanced (prevalence 10%), and extremely unbalanced (prevalence 0.2%). Given a large sample size of 10,000 and genotype odds ratio of 2.23, we observed the statistical power increased from 0.55 to 0.98 compared extremely with a moderate unbalanced case-control scenario. When the case-control is extremely unbalanced, the statistical power increased from 0.55 to 0.82 when the sample size increased from 10,000 to 20,000 given a genotype odds ratio of 2.23. When the sample size is as small as 1,000, power reached 0.71 when the genotype log odds ratio reached 12.18. For the Poisson model, we simulated count data with a target mean count of 0.1 for a genotype rate ratio of 1.65 and MAF of 0.01. The type I error of Giddyup is calibrated across varying sample sizes. The statistical power increased from 0.1 to 0.76 when the sample size increased from 1,000 to 10,000. When the mean count was reduced to 0.02, the statistical power maintained at 0.76 when the sample size increased to 10,000 and the genotype rate ratio increased to 2.72. Taken together, Giddyup provides a flexible and robust approach for genetic association tests using saddlepoint approximation in GLM models.>

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4055T: Enhancing non-linear TWAS performance via trait imputation with applications to Alzheimer's disease

Authors:

W. Pan[1], R. He[2], **J. Ren**[3], M. M. Malakhov[1]; [1]Univ. of Minnesota, MINNEAPOLIS, MN, [2]Univ. of Minnesota, Twin Cities, Minneapolis, MN, [3]Univ. of Minnesota, Minneapolis, MN

Abstract:

Genome-wide association studies (GWAS) performed on large biobank datasets have identified numerous genetic loci associated with Alzheimer's disease. However, the younger demographic of biobank participants relative to the typical AD age of onset has resulted in an insufficient number of AD cases, limiting the statistical power of GWAS and any downstream analyses. To mitigate this limitation, several trait imputation methods have been proposed to impute the expected future AD status of individuals who may not have yet developed the disease. This paper explores the use of imputed AD status in nonlinear transcriptome-wide association studies (TWAS) to identify genes and proteins whose genetically regulated expression is associated with AD risk. In particular, we considered the TWAS method DeLIVR, which utilizes deep learning to model the nonlinear effects of expression on disease. We trained transcriptome and proteome imputation models for DeLIVR on data from the Genotype-Tissue Expression (GTEx) Project and the UK Biobank (UKB), respectively, with imputed AD status in UKB participants as the outcome. Next, we performed hypothesis testing for the DeLIVR models on diagnosed AD outcomes using the Alzheimer's Disease Sequencing Project (ADSP) dataset. Our results demonstrate that nonlinear TWAS trained with imputed AD outcomes in UKB successfully identified known and putative AD risk genes, and that training with imputed outcomes yielded more discoveries than training with diagnosed cases alone. Notably, we found that different AD imputation methods yield complementary results in association analyses, suggesting that investigations relying on a single imputation method may miss important pathways involved in neurodegeneration.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4056T: Epistatic SNPs land in regulatory regions and possess a similar risk as GWAS SNPs in a common liver disease, MASLD

Authors:

A. Kar[1], B. Fu[1], M. G. Sukhatme[1], S. T. Lee[1], R. Border[1], K. M. Garske[1], Z. Chen[1], D. Kaminska[2], V. Männistö[2], J. Pihlajamäki[2], S. Sankararaman[1], P. Pajukanta[1]; [1]UCLA, Los Angeles, CA, [2]Univ. of Eastern Finland, Kuopio, Finland

Abstract:

Epistasis, the interactive genetic effect on phenotypes (GxG), has been suggested to contribute to disease heritability; however, its effect sizes and mechanisms are poorly understood in humans. Here, we hypothesized that epistasis may involve epigenetic regulatory regions to confer disease risk. To this end, we used the QuadKAST marginal epistasis tool to identify 31 pairs of biomarkers from the UK Biobank (UKB) and regional SNP sets with epistasis (p<0.05) in UKB. Upon functionally assessing the 23 sets with >10 SNPs, including tight LD proxies (r^2>0.95), we found that 74% of the sets are significantly enriched (p_{perm}10000<0.05) for SNPs with strong evidence of regulatory importance (regulomeDB score≥1f) compared to same-sized random sets of significant ($p<5\times10^{-8}$) SNPs from a genome-wide association study (GWAS) for the phenotype. As we detected epistasis for cardiometabolic and liver biomarkers, including alanine aminotransferase (ALT) and lipoprotein A, we further investigated the identified sets for enrichment of regulation in liver and subcutaneous adipose tissue (SAT). We determined 61% and 43% of the sets to be enriched (p_{perm}10000<0.05) for significant *cis*-eQTL SNPs (FDR<0.1) in SAT and liver bulk RNA-seq data, respectively, and 22%, 17%, and 30% of SNPs landing in open chromatin regions from preadipocyte, adipocyte, and bulk liver ATAC-seq data (p_{perm}10000<0.05), taken together corroborating the regulatory nature of regions involved in epistasis. Subsequently, we searched for cross-chromosomal epistasis on disease risk, focusing on 7 most robustly replicated GWAS SNPs for metabolic dysfunction-associated steatosis liver disease (MASLD), which alter function, expression, or protein production, and include SNPs in 4 lipid droplet-related genes, *PNPLA3*, *MBOAT7*, *TM6SF2*, and *HSD17B13*. We identified 3 pairwise interactions on MASLD in UKB, between rs738409 and rs58542926 (odds ratio (OR)=1.12; $p=1.8\times10^{-4}$), rs738409 and rs72613567 (OR=0.93; $p=1.2\times10^{-4}$), and rs641738 and rs4374383 (OR=0.97; p=0.03), which all replicated as significant (p_{adj}<0.05) for the MASLD biomarker ALT. Of note, we observed that these epistatic effect magnitudes are comparable to the additive effects of the 94 MASLD GWAS SNPs, ranking among the 95th, 75th, and 7th percentiles of the SNP effects. We also found 9 nominally significant interactions (p<0.05) between these 5 SNPs and the 59 other independent MASLD GWAS SNPs (r^2<0.001), similarly ranking between the 8th and 90th percentiles of effects from the 59 SNPs. In summary, epistatic SNPS are significantly enriched for regulatory regions, and the GxGs of MASLD confer the same amount of risk as MASLD GWAS SNPs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4057T: Proteome-wide Mendelian randomisation of 2,167 disease outcomes in UK Biobank and FinnGen

Authors:

J. Burton[1], D. Gadd[2], J. Fisher[3], Z. Zheng[4], Z. Kuncheva[5], F. Finngen[6], G. Hemani[1], B. Sun[7], M. Kurki[8], T. Gaunt[1], D. Baird[7], C-Y. Chen[7]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Univ. of Edinburgh, Ediburgh, United Kingdom, [3]Harvard Univ., Cambridge, MA, [4]Broad Inst. of MIT and Harvard, Boston, MA, [5]Optima Partners Ltd, London, United Kingdom, [6]FinnGen, Helsinki, Finland, [7]Biogen, Cambridge, MA, [8]Broad Inst., Boston, MA

Abstract:

Protein quantitative trait loci (pQTL) enable prediction of causal effects of variation in protein levels within serum. Here, we performed a comprehensive Mendelian randomization (MR) and colocalization analysis to investigate the causal effects of plasma protein levels on a range of disease outcomes, by leveraging data from two large biobanks, UK Biobank (UKB) and FinnGen. We performed discovery two-sample MR and colocalization analysis using 3,580 cis- and 15,517 trans-pQTLs for 1,102 proteins from the UKB Pharma Proteomics Project (UKB-PPP) as exposures (N = 34,557) against 2,167 disease outcome genome-wide association studies (GWAS) from FinnGen (N = 453,733). Cis-pQTLs were, on average, stronger genetic instruments (higher f-statistics) than trans-pQTLs, and inclusion of trans-pQTLs did not substantially improve overall instrument strength. A total of 246 proteins were instrumented with a single SNP, with the remaining 676 instrumented with multiple SNPs (median = 4). To account for allelic heterogeneity, when multiple cis-pQTL instruments were available for a protein, the Wald ratio estimates were pooled using the Inverse Variance Weighted method (IVW) and our colocalization procedure used coloc-SuSiE which conducts fine-mapping to perform coloc at single variant resolution (PP4 for each credible set). Of the 141 causal effects identified with cis-pQTL MR, 42 showed colocalization evidence (PP4>0.7 for any credible set), representing 9 proteins and 40 diseases. Of the 40 diseases, the majority (N=38) only colocalized with a single protein. The diseases that colocalized with the most proteins were wet form age-related macular degeneration and mucosal proctocolitis (N=2). The ICD10 code with the highest proportion of colocalized proteins ($N_{proteins}$ = 3, $N_{diseases}$ = 11) represent diseases of the digestive system in chapter 11. Of the 9 proteins, 5 colocalised with a single disease, and the remaining colocalised with multiple proteins to varying degrees (mean = 16.5, ranging between 2 and 53). APOE was the most pleiotropic protein colocalising with diseases spanning 8 different ICD10 categories. To replicate our findings, we performed the reciprocal MR analysis using FinnGen pQTL to instrument proteins (N = 1,732) and 681 UKB disease outcome GWAS (N = 420,531). Nine causal effects were replicated, of which 8 showed evidence of colocalization in the discovery analysis and all 8 shared the same direction of effect between discovery and replication (R^2 = 0.92, b = 0.91). Our study demonstrates the potential of large pQTL datasets to generate new understanding of disease mechanisms and identify potential therapeutic and disease prevention targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4058T: Estimating the environmental effect of social determinants of health and race/ethnicity on human trait and disease outcomes

Authors:

O. Risse-Adams[1], N. Sinnott-Armstrong[2], S. Musharoff[1]; [1]Cornell Univ., Ithaca, NY, [2]Fred Hutch Cancer Ctr., Seattle, WA

Abstract:

Differences in traits and disease outcomes in humans are caused by genetic and environmental factors as well as their interactions. To assess the contribution of non-genetic factors, specifically Social Determinants of Health (SDOH), we compare traits with shared genetic effects between individuals experiencing health disparities. For this pilot study, we focused on quantitative trait data related to known health disparities, including BMI, systolic blood pressure, heart rate, heart rhythm, urate levels, and waist circumference measures. We analyzed a random multiethnic subset of 10,000 participants in the All of Us Research Program, self-identifying with a range of races/ethnicities.
Initially, we estimated genetic and environmental variance using GCTA and adjusted for gender, age, and genetic principal components (PCs) 1-3. We then incorporate data from the SDOH survey, answered by ~25% of All of Us participants, by performing principal component analysis on the survey results. We find that SDOH PC1 captures individuals' sentiments about their neighborhood (e.g. safety) and general living conditions (e.g. presence of mold). After adjusting for SDOH PC1, the estimates of genetic (7.8-fold higher) and environmental variance (4.3-fold lower) were substantially different for systolic blood pressure relative to no SDOH PC adjustment, despite a similar total trait variance, supporting the notion that most differences in SDOH result in changes to the environmental variance of health disparities. This trend was still present but less pronounced for both waist circumference (3.0-fold higher genetic and 1.3-fold lower environmental variance) and heart rate (1.8-fold higher genetic and 1.6-fold lower environmental variance). When adjusting for all SDOH PCs (1-20) within a subset of European-ancestry individuals, these changes in estimates persisted for urate and waist circumference (6.8-fold lower genetic and 1.1-fold lower environmental variance, respectively, relative to no SDOH adjustment). We repeated analyses and compared results for diverse race/ethnic subsets. In conclusion, we found that surveys of SDOH provide an explicit, and orthogonal, source of information from those historically included in genomic studies. We find that stratifying

individuals by their lived social determinants of health indicates substantial contributions to environmental variance among traits known to exacerbate health disparities. By jointly analyzing survey information of SDOH and genomic data, we can improve the accuracy and interpretability of genetic associations, particularly in diverse populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4059T: Evaluating common practices for Mendelian randomization studies in East Asians: current practices, challenges and recommendations

Authors:

J. Ong[1,2], A-Y. Lim[1,2], S. MacGregor[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Univ. of Queensland, Brisbane, Australia

Abstract:

Recent advances in Mendelian randomization (MR) studies, facilitated by large-scale biobanks and genome-wide association studies (GWAS) datasets, have provided significant insights into causal inference for complex diseases, pharmacogenetics, and intervention strategies. However, unlike the well-defined guidelines for trans-ethnic polygenic risk prediction frameworks, standards for cross-ancestry MR studies, particularly for disease endpoints in non-European populations, are limited. This gap has led to variability and inconsistency in reporting MR findings, such as the selection of instrumental variables (IVs) when GWAS data in the target population is limited, and sensitivity analyses for addressing violation of IV assumptions. Consequently, cross-ancestry MR studies suffer from poor interpretability, exacerbated by the lack of consideration for potential differences in genetic architecture between local ancestries. To address this, we conducted a scoping review of MR studies focusing on disease inference in East Asians. We included English-language MR studies published between 2014 and 2023, resulting in a comprehensive collection of 350 studies. These studies were categorized into three types based on the origin of their instruments: Type-I (European-derived instruments applied to East Asian populations), Type-II (European-derived instruments validated with East Asian datasets), and Type-III (East Asian-derived instruments). Additional information appraised includes target populations, sample sizes, data availability, individual versus summary-level data approaches, and types of sensitivity analyses conducted.
In the meeting, we will provide an extensive review of current practices in East Asian MR literature. Using scenario-based approaches and simulation studies, we will illustrate how ancestry-specific genetic effects (including allele frequency discrepancies) impact MR inferences

across the three study types. Based on these findings, we will propose enhanced sensitivity analyses to improve the robustness of MR results. Finally, we will offer best-practice recommendations to guide MR researchers in resource-constrained settings, further highlighting the importance of expanding biobank resources to better capture the genetic and cultural diversity of East Asian populations. Our work underscores the need for more refined and validated MR practices, advocating the need to develop additional resources to strengthen the reliability and interpretability of MR studies in diverse populations.

---PAGEBREAK---

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4060T: Exploring platform validation of modern microarrays with applications in polygenic risk scoring at-scale in a large institutional biobank.

Authors:

J. Shortt[1,2], M. Fisher[3], M. Lin[2], M. Brasher[3], N. Rafaels[4], C. Sanchez Wild[1], N. Pozdeyev[5], T. Brunetti[2], K. Crooks[2], C. Gignoux[2], Colorado Center for Personalized Medicine; [1]Colorado Ctr. for Personalized Med., Aurora, CO, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Univ. of Colorado Anschutz, Aurora, CO, [4]Univ. of Colorado Sch. of Med., Aurora, CO, [5]Univ. of Colorado, Aurora, CO

Abstract:

Polygenic risk scores (PRS) use the cumulative effect of dozens to thousands of single nucleotide polymorphisms (SNPs) to predict risk of complex disease and have the potential to transform personalized healthcare. Among the challenges facing clinical implementation of PRS at population scales is the known confounding of PRS distributions because of population stratification. These differences can be corrected using a number of proposed PRS recalibration methods, many of which depend on some form of individual-level genetic ancestry information, but this poses the additional implementation challenge of generating robust genetic similarity information for each individual. Here, we present our approach to validation of the Global Diversity Array (GDA) in a large institutional biobank to allow for improved genetic similarity PRS calibration across individuals. Following extensive quality control including comparisons to gnomAD and GIAB, we selected 1,151,179 variants from the GDA for platform validation, which can serve as a pool of high-quality variants for PRS model selection and for genetic ancestry-based PRS calibration. Across 24 control samples with known genotypes assayed more than 75

times each, these variants have high allelic concordance (mean concordance=99.99%) compared to variants that were removed during quality control (n=167629, mean concordance=99.04%) (p-value < 2.2e-16). We genotyped 48634 individuals on the GDA and projected them onto principal components (PCs) created with a linkage-pruned subset of high-quality GDA variants in a diverse reference panel (1000G and HGDP) to evaluate robustness to variant missingness. We assess whether these PCs are sufficiently robust to recalibrate PRS for accurate clinical risk assessment for all individuals regardless of genetic ancestry. Importantly, this approach to genotype validation supports efforts to recalibrate PRS along a genetic ancestry continuum without the need to assign individuals to distinct genetic similarity groups. Such efforts are vital in ensuring that advances in personalized medicine are implemented with equity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4061T: FastGxC: A fast and powerful statistical method for mapping context-specific regulatory variants in bulk and scRNA-seq studies ⭐

Authors:

L. Krockenberger[1], A. Lu[1], M. Thompson[2], C. Liu[1], G. Gordon[3], A. Dahl[4], C. Ye[3], N. Zaitlen[1], B. Balliu[1]; [1]UCLA, Los Angeles, CA, [2]Ctr. for Genomic Regulation (CRG), Barcelona, Spain, [3]UCSF, San Francisco, CA, [4]UChicago, Chicago, IL

Abstract:

Context-specific cis expression quantitative trait loci (sp-eQTLs) underlie genetic risk factors for complex diseases. To limit cost and experimental heterogeneity, bulk and single cell eQTL studies often gather multiple samples across contexts (tissues and cell types) from the same donors. Linear mixed models (LMMs) are a natural analysis choice for such studies because they model intra-individual residual correlation and directly identify sp-eQTLs through a genotype-by-context (GxC) interaction term. However, LMM-GxC are computationally infeasible for eQTL studies. Hence, researchers rely on context-by-context eQTL mapping with post hoc examination of summary statistics to identify sp-eQTLs. While fast, these approaches are underpowered, as they do not account for shared noise within individuals, and rely on ad hoc definitions of sp-eQTLs, e.g. presence of eQTL in a single context. These shortcomings hinder characterization of sp-eQTLs and interpretation of disease-associated variants. Here, we introduce FastGxC, a fast and powerful method for sp-eQTL mapping in studies with repeated sampling. FastGxC removes intra-individual correlation by decomposing gene expression per individual into context-shared

and context-specific components and estimates genetic effects on these components separately using ultra-fast implementations of linear regression. FastGxC eQTL effect sizes are a re-parameterization of LMM-GxC effect size estimates, enabling direct mapping of sp-eQTLs without post hoc analyses. FastGxC provides a global test to identify variants with heterogeneous effect sizes across contexts and a marginal test to identify the context(s) driving this heterogeneity. Its output integrates seamlessly with existing methods like mash and METASOFT, to improve the statistical power of eQTL mapping. In simulations, FastGxC is orders of magnitude faster and more powerful than previous approaches. We apply FastGxC to bulk multi-tissue RNA-seq data from GTEx (N=948) and PBMC single-cell RNA-seq data from the CLUES (N=234) and OneK1K (N=982) cohorts, generating a comprehensive tissue- and PBMC cell type-specific cis-eQTL map. FastGxC identifies eQTLs for 70.2% and 36.1% of tested genes across tissues and PBMC cell types, 50.6% and 22.5% of which show specificity across tissues and cell types, respectively. FastGxC sp-eQTL are enriched in context-specific genomic annotations and provide a three-fold increase in precision to identify relevant contexts for GWAS variants across 138 complex traits. In summary, FastGxC sp-eQTLs can be used to understand context-specific gene regulatory mechanisms underlying complex human diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4062T: Finding Structure In Gene Regulatory Networks: A Graph Neural Network Algorithm for Hierarchical Community Detection

Authors:

J. Kvamme, A. Fu; Univ. of Idaho, Moscow, ID

Abstract:

Gene regulatory networks (GRNs) are a class of biological networks which focus on the regulation of gene expression where nodes represent genes and edges represent interactions between genes or other biological elements. Causal GRNs are directed networks where edges represent valid causal relationships. Unlike traditional GRNs, causal GRNs aim to identify the direction of the regulatory relationships between genes. These networks provide intuitive models of how genes or genetic factors influence the activation or suppression of others, highlighting the mechanisms driving gene expression and genetic disease. However, current methods for learning causal GRNs typically handle only small networks with just a few genes and variants, ignoring the global network regulating gene expression
One solution is to simplify a GRN by partitioning genes into a hierarchy which maps genes into a new set of community representations, reducing the number of nodes in the network. To tackle

this problem, we introduce Hierarchical Community Detection (HCD) which utilizes graph neural networks to partition a GRN in a set of hierarchical community representations. Our HCD model first processes an attributed graph using graph attention autoencoder to learn a low dimensional embedding of the node attributes (i.e gene expression). This embedding is then passed to an inference module, composed of feed-forward linear layers. Each layer in this module refines the input from the previous layer into a set number of communities where each set of community predictions is nested within the communities of the next layer, forming a hierarchy. Like dimensionality reduction techniques, this process maps genes in the original network to a smaller set of nodes representing communities of genes with similar expression profiles, simplifying the original network without loss of information about the genes or their potential relationships.
We present preliminary results from applying HCD on simulated hierarchical networks of gene expression. Our evaluation considers parameters such as graph sparsity, noise, and subgraph architectures (e.g., small-world, scale-free, random graph). We focus our efforts on hierarchies with one or two hierarchical layers and apply HCD using both the true graph as input and estimates of the true graph to replicate real-world conditions where the graph is unknown. Our preliminary results are promising and indicate that HCD is capable of accurately predicting the true communities at each hierarchical layer using only an approximation of the true graph.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4063T: FISHNET: Finding Significant Hits In Networks

Authors:

S. Acharya[1], Y. Kang[2], S. Liao[3], W. Jung[1], V. Akbary Moghaddam[4], M. Brent[1]; [1]Washington Univ. in St Louis, Saint Louis, MO, [2]Washington Univ. in Saint Louis, Saint Louis, MO, [3]Washington Univ. in St.Louis, St.Louis, MO, [4]Washington Univ. in St. Louis, St. Louis, MO

Abstract:

The Bonferroni correction sets a stringent significance threshold in GWAS, necessitating large samples and effect sizes to achieve significance. We propose FISHNET, an algorithm that uses prior biological knowledge embodied in gene interaction networks and gene function annotations to identify genes that do not meet the Bonferroni threshold but replicate nonetheless. Its input is gene-level P-values obtained by omicsWAS or by aggregating SNP P-values. It is based on the idea that genes whose P-values are low due to sampling error are distributed randomly across networks and functions, so suggestive P-values that cluster in densely connected subnetworks and share common functions are unlikely to reflect sampling

error and more likely to replicate. FISHNET combines network and gene set enrichment analysis with permutation-based P-value thresholds to identify a small set of exceptional genes that we call FISHNET genes.
We applied FISHNET to 23 GWAS discovery datasets and evaluated the results using 23 replication datasets. In the discovery sets, we identified 583 FISHNET genes. Of these, 268 (46%) were also FISHNET genes in the replication datasets, and 268 (46%) met Bonferroni replication threshold in replication datasets. Notably, 49/268 Bonferroni-replicated genes were not genome-wide significant in the discovery set but were genome-wide significant in replication sets. The replication rate of FISHNET genes matched or exceeded that of all genes at various statistical thresholds in the range ($p \leq 2.9\times10\text{-}6$, $p \leq 2.9\times10\text{^-}2$). By combining summary statistics with biological knowledge, FISHNET identifies replicable gene-trait associations that fail traditional P-value thresholds.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4064T: Gene-Environment and Metabolite-Environment Interaction Effects on Intraocular Pressure in the Canadian Longitudinal Study on Aging

Authors:

R. Lelievre[1], M. Rakesh[1], P. Hysi[2], J. Little[3], E. E. Freeman[1], M-h. Roy-gagnon[1]; [1]Univ. of Ottawa, Ottawa, ON, Canada, [2]KCL St. Thomas Hosp., London, United Kingdom, [3]Univ Ottawa, Ottawa, ON, Canada

Abstract:

Intraocular pressure (IOP) is an important risk factor for glaucoma which is influenced by genetic and environmental factors. While there have been several studies on both the genetic and environmental contributions to IOP, a large portion of the variance in IOP is still unexplained. Metabolites are compounds involved in metabolism which provide a link between the internal (genetic) and external environments. Past research has found that O-methylascorbate, a vitamin C related metabolite, is associated with IOP. Environmental factors related to vitamin C and diet have also been linked to IOP and glaucoma. In addition, researchers have identified several genetic variants which are associated with metabolite concentrations, including O-methylascorbate and another vitamin C related metabolite, ascorbic acid 2-sulfate. Here we aimed to understand how O-methylascorbate and ascorbic acid 2-sulfate, or genetic variants associated with these metabolites, modify the associations between dietary environmental factors and IOP. We used data from 8060 participants of the Canadian Longitudinal Study on

Aging. Using both linear and linear mixed models adjusted for relevant covariates, we tested for interactions between six genetic variants previously found to be associated with O-methylascorbate and ascorbic acid 2-sulfate and four environmental factors related to diet (alcohol consumption frequency, smoking status, fruit consumption, and vegetable consumption). We also tested for interactions between serum concentrations of O-methylascorbate and ascorbic acid 2-sulfate and the four environmental factors. We used a permutation approach to correct for the 32 interaction tests performed. Three interactions were statistically significant after multiple testing correction (adjusted p-value $\leq$ 0.05): rs8050812 and alcohol consumption frequency, rs12414734 and alcohol consumption frequency, and rs61484427 and smoking status. Understanding how genetic variants and metabolites interact with the environment could shed light on biological pathways controlling IOP and lead to improved prevention and treatment of glaucoma.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4065T: Genes, Machines, and Missteps: The Unseen Risks of Machine Learning Derived Phenotypes

Authors:

S. Mukherjee[1], Z. McCaw[2], D. Amar[3], H. Somineni[2], C. O'Dushlaine[2]; [1]Insitro, Issaquah, WA, [2]Insitro, South San Francisco, CA, [3]Insitro, Sunnyvale, CA

Abstract:

Machine Learning (ML) is increasingly employed to generate phenotypes for genetic discovery, either by imputing existing phenotypes to larger cohorts or by creating novel phenotypes. While these ML-derived phenotypes can significantly empower genetic discovery, they can also uncover false-positive associations. Recent research has focused on developing estimators that use both true and ML-derived phenotypes to properly control the type-I error. Our work complements these efforts by exploring under what causal scenarios ML-derived phenotypes incur spurious associations.

Within a simulation-based framework, we generate ML-derived phenotypes while modulating the genetic and environmental correlation with the true phenotype. We evaluate ML-phenotypes based on recovery of variants causal for the target phenotype (recall), and the proportion of significant variants that are true positives (precision). We show that evaluating ML-derived phenotypes based on the out-of-sample predictive performance (e.g. phenotypic R2) is a poor proxy for variant precision/recall. Next, we simulate a scenario where a proxy phenotype is causally downstream of a true phenotype. We demonstrate that the recall

depends on several factors, including the heritability of the true phenotype and the genetic correlation between the true and ML phenotypes. Finally, we compare performance for two scenarios where ML is used to impute a phenotype from associated biomarkers: i) when the biomarkers are causally upstream of the true phenotype, ii) when they are causally downstream. We show that the former scenario does not inflate type-I error rates, while the latter can. However, increasing the correlation between the true and imputed phenotypes can limit type-I error rates even in the latter scenario.
After demonstrating these common pitfalls of using ML-derived phenotypes through simulations, we illustrate them using real-world data from the UK Biobank.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4066T: Genetic & phenotypic associations of frailty with cardiovascular indicators & behavioral characteristics

Authors:

H. Zhao; Sun Yat-sen Univ., Guangzhou, China

Abstract:

Introduction: Frailty Index (FI) is a common measure of frailty, which has been advocated as a routine clinical test by many guidelines. The genetic and phenotypic relationships of FI with cardiovascular indicators (CIs) and behavioral characteristics (BCs) are unclear, which has hampered ability to monitor FI using easily collected data.**Objectives**: This study is designed to investigate the genetic and phenotypic associations of frailty with CIs and BCs, and further to construct a model to predict FI.**Method:** Genetic relationships of FI with 288 CIs and 90 BCs were assessed by the cross-trait LD score regression (LDSC) and Mendelian randomization (MR). The phenotypic data of these CIs and BCs were integrated with a machine-learning model to predict FI of individuals in UK-biobank. The relationships of the predicted FI with risks of type 2 diabetes (T2D) and neurodegenerative diseases were tested by the Kaplan-Meier estimator and Cox proportional hazards model.**Results:** MR revealed putative causal effects of seven CIs and eight BCs on FI. These CIs and BCs were integrated to establish a model for predicting FI. The predicted FI is significantly correlated with the observed FI (Pearson correlation coefficient = 0.660, P-value = 4.96×10^{-62}). The prediction model indicated “usual walking pace” contributes the most to prediction. Patients who were predicted with high FI are in significantly higher risk of T2D (HR =2.635, $P < 2\times10^{-16}$) and neurodegenerative diseases (HR = 2.307, P = 1.62×10^{-3}) than other patients.**Conclusion:** This study supports associations of FI with CIs and BCs from

genetic and phenotypic perspectives. The model that is developed by integrating easily collected CIs and BCs data in predicting FI has the potential to monitor disease risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4067T: Genetic overlap between impaired kidney function and respiratory function: a large-scale genome-wide cross-trait analysis

Authors:

S. Su[1,2], N. Xiang[1,2], Y. Yang[3], P. Yao[4], T. Fu[1,2,5], L. Wang[1,2,5], S. Huang[6], G. Liao[7], L. Du[8], J. Huang[1], Y. Lin[1,9]; [1]Dept. of Urology, Innovation Inst. for Integration of Med. and Engineering, West China Hosp., Chengdu, China, [2]Frontiers Sci. Ctr. for Disease-related Molecular Network, West China Hosp., Sichuan Univ., Chengdu, China, [3]Hlth.Management Ctr., Gen. Practice Med. Ctr., Innovation Inst. for Integration of Med. and Engineering, West China Hosp., Chengdu, China, [4]Dept. of Respiratory and Critical Care Med., West China Hosp., Chengdu, China, [5]Key Lab. of Bio-Resource and Eco-Environment of Ministry of Ed., Coll. of Life Sci., Sichuan Univ., Chengdu, China, [6]West China Sch. of Publ. Hlth., Chengdu, China, [7]State Key Lab. of Oral Diseases, Natl. Clinical Res. Ctr. for Oral Diseases, West China Hosp. of Stomatology, Chengdu, China, [8]Chinese Evidence-Based Med. Ctr., West China Hosp., Chengdu, China, [9]Program in Genetic Epidemiology and Statistical Genetics, Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Background Increasing clinical and epidemiological evidence has highlighted phenotypic association between kidney dysfunction and pulmonary injury. Both kidney and respiratory function are highly heritable and show parallel epidemic trends, suggesting shared genetic and environmental components. However, the extent of shared genetic components between them remains unclear. Thus, we investigated the genetic correlation between kidney and respiratory function traits to describe shared genetic loci and biological pathways between traits. **Method** We conducted a cross-trait genome-wide association study (GWAS) to investigate genetic overlap between 5 kidney function and 4 respiratory function traits. The kidney traits included: glomerular filtration rate estimated from creatinine (eGFRcrea, N=1,004,040) and from cystatin C (eGFRcys, N=460,826), blood urea nitrogen (BUN, N=852,678), urinary albumin-to-creatinine ratio (UACR, N=288,649), urate (N=547,361), sourced from the CKDGen Consortium and UK Biobank. The respiratory traits from UK Biobank were: forced vital capacity (FVC, N=321,047), forced expiratory volume in one second (FEV1, N=321,047), the FEV1/FVC ratio (N=321,047), and peak expiratory flow (PEF, N=321,047). We

evaluated genetic and causal correlations by linkage disequilibrium score regression (LDSC) and Mendelian Randomization (MR). A multi-trait analysis of GWAS (MTAG) was conducted to test if a single SNP can influence multiple traits. Finally, we performed colocalization analysis and transcriptome-wide association studies (TWAS) to identify causal variants and link genes to tissues. **Results** We revealed negative genetic correlations between respiratory function and eGFRcrea (rg = -0.064 to -0.043) and urate (rg = -0.086 to -0.111), along with positive correlations between respiratory function and eGFRcys (rg = 0.049 to 0.105). Through MATG, we determined 146, 83, 26, 14 and 22 overlapping genetic loci for eGFRcrea, eGFRcys, BUN, UACR and urate (Pmtag<5e-8, Psingle_trait<5e-6), 9 of which had strong evidence of colocalization (H4_P>0.95). Moreover, the pleiotropic loci were enriched in brain-related tissues, while 2 shared genes, *ABO* and *RF00017*, which were found through MTAG, colocalization and TWAS, exhibited notable enrichment in kidney cortex. MR analyses showed a significant causal association of respiratory function on urate (OR = 0.79 to 0.96). **Conclusion** This study demonstrated a shared genetic architecture and a causal association between kidney and respiratory function traits, providing novel insights into the molecular mechanisms underlying the association between kidney and lung.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4068T: Genome-wide aggregated *trans*-effects analysis on the risk of polymyalgia rheumatica and giant cell arteritis in UK Biobank

Authors:

A. Nwagbata[1], A. Iakovliev[1], A. Morgan[2], P. McKeigue[1], A. Spiliopoulou[1]; [1]Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Univ. of Leeds, Leeds, United Kingdom

Abstract:

Background: Polymyalgia rheumatica (PMR) and giant cell arteritis (GCA) are frequently overlapping age-associated inflammatory diseases. PMR manifests with pain, stiffness, and inflammation in the muscles around the shoulders, neck, and hips, whereas GCA primarily affects the arteries, notably the aorta and its extracranial branches. Our objective is to identify potential “core” genes involved in the pathogenesis of PMR and GCA through a genome-wide aggregated trans-effects analysis, aimed at facilitating drug discovery efforts.
Methods: We used eQTL (eQTLGen) and pQTL (deCODE, UKB-PPP) summary statistics to compute genotypic scores for trans-regulated gene expression or protein levels in a cohort comprising 6237 PMR/GCA cases and 458530 controls from the UK-Biobank. These genotypic scores were aggregated per gene or protein (GATE scores) and then tested for association with

PMR/GCA. Core genes were further identified based on the strength of the association, the number of independent trans-loci contributing to each GATE score, and the lack of correlation between core genes. Subsequent validation included replication of associations in an independent cohort of 1878 GCA cases and 5520 controls, along with an analysis of measured protein levels of these core genes in 668 PMR/GCA cases and 50,626 controls from the UK-Biobank.
Result: Our analysis identified seven putative core genes at a significance threshold of P<10^(-05): *ANKRD55* and *LEF1* from the eQTL GATE analysis,
and *CXCL9*, *LECT2*, *PLG*, *MST1*, *CXCL10* from the pQTL GATE analysis. Five of these genes are involved in pathways mediating inflammation and immunity.
Furthermore, *ANKRD55* and *PLG* have been previously implicated in GWAS studies of GCA/PMR, providing independent validation due to the exclusion of cis-acting SNPs from the GATE scores. *LECT2* showed the strongest association at P<10^(-09) in the independent cohort replication analysis. Additionally, the measured protein levels of four out of five pQTL core genes were associated with PMR/GCA.
Conclusion: Using GATE analysis, we identified core genes for PMR/GCA, highlighting significant pathways underlying these conditions. *LEF1*, *LECT2*, *CXCL9*, and *CXCL10* play critical roles in inflammation and immune response modulation, while *MST1* regulates dendritic cell activity, crucial in antigen presentation to T cells. These findings offer promising avenues for drug target identification or repurposing once clinically validated.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4069T: Genome-wide association study of fibrotic disease across organs and systems

Authors:

E. Joof[1], G. M. Massen[2], O. C. Leavy[1], G. Parcesepe[1], I. D. Stewart[2], G. P. Aithal[3], C. J. Scotton[4], D. P. Auer[3], S. Francis[3], G. R. Jenkins[2], J. K. Quint[2], H. Longhurst[5], L. V. Wain[1], R. J. Allen[1], DEMISTFI Consortium; [1]Univ. of Leicester, Leicester, United Kingdom, [2]Imperial Coll. London, London, United Kingdom, [3]Univ. of Nottingham, Nottingham, United Kingdom, [4]Univ. of Exeter, Exeter, United Kingdom, [5]Dyskeratosis Congenita (DC) Action, London, United Kingdom

Abstract:

Rationale: Fibrosis, i.e., the thickening or scarring of tissue, is a major factor in global mortality and can affect virtually every organ system. However, there is no good treatment for fibrosis. Rather than considering single fibrotic diseases, we can group diseases by organ; then leverage

this to identify shared mechanisms of fibrosis across organs. This may be beneficial for new treatment development, drug repurposing and drug adverse effects management.
Objectives: We sought to identify shared genetic signals of fibrosis across different organs and systems.
Methods: We included unrelated European ancestry individuals from UK Biobank. We defined a disease as "fibrotic" if there was clinical consensus that it was either "always fibrotic" or "can develop fibrosis" from a published Delphi survey. The fibrotic diseases were then assigned to one of the following 11 organ groups (pulmonary, liver, biliary, skin, cardiovascular, intestinal-pancreatic, skeletal, systemic, reproductive, renal and diabetes). Individuals with at least one fibrotic disease recorded in their hospital episode statistics (HES) or mortality records were defined as cases (all others were controls). Genome-wide association studies (GWAS) for each organ were performed separately and meta-analysed with publicly available GWAS results of corresponding single fibrotic diseases from non-UK Biobank cohorts. To assess whether the signals shared causal variant(s) between organ groups, colocalisation analysis with the Coloc package was performed using a colocalisation probability threshold of 70%.
Results: We identified four signals shared between two or more organ groups at $p<5x10^{-6}$. A signal near *PTPN22* was shared across skeletal fibrosis, skin fibrosis and diabetes, a signal near *FCGR2A* was shared between skeletal and skin fibrosis, a signal near *GCKR* was shared between liver and renal fibrosis, and a signal near *SH2B3* was shared across four organs (renal, biliary, skeletal and skin).
Conclusions: These shared signals may reveal potential causal pathways of general fibrosis and provide new insight into mechanisms shared by different fibrotic diseases which can be leveraged in the development of new therapies. We are currently expanding the analyses to other ancestry groups to better define the extent of shared genetic signals for fibrosis across all populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4070T: Identification of moderate effect size genes in autism spectrum disorder through a novel gene pairing approach

Authors:

M. Caballero[1], F. Satterstrom[2], J. Buxbaum[3], B. Mahjani[3]; [1]Icahn Sch. of Med. at Mt Sinai, New York, NY, [2]Broad Inst., Cambridge, MA, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Autism Spectrum Disorder (ASD) arises from complex genetic and environmental factors, with inherited genetic variation playing a substantial role. This study introduces a novel approach to uncover moderate effect size (MES) genes in ASD, which individually do not meet the ASD liability threshold but collectively contribute when paired with specific other MES genes. Analyzing 10,795 families from the SPARK dataset, we identified 97 MES genes forming 50 significant gene pairs, demonstrating a substantial association with ASD when considered in tandem, but not individually. Our method leverages familial inheritance patterns and statistical analyses, refined by comparisons against control cohorts, to elucidate these gene pairs' contribution to ASD liability. Furthermore, expression profile analyses of these genes in brain tissues underscore their relevance to ASD pathology. This study underscores the complexity of ASD's genetic landscape, suggesting that gene combinations, beyond high impact single-gene mutations, significantly contribute to the disorder's etiology and heterogeneity. Our findings pave the way for new avenues in understanding ASD's genetic underpinnings and developing targeted therapeutic strategies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4071T: Identification of potential drug targets for cardiac congenital anomalies through proteome-wide mendelian randomization analysis

Authors:

Y. Wan[1], Q. Zhou[2], T. Gong[1], H. Le[1], J. Ma[1], H. Deng[1], **Z. Zhong**[1], C. Sun[1], X. Lei[1], C. Zhang[1]; [1]The First Affiliated Hosp., Jiangxi Med. Coll., Nanchang Univ., Nanchang, China, [2]The First People's Hosp. of Fuzhou, Fuzhou, China

Abstract:

Cardiac congenital anomalies (CCAs) pose significant health challenges, necessitating the exploration of novel therapeutic strategies. Proteome-wide Mendelian randomization (PW-MR) analysis offers a promising approach to uncover causal relationships between plasma proteins and CCAs, and potentially reveal new targets for therapeutic intervention. This study utilized comprehensive genome-wide association study (GWAS) summary data to conduct PW-MR analysis, employing plasma protein levels as exposure and CCAs as outcome variables. Data from two cohorts, UK Biobank Pharma Proteomics Project (UKB-PPP) and deCODE Health study, were analyzed. The analysis encompassed validation through reverse MR analysis, biological significance exploration via gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, colocalization analysis for shared genetic associations, druggability

assessment, and phenome-wide association study (PheWAS) to assess therapeutic potential. PW-MR analysis identified 88 plasma proteins causally linked to CCAs, highlighting *LPA, NTRK3*, and *LEFTY2* as promising therapeutic targets. Colocalization analysis confirmed shared genetic associations between these proteins and CCAs. GO and KEGG enrichment analysis revealed the involvement of identified proteins in immune responses, cell adhesion, and hematopoietic cell lineage pathways. Druggability assessment identified existing drugs targeting *LPA* and *NTRK3*. PheWAS analysis showed no significant causal links between identified proteins and other health conditions. This study underscores the potential of PW-MR analysis to identify novel therapeutic targets for CCAs. *LPA*, *NTRK3*, and *LEFTY2* emerge as promising candidates for targeted intervention in CCAs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4072T: Identification of Target Patient Populations Using Genetic Proxies of Drug Targets and Time-to-event analyses

Authors:

L. Zhang, P. Kulkarni, F. Farshidfar, W. Tingley, T. Hoey, W. Wang, J. Priest, S. Figarska; Tenaya Therapeutics, South San Francisco, CA

Abstract:

Background: Human genetics is an important tool for identifying genes as potential drug targets, and the extensive genetic study of cardiovascular disease provides an opportunity to leverage genetics to match specific patient populations to specific drug targets to improve prioritization of patient selection for clinical studies. **Methods:** We selected well described genetic variants in the region of PCSK9 (rs11591147 and rs562556), ADRB1 (rs7076938), ACE (rs4968782 and rs4363) and BAG3 (rs2234962) for use as proxies for the effects of drugs. Time-to-event analyses were utilized to evaluate their effects on atrial fibrillation (AF) and heart failure (HF) death and/or re-hospitalization using real-world longitudinal dataset. To mitigate the effect of confounding factors for cardiovascular (CV) outcomes, we employed propensity score matching.Results: After matching, a genetic proxy for PCSK9 inhibition (rs11591147) improved survival from CV death/heart transplant in individuals following a diagnosis of ischemic heart disease (Hazard Ratio (HR)=0.78, P=0.04). A genetic proxy for beta-blockade (rs7076938) improved freedom from rehospitalization or death in individuals with AF (HR=0.92, P=0.001), and a genetic proxy of ACE inhibition (rs7076938) improved freedom from death or rehospitalization for HF (HR=0.84, P=0.017) and AF (HR=0.85, P=0.0014). A protective variant in BAG3 (rs2234962) showed decreased risk of HF rehospitalization or CV death/heart transplant

composite outcome within 10 years in HF patients (HR=0.96, P=0.033). Notably, despite smaller cohort sizes after matching, we often observed numerically smaller HRs and reduced P, indicating more pronounced effects and increased statistical association. However, not all genetic proxies replicated known treatment effects. **Conclusions:** Genetic proxies for well-known drugs corroborate findings from clinical trials in cardiovascular disease. Our results may demonstrate a novel analytical approach that leverages genetic evidence from a large cohort to effectively select patient populations where specific drug targets may be most effective.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4073T: Identifying Rheumatoid Arthritis Gene-regulatory Enhancers with RNA+ATAC Multiome Data

Authors:

L. Liu[1,2], T. Park[1], S. Mitra[1], M. Smith[3,4], L. Donlin[3,2], K. Dey[1,2], C. Leslie[1,2]; [1]Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Weill Cornell Graduate Sch. of Med. Sci., New York, NY, [3]Hosp. for Special Surgery, New York, NY, [4]Weill Cornell Med. Coll., New York, NY

Abstract:

Joint profiling of chromatin accessibility and gene expression in 10X scRNA+ATAC multiome data can be used to uncover the enhancer-driven regulatory architecture of genes specific to different cell types. Our recently developed model, SCARlink (Mitra et al. 2024 Nat Genet), can be used to predict gene expression from chromatin accessibilities of proximal and distal genomic tiles and assess the relative importance of each such tile. We applied SCARlink to a 10X multiome dataset from synovial tissue of 11 Rheumatoid Arthritis (RA) patients (32,554 cells) sorted into equal proportions of macrophages, fibroblasts, and T cells. RA is a chronic autoimmune disease in synovium with symptoms of joint damage, swelling and reduced function. Notably, this dataset was sequenced in three patient batches contributing 12%, 37% and 51% of cells, respectively. Therefore, we performed a comprehensive analysis applying SCARlink to data from each batch separately, as well as integrated data across all batches. In total, we trained 3,476 genes using 4,188,066 genomic tiles and attributed tile importance scores based on cell-type agnostic and cell-type level annotations.
We assessed disease relevance of SCARlink identified tiles by exploring their overlaps with 581 variants in GWAS 95% credible sets from a multi-ancestry meta-analysis study for RA (N-case=35,871, N-controls = 240,149) (Ishigaki et al. 2022 Nat Genet) and 186 OpenTargets high “locus-to-gene” (L2G > 0.5) maps corresponding to RA (Carvalho-Silva et al. 2019 NAR). We pinpointed 11 genomic tiles from 6 genes (Fisher's test, p = 6.91*10^-12) overlapping these

disease-relevant variants. One such example is a causal variant rs798000 (PIP=1) residing in a tile of CD2 gene in T cells, which has been identified across multiple GWAS studies focusing on RA cases with European ancestry. CD2 plays a regulatory role in immune response, and genetic variations proximal to this gene could lead to the change of immune signaling. Another example is rs13330176 (L2G = 0.74) harbored by a macrophage-specific tile nearby IRF8, which has been implicated in multiple studies focusing on RA. IRF8 is recognized as a pivotal lineage-determining transcription factor essential for immune response regulation.Our findings proposed potential enhancers of RA immune regulations, and demonstrated the advantage of leveraging tile-level chromatin accessibilities for predicting gene expression and identifying cell-type-specific enhancers without peak-calling, shedding lights on effective method for investigating enhancer's regulatory impact on target gene expression.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4074T: Imp-Map: A novel biobank scalable approach to fine-mapping and imputation

Authors:

C. Foley[1,2], Z. Kuncheva[3,2], A. Hutchinson[4], S. Afyouni[3,2], M. Dissanayake[1,2], D. McCartney[2,1], A. Cortes[5]; [1]bioXcelerate AI, Edinburgh, United Kingdom, [2]Univ. of Edinburgh, Edinburgh, United Kingdom, [3]bioXcelerate AI, London, United Kingdom, [4]GSK, London, United Kingdom, [5]GSK, Heidelberg, Germany

Abstract:

Statistical fine-mapping and imputation methodologies are routinely used to enhance biological insights from genome-wide association studies (GWAS) and often embedded in genetics-driven drug discovery pipelines. Their seemingly ubiquitous use is a consequence of impact: helping maximise coverage of genotype to trait associations (imputation); and assisting in the identification of candidate causal variants in a genomic region (fine-mapping). The connection between these approaches is under-explored and significant bottlenecks in their successful deployment persist. Most notably, the dramatic increase in cohort sizes and diverse data types have surfaced an immediate need for novel, biobank scalable approaches: applications to these scenarios often result in computationally sluggish outcomes or no biological insights at all.
To meet this challenge, we present new and complimentary methodologies. First, we describe Switch-Step, a novel and ultra-fast fine-mapping method that uses greedy variable selection with adaptive switching to identify the most likely candidate causal variants. Second, we showcase Imp-Map, a uniquely scalable algorithm for imputing GWA summary statistics (Z-

scores, effect sizes and standard errors). We highlight the computational and statistical links between the two methodologies which are fundamental to the significant efficiency and statistical robustness gains we achieve.
Extensive empirical tests and simulation scenarios demonstrate that Switch-Step results are highly comparable to traditional regression-based approaches, but are over 100 times faster. Further still, we demonstrate that the Imp-Map procedure is more than 1000 times faster relative to classical approaches, while also benefiting from increased accuracy.
Owing to the significant computational and accuracy gains identified, we anticipate a broad use of Switch-Step and Imp-Map in research and industry settings: helping robustify and improve computational frameworks in genetics-driven drug discovery.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4075T: ImputeMLP: An efficient and accurate deep-learning method for genotype imputation method

Authors:

U. An, J. Fu, S. Sankararaman, E. Eskin; Univ. of California Los Angeles, Los Angeles, CA

Abstract:

Genotype imputation methods use a statistical model to infer missing genetic variants from observed ones and are widely used to increase the power to detect genetic variants associated with traits and diseases. Current state-of-the-art methods for genotype imputation use a Hidden Markov Model (HMM) to model LD between observed and target variants. The ability of deep learning-based models to capture complex non-linear relationships has motivated recent works that use Deep Learning (DL) to impute genotypes. However, current DL-based genotype imputation methods often rely on synthetic settings that don't reflect common genotyping panels, overlook non-trivial panel correlations, simulate too few missing SNPs which do not reflect a practical use case, or struggle to scale with large datasets.We propose ImputeMLP, an accurate and efficient imputation method based on Multi-Layer Perceptrons (MLP). Unlike prior approaches that train models and perform denoising on an arbitrary set of observed SNPs, which lead to a large number of parameters that need to be estimated, our model is configured to perform imputation based on SNPs that are genotyped on standard arrays. Our approach is lighter in trainable parameters than the prior works, allowing us to explore the benefits of observing a wider neighborhood of SNPs more efficiently. We demonstrate the benefits of our approach in simulations using genotypes acquired using the UKB Axiom Array across 339,288 white British unrelated individuals from the UK Biobank (UKB) that were imputed into SNPs on

chromosome 22 discovered in the UK10K and Haplotype Reference Consortium. In this setting, ImputeMLP obtained a median increase of 5% in imputation accuracy (r^2) across 14,270 SNPs in comparison to the next best method (Minimac4) with a 4.6-fold increase in imputation accuracy for rare SNPs (MAF $<1 \times 10^{-4}$). In total, 3.2x more SNPs increase in accuracy with significance than decrease. To quantify whether these results generalize in settings beyond UKB, we explored the accuracy of imputing SNPs in the 1000 Genomes European individuals with models trained in the UKB. In this setting, ImputeMLP obtains accuracies that are comparable to HMM-based methods; however, our approach does not require access to the reference panel, allowing portability and faster inference times. Finally, we show that the accuracy of ImputeMLP grows with increasing sample size, outperforming HMM-based methods when the sample size exceeds 50k, suggesting that the advantage of deep-learning methods is expected to increase with growing sample sizes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4076T: Incorporating additive genetic effects and full LD information to discover genome-level gene-environment interactions with summary statistics of complex traits

Authors:

Z. Dong, W. Jiang, J. Shen, A. Dewan, H. Zhao; Yale Sch. of Publ. Hlth., New Haven, CT

Abstract:

Title: Incorporating additive genetic effects and full LD information to discover genome-level gene-environment interactions with summary statistics of complex traitsAbstract: Uncovering the environmental factors interacting with genetic factors to influence complex traits is important in understanding disease etiology and improving disease prevention and treatment. Existing methods examining gene-environment (GE) interactions either test GE interaction effects ignoring the correlations with additive genetic effects or use only partial information of Linkage Disequilibrium (LD), leading to the loss of statistical power to uncover interacting factors. We introduce the BiVariate Linkage-Disequilibrium Eigenvalue Regression for Gene-Environment interactions (BV-LDER-GE), a novel statistical method that detects the overall contributions of GE interactions in the genome using summary statistics of complex traits. BV-LDER-GE harnesses both the correlations with additive genetic effects and full LD information to enhance its statistical power to detect genome-scale GE interactions. Extensive simulations demonstrate that BV-LDER-GE is more powerful than existing methods while the type-I error rate is well-controlled. Applying BV-LDER-GE to the UK Biobank data, we examined 151

environmental covariate-phenotype (E-Y) pairs including seven common environmental covariates and 22 widely studied phenotypes and diseases. BV-LDER-GE identified 63 statistically significant interacting environmental covariate phenotype pairs (E-Y pairs), compared to 35 pairs identified by existing methods using summary statistics. Our results suggest the pervasive existence of GxE interaction effects on complex traits. The power of BV-LDER-GE comes from two aspects: the joint modeling that gathers partial information from additive genetic effect, whose variant-level magnitude is generally stronger than the G×E interaction effect; and the exploit of full LD information improves the efficient of model parameter inference and subsequent testing.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4077T: Integrating genetic data and electronic health records for health risk predictions using a graph convolution network

Authors:

Y. Miao[1], S. Wang[2], C. Weng[1], K. Kiryluk[3]; [1]Columbia Univ., New York, NY, [2]Columbia Universiy, New York, NY, [3]Columbia Univ. Hlth.Sci., New York, NY

Abstract:

Electronic Health Records (EHR) and molecular profiles such as genetic or gene expression data of patients are different aspects of patient information, each providing valuable and unique information in personalized health risk predictions. However, due to the distinct data dimensions of the two data types, it is methodologically challenging to integrate them effectively for prediction purposes. Here we propose GCN-EPI: a Graph Convolution Network for EHR data and Patient graph from molecular profiles Integration. With patient graphs from patient molecular profiles and through graph convolution, GCN-EPI aggregates EHR data and molecular profiles by weighting EHR information using neighboring patients through patient graphs adaptively to improve patient risk predictions. Simulation studies suggested superior prediction performance of GCN-EPI over that of competing methods including Lasso and random forest on concatenated features from molecular profiles and EHR, and similarity network fusion (SNF) that fuses two similarity matrices from molecular profiles and EHR. We applied the proposed GCN-EPI algorithm and competing methods to predict incident end-stage kidney disease (ESKD) using genetic data and EHR data extracted from the data warehouse of Columbia University Irving Medical Center and observed improved prediction results of the proposed GCN-EPI.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4078T: Integration of CNVs with SNPs enhances exome-wide association study analysis of 770,000 samples ⭐

Authors:

O. Delaneau[1], J. Mbatchou[1], M. Kessler[1], N. Parikshak[1], X. Bai[1], E. Maxwell[1], M. Suciu[1], V. Rajagopal[1], J. Torres[2], R. Collins[3], J. Emberson[3], J. Berumen[4], P. Kuri-Morales[5], J. Alegre-Díaz[6], R. Tapia-Conyer[6], Mexico City Prospective Study, GHS-RGC DiscovEHR Collaboration, N. Verweij[7], W. Salerno[8], J. Reid[9], G. Abecasis[1], J. Marchini[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]The Univ. of Oxford, Oxford, United Kingdom, [3]Univ. of Oxford, Oxford, United Kingdom, [4]Univ. Natl. Autónoma de México, México, Mexico, [5]Tecnológico de Monterrey, Monterrey, Mexico, [6]Univ. Natl. Autónoma de México, Mexico city, Mexico, [7]Regeneron Genetics Ctr., New York, NY, [8]Regeneron, New York, NY, [9]Regeneron Pharmaceuticals, Tarrytown, NY

Abstract:

The incorporation of copy-number variants (CNVs) into the analysis of Exome-wide Association Studies (ExWAS) has not been widely adopted in the field. However, recent studies of CNVs have demonstrated their significant role as loss-of-function (LoF) events of protein-coding genes and their widespread association with multiple complex traits. In this work, we explored the impact of integrating CNVs with SNPs and short indels in the context of meta-analysis of large-scale ExWAS. To detect CNVs, we employed two methods. One method that uses coverage profiles to identify CNVs spanning multiple exons and a second method that uses SNP variation and haplotype sharing among individuals to refine the CNV profiles discovered from coverage data alone. Our second method utilizes a scaffold of haplotypes, derived from SNP array data, to pinpoint local haplotype sharing between individuals. Then, calls are made for an individual by borrowing information from other individuals who share local haplotype segments. We validated this approach at multiple levels by looking at (i) depletion of SNP heterozygosity in putatively deleted regions, (ii) metrics of gene tolerance to loss-of-function events, (iii) comparison with Whole Genome Sequencing based analysis of the same samples, and (iii) Mendelian inconsistencies in parent offspring trios. All these analyses confirmed the high accuracy of our calling strategy. We applied our approach to the UK Biobank (UKB, N=464,478), the Mexico City Prospective Study (MCPS, N=137,657) and the Geisinger Health System (GHS, N=168,013), comprising both WES data and SNP array data for >770,000 samples in total. Integrating CNVs resulted in an effective increase of ~10% in the number of putative loss-of-function events. To test the impact on discovery power, we examined ~80 complex traits (such as height, LDL, and blood pressure) within the UKB dataset and a subset of those in the other two cohorts. We compared scans incorporating CNVs with scans using SNPs and short indels

only. Overall, CNV loss-of-gene function events enhanced the signal in 20% of the gene-based association signals and resulted in a net 8% increase in the yield of significantly associated genes. One notable example of this improvement is the deletion of PCSK9 exon 1, which increased the number of pLoF carriers for this gene in the UKB dataset by 17.5%, leading to a better association signal with LDL (before: 8.7×10^{-159}, after: 1.1×10^{-182}). In summary, this research presents new, efficient tools to detect small CNVs in WES data. It demonstrates that significant improvements can be achieved in exome wide association studies by incorporating CNVs, along with SNPs and indels.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4079T: Integration of multi-omics data & eQTL summary statistics with omicsSTAAR boosts power in rare variant association tests of non-coding regions ⭐

Authors:

E. Van Buren[1], X. Li[2], Z. Li[3], P. Orchard[4], H. Zhou[5], A. Reiner[6], L. Raffield[7], X. Lin[8], NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium, TOPMed Multi-Omics Working Group; [1]Harvard Univ., CAMBRIDGE, MA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Northeast Normal Univ., Changchun, China, [4]Univ. of Michigan, Ann Arbor, MI, [5]Harvard, Boston, MA, [6]Univ of Washington, Seattle, WA, [7]UNC - Chapel Hill, CHAPEL HILL, NC, [8]Harvard T.H. Chan Sch Pub Hlth, Boston, MA

Abstract:

Introduction
Through our previously developed cellSTAAR method, we demonstrated that integration of single-cell-sequencing-based epigenetic data can boost the power of gene-centric Rare Variant (RV) association tests (RVATs) to detect associations of candidate Cis-Regulatory Elements (cCREs) in complex human diseases. Integrating additional kinds of multi-omics data to capture additional sources of functional variability that exists in the non-coding genome may further increase power.

Methods
We propose omicsSTAAR as a new method to robustly integrate several kinds of multi-omics data into gene-centric RVATs of non-coding regions. First, omicsSTAAR can integrate variant-level multi-omics datasets, such as from methylation or expression QTL summary statistics, to create custom variant sets of the most likely causal variants weighted with corresponding functional annotations. Association p-values from each variant set are aggregated using the Cauchy

Combination Test to create an omnibus p-value summarizing evidence across different categories of multi-omics data. Second, omicsSTAAR can integrate gene-level multi-omics datasets, such as RNA-seq and proteomics experiments, to weight omnibus gene-centric association p-values using "side-information" approaches such as Independent Hypothesis Weighting (IHW). Using such approaches, omicsSTAAR can account for the biological activity of each gene as measured by expression or protein abundance in relevant tissues.

Results

We applied omicsSTAAR on Freeze 8 (N = 60,000) of the NHLBI Trans-Omics for Precision Medicine (TOPMed) consortium data of four hematological traits: hemoglobin (HGB), hematocrit (HCT), platelet count (PLT), and white blood cell count (WBC). To demonstrate omicsSTAAR, we collected single-cell ATAC-seq data and two TOPMed blood-based datasets: RNA-seq from the WHI and FHS TOPMed cohorts (N = 2,072) and eQTL summary statistics based on 6,602 TOPMed participants. Our analysis reveals associations in several known genes for hematological traits, including HBQ1 and CD84, while showing variability in which kinds of omics data detect each association. We also demonstrate a substantial increase in the number of discoveries at a reduced significance threshold when combining the variant-level multi-omics data (scATAC-seq and eQTL summary statistics) association results into an omnibus association p-value and when using gene-level multi-omics data (RNA-seq) to weight the gene-centric omnibus p-values.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4080T: Interaction analysis of smoking, prostate specific antigen, particulate matter, and polygenic risk score on prostate cancer incidence in Korea

Authors:

S-H. Jee[1], U. Kim[2], J. Shin[1]; [1]Graduate Sch. of Publ. Hlth., Inst. for Hlth.Promotion, Yonsei Univ., Seoul, Korea, Republic of, [2]Dept. of Biostatistics and Computing, Coll. of Med., Yonsei Univ., Seoul, Korea, Republic of

Abstract:

Prostate cancer is one of the most rapidly increasing cancers in South Korea. Smoking and Prostate Specific Antigen are well-established risk factors for prostate cancer. This study aims to investigate the interactions between the Polygenic Risk Score for prostate cancer and smoking, PSA, and particulate matter.Data were obtained from the Korean Cancer Prevention Study-II, including male participants with smoking history and PSA test results, for whom PRS

calculations were available. The study sample comprised 10,327 individuals with 327 cases. The PRS of prostate cancer was calculated using 451 SNPs identified in a previous study. Interaction analyses were conducted using generalized linear mixed models, GLMM, with age included as a covariate. Random effects included random intercepts for smoking, PSA, and PM, as well as random slopes for the interaction with PRS. Interaction effects were assumed to reflect the heterogeneous influence of environmental factors on prostate cancer risk based on PRS. Statistical significance of interactions was assessed using likelihood-based tests, and the fraction of genetic variance explained by interactions was estimated. Fitting the GLMM showed that the fixed effect for the second tertile of PRS had a beta coefficient of 1.246, and the third tertile had a beta coefficient of 2.603 with p value&lt0.001. The random slope variance for PRS and smoking was 1.033 for the second tertile and 1.217 for the third tertile, the highest among the environmental factors. Interaction analysis revealed significant interactions between PRS and smoking with p value&lt0.001 and PRS and PSA with p value&lt0.001, while no significant interaction was observed with PM. The gene environment interactions analysis indicated significant heterogeneous genetic associations for PM, smoking with p value&lt0.001, and PSA with p value&lt0.001. The estimated gene environment interactions values were highest for smoking, followed by PSA, and PM. Higher PRS for prostate cancer are associated with increased odds of prostate cancer incidence. The effect of PRS on prostate cancer varies with environmental factors such as smoking, PSA, and PM, with the greatest heterogeneity observed for smoking. These findings suggest that random effect models considering genetic and environmental heterogeneity are more appropriate than single effect models for prostate cancer. This underscores the existence of gene environment interactions, indicating the need for future research to explore more candidate environmental factors and develop optimal models considering these interactions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4081T: Intraplaque haemorrhage prediction and MACE risk modelling using attention based multiple instance learning

Authors:

F. Cisternino[1], Y. Song[2], T. Peters[3], R. Westerman[3], N. Van den Dungen[3], G. J. de Borst[4], E. Diez Benavente[3], P. H. Van der Kraak[5], D. P. V. de Kleijn[4], J. Mekke[4], M. Mokry[3], G. Pasterkamp[3], H. M. den Ruijter[3,6], E. Velema[6], C. L. Miller[2], S. van der Laan[3,2], C. A. Glastonbury[1]; [1]Population & Med. Genomics Res. Ctr., Human Technopole, Milan, Italy, [2]Ctr. of Population Hlth.and Genomics, Univ. of Virginia, Charlottesville, VA, [3]Central Diagnostics Lab., Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [4]Vascular surgery, Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [5]Pathology, Univ.

Med. Ctr. Utrecht, Utrecht, Netherlands, [6]Experimental Cardiology, Univ. Med. Ctr. Utrecht, Utrecht, Netherlands

Abstract:

Atherosclerosis is a chronic inflammatory process affecting arterial walls and is a leading cause of cardiovascular diseases. A key aspect of atherosclerosis is the development of intraplaque haemorrhage (IPH), which marks the transition of a stable plaque into a vulnerable one. Vulnerable plaques have a higher propensity to rupture, often leading to acute cardiovascular events. Identifying the presence and extent of IPH as well as its biological origins is critical in predicting and mitigating the risk of Major Adverse Cardiovascular Events (MACE) post intervention. Current approaches to assess IPH rely on manual scoring. To automatically detect and quantify IPH and understand its prognostic utility in predicting MACE and the underlying biology, a large cohort (~13K) of plaque samples were collected from 2,461 patients undergoing carotid arterial endarterectomy. Samples were stained for several markers and histological stains (H&E, SMA - Smooth Muscle Actin, EVG - elastic fibers, glycophorin C, fibrin, CD68, CD66b, CD34 and SR for fibrosis). In addition to these Whole Slide Images (WSI), plaque RNA-Seq (N=1,093) and germline genotypes (N=1,992) were collected in a subset of donors. To quantify IPH from plaque histology, an additive Multiple Instance Learning model was utilised, which provides clear interpretability of IPH presence by highlighting regions of the plaque which contribute to the sample-level prediction of IPH. These patch level scores can be used to quantify haemorrhage extent. Across all stains tested, H&E resulted in the highest AUC when classifying IPH events (N=2,223, AUC=0.86). We also assessed the gain of modelling multiple stains and observed that H&E + EVG + SR resulted in a 9% gain in AUC. Little is known about gene expression regulation in regions of IPH. To study this, we fit differential expression models for IPH, adjusting for age, sex and batch. Overall, we identified a total of 43 upregulated and 45 downregulated genes (FDR 5%), which are enriched for macrophages (p=0.003) and include known enzymes (e.g. histidine decarboxylase - HDC, produced by macrophages and mast cells in atherosclerotic lesions). Finally, we performed time-to-event analysis using IPH area to predict MACE and found a significant stratification of risk (p=0.04). In conclusion, we have conducted the largest study of IPH and its role in MACE to date, by integrating imaging, transcriptomics and clinical data to automate the quantification of IPH in plaques, as well as find biomarkers as biological proxies of IPH. On-going work aims to identify common genetic variation associated with IPH severity and plaque vulnerability through genome-wide association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4082T: Investigating the determinants of frailty: The intersection of genetics and environment in Mexican Americans from South Texas

Authors:

N. Garcia-Rodriguez, V. Diego, M. Almeida, J. Galan, S. Laston, J. Curran, T. Howard, S. Williams-Blangero, J. Blangero, E. MANUSOV; Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX

Abstract:

Background: The projected worldwide population of people aged 65 and over is anticipated to hit 1.6 billion by 2050. Frailty, marked by increased susceptibility to adverse health outcomes, mortality, and decreased health-related quality of life (HrQoL), poses significant challenges to an aging population. The Frailty Index (FI), is the ratio of the number of accumulated health deficits to the total number of deficits considered, is used to predict frailty.
Methods: Utilizing data from the Mexican American Family Study, we investigated the heritability and GxE interactions influencing frailty in 1,029 Mexican American participants. We calculated the FI using 34 variables, including phenotypes, social determinants of health, medical history, and metabolism-related behaviors. Environmental factors included dietary, physical activity, socioeconomic, and dichotomous disease variables. We employed a linear mixed model approach to estimate heritability, and to investigate GxE interactions, incorporating dichotomous and continuous environmental measures. Statistical analysis utilized likelihood-based estimation and hypothesis testing.
Results: The FI demonstrated moderate heritability (h^2=0.36, p=6.2×10-12). We found significant GxE interactions across six discrete “environments” (obesity, hypertriglyceridemia, low high-density lipoprotein, impaired fasting glucose test, diabetes, and hypertension) and nine continuous measures of dietary environment (polyunsaturated fat, carbohydrates, calories, monounsaturated fat, fiber, protein, saturated fat, and cholesterol). Between-environment genetic correlation analysis suggested distinct gene sets influencing frailty under different conditions, with significant genetic contributors (inflammation and diet) to frailty.
Discussion: Our findings highlight the complex interplay between genetic and environmental factors in the development and progression of frailty. The results help direct personalized approaches in frailty management. Our discussion includes novel statistical methods and focus on how our results can help direct aging research and healthcare.
Conclusion: The study investigates the heritability and gene-environment interaction (GxE) in frailty, showing how lifestyle and environment can affect gene expression and contribute to frailty. The findings highlight the importance of considering heritability and GxE interactions in frailty research and advocates for future exploration of specific genes and mechanisms involved in frailty development and progression.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4083T: LEGO: Local ancestry Enable predicting Gene expression Optimally

Authors:

D. Chen, L. Wang, X. Wu, D. Liu; Penn State Coll. of Med., Hershey, PA

Abstract:

Accurate prediction of genetically regulated gene expression is essential for transcriptome-wide association studies (TWAS). However, gene expression prediction models developed primarily with European ancestry samples demonstrate significantly reduced accuracy when applied to non-European populations. Despite ongoing efforts to generate eQTL datasets in diverse ancestries, the availability of non-European eQTL datasets remains limited. To tackle these issues, we propose to integrate eQTL datasets from multi-ancestries to borrow strength of large sample size and improve the gene expression prediction. Our approach, LEGO, exploits the insight that eQTL effects within admixed populations represent a weighted combination of effects from ancestral populations. For instance, in African Americans, eQTL effects are a weighted average of that of European and African ancestries, while in Hispanics, they comprise effects from European, African, and Native American ancestries. With decomposed eQTL effects of admixed populations, we can jointly analyze them with other non-admixed populations, which improves the marginal eQTL effect estimates and hence improves the prediction of eQTL effects. We performed extensive preliminary evaluation of the proposed method LEGO. We simulate gene expression level based on the multi-ancestry TOPMed samples and assumed eQTL effect sizes. The extensive simulation studies demonstrate that LEGO maintains calibrated type I error, and powerfully identifies more significant *cis*-eQTLS compared to traditional eQTL mapping methods using linear models, analzying each ancestry separately. With multi-ancestry GWAS datasets, we will further apply LEGO to generate refined eQTL effect estimates and use our recently developed method PUMICE to integrate epigenetic and 3D genome data and predict gene expressions for downstream TWAS analyses. In summary, we propose a local ancestry-aware approach that can improve the eQTL effect estimates, offering the potential to facilitate a more comprehensive understanding of genetic regulation among diverse populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4085T: Leveraging causal cell-type-specific regulatory annotations improves the accuracy of functionally-informed fine-mapping in complex traits

Authors:

A. Kim[1], S. Gazal[2]; [1]Univ. of Southern California, Los Angeles, CA, [2]USC, Los Angeles, CA

Abstract:

While genome-wide association studies (GWASs) have successfully identified thousands of genetic variants associated with human complex diseases, pinpointing the precise causal variants and downstream biological mechanisms remains a core challenge. Fine-mapping techniques have improved our ability to nominate causal variants from GWAS signals, but the power of these methods is limited due to strong linkage disequilibrium between associated variants. Functionally informed fine-mapping (FIFM) methods prioritize variants in functional annotations enriched for trait heritability. However, while heritability is strongly enriched in candidate regulatory elements (CREs) of disease-relevant cell types, current FIFM approaches employ a pre-defined baseline model of cell-type-agnostic annotations, potentially limiting FIFM power and accuracy. To address this limitation, we propose a novel framework (CT-PolyFun) integrating our recent method of causal cell-type-specific CRE inference (CT-FM; Kim et al. 2024 *Medrxiv*) with the state-of-the-art FIFM PolyFun method (Weissbrod et al., 2020 *Nat Genet*). CT-PolyFun relies on the hypothesis that: 1) most complex traits are context-specific, *i.e.*, genetic variants predispose to the trait by acting in a limited number of cell-types; and 2) a prior selection of disease-relevant cell-type-specific CRE annotations can improve the accuracy of FIFM. First, we updated the existing PolyFun baseline model by jointly leveraging recent datasets of regulatory elements including CRE annotations from ENCODE4. Applied to GWAS summary statistics of 11 traits from UK Biobank, our new updated model confidently fine-mapped (PIP > 0.5) a total of 4,664 causal SNPs, representing a slight increase over the baseline model (~3 % increase). Second, we applied CT-PolyFun to the same set of traits. While we observed a slightly increased number of confidently fine-mapped variants (6 % increase), 22 % of these variants were not identified by PolyFun, suggesting that modifying priors mainly modify the set of SNPs that are prioritized. Finally, we leveraged functional data from massively parallel reporter assays (MPRA) and showed that CT-PolyFun identifies a substantially greater proportion of expression-modulating functional variants (39.2% vs. 21.0% for Polyfun), indicating a higher accuracy of CT-PolyFun in fine-mapping causal SNP with clear cellular impact. Our results indicate the relevance of context-specific FIFM and propose a new fine-mapping strategy that improves the identification of functional causal variants, thus helping the translation of GWAS findings into discoveries that will enhance disease treatment.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4086T: Leveraging genetic and proteomic data to decipher smoking-induced pathology ⭐

Authors:

S. Xiao[1], M. Argentieri[2], B. Liu[1], C. Shovlin[3], L. Belbasis[1], R. Mosaoa[4], A. Nevado-Holgado[1], C. Zhengming[1], D. Bennett[1], N. Amin[1], C. Van Duijn[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Imperial Coll. London, London, United Kingdom, [4]King Abdulaziz Univ., Jeddah, Saudi Arabia

Abstract:

Over 8 million people a year die as a result of smoking. Our study aimed to a) construct a machine learning-based proteomic profile of smoking; b) determine its association with mortality and 24 diseases; c) investigate its genetic and exposomal architecture. Using 2,917 proteins of the Olink Discovery panel in UK Biobank (n=43,914), we applied gradient boosting to generate a proteomic Smoking INdex (pSIN) to discriminate current (N=4,732) from never smokers (N=23,778). We found 51 proteins were able to predict smoking status (AUC=0.95; CI:0.94-0.95) in the test dataset. pSIN was externally validated in 3,977 China Kadoorie Biobank participants (AUC=0.90; 95% CI: 0.89-0.92). Proteins selected were differentially expressed in tissues known to be affected by smoking, i.e., lung, salivary gland, colon, esophagus and adipose. A standard deviation increase in pSIN was associated with a 32% increased risk of mortality (95% CI:28-36), and 17 out of 24 major diseases including lung cancer (HR=1.97, CI:1.83-2.10), chronic obstructive pulmonary disease (COPD, HR=1.72, CI:1.67-1.78), and head and neck cancer (HR=1.64, CI:1.44-1.86) adjusting for smoking history. A genome-wide association study of pSIN identified 95 genome-wide significant variants mapped to 129 genes of which 8 *(ALPP, CST5, IL12B, ACVRL1, IL7R, SCGB1A1, NCAM1, ICAM5* and*)* encoded one of the selected proteins. Further, 32 out of 95 variants were cis-pQTLs and 10 genes were previously identified as GWAS smoking loci. Of the 129 genes, 119 (92%) were novel and co-localize with body mass index (BMI), diabetes, immune related traits, and cancer development. LD score regression analysis highlighted strong genetic correlations of pSIN with ever smoking (r=0.78, $p=1.08*10^{-98}$), maternal smoking around birth (r=0.66, $p=2.08*10^{-42}$), smoking related disorders including lung cancer (r=0.71, $p=1.14*10^{-06}$), COPD (r=0.54, $p=1.04*10^{-31}$), and BMI (r=0.33, $p=1.832*10^{-22}$). Overall, genome, exposome (i.e., all exposures measured in UKB) and smoking-related variables (e.g., smoking status, number of cigarettes, pack years) combined explained 69% (CI:68%-70%) of pSIN variability, with genetic factors contributing 22% to the pSIN

prediction. We found that pSIN, a powerful proteomics score discriminating smoking status with high accuracy. PSIN associated to mortality and lung and cardiometabolic disorders independent of self-reported smoking exposure. pSIN was partially genetically driven by variants mapped to *ALPP, CST5* and *IL12B* among others including 119 novel loci. GWAS findings suggested a shared genetic aetiology of smoking with maternal smoking, obesity, diabetes, immune traits, and cancer.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4087T: Leveraging haplotype information in heritability estimation and polygenic prediction

Authors:

J. Meisner[1,2], M. Benros[2], S. Rasmussen[1]; [1]Univ. of Copenhagen, Copenhagen, Denmark, [2]Copenhagen Univ. Hosp., Copenhagen, Denmark

Abstract:

Polygenic prediction has yet to make a major clinical breakthrough in precision medicine and psychiatry, where the application of polygenic risk scores are expected to improve clinical decision-making. Most widely used approaches for estimating polygenic risk scores are based on summary statistics from external large-scale genome-wide association studies, which relies on assumptions of matching data distributions. This may hinder the impact of polygenic risk scores in modern diverse populations due to small differences in genetic architectures. Reference-free estimators of polygenic scores are instead based on genomic best linear unbiased predictions and models the population of interest directly. We introduce a framework, named hapla, with a novel algorithm for clustering haplotypes in phased genotype data to estimate heritability and perform reference-free polygenic prediction in complex traits. We utilize inferred haplotype clusters to compute accurate SNP heritability estimates and polygenic scores in a simulation study and the iPSYCH2012 case-cohort for depression disorders and schizophrenia. We demonstrate that our haplotype-based approach robustly outperforms standard genotype-based approaches, which can help pave the way for polygenic risk scores in the future of precision medicine and psychiatry.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4088T: Leveraging phenotypic, genetic, and environmental correlations between complex traits to improve risk predictions

Authors:

K. Zhang, R. Bierman, J. Akey; Princeton Univ., Princeton, NJ

Abstract:

The massively large number of genome-wide association studies (GWAS) performed to date have identified thousands of loci across the genome that are correlated with phenotypic variation and disease susceptibility. An important application of GWAS results is the development of genetic based predictions of disease risk. Genetic risk predictions can help identify high-risk patients for intervention and guide clinical decision-making. Nonetheless, even traits with high heritability, variants with moderately large effect sizes, and large sample sizes often have risk predictors that perform poorly. Inspired by the widespread pleiotropy observed in GWAS data and the development of large cohorts that have been deeply phenotyped for many traits, we developed an approach to leverage correlated traits to improve the predictive accuracy of risk prediction. To evaluate our approach and the parameter space in which correlated traits improve risk prediction, we built a flexible simulation framework to investigate the performance and behavior of phenotypic, genetic, and environmental correlations between traits. We also investigated how other features, such as the heritability or genetic architecture of traits influenced the accuracy of correlated traits in risk prediction. Furthermore, using the UK Biobank data, we explore whether environmental correlations can be used to develop hypotheses on the shared environmental determinants of phenotypes. We incorporate our correlated trait framework into both linear and non-linear machine learning methods, and show how their performance varies depending on the characteristics of the trait being analyzed. Finally, we empirically compute phenotypic and genetic correlations for metabolic and psychiatric traits in the UK Biobank data and apply our methodology to demonstrate the ability of correlated traits to significantly improve risk prediction in real data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4089T: Leveraging protein interaction networks for powerful rare-variant gene set association analysis

Authors:

S. Kim, S. Lee; Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

To enhance statistical power and interpretation for complex phenotypes, genetic association studies for rare variants require analytical strategies beyond traditional single-variant and gene-based tests. We present a novel approach integrating the gene-based rare variant test, SAIGE-GENE+, and the gene-set analysis method, GAUSS, with protein-protein interaction (PPI) data to enhance rare-variant association analysis. We constructed gene sets from the STRING database, including the top 10 interacting genes for each 19,385 human protein-coding genes. These gene sets are analyzed against a reference panel empirically estimating gene correlations across 1,941 phenotypes from the White British samples of UK Biobank 470k Whole Exome Sequencing data using 18,378 gene-based p-values for each phenotype estimated by SAIGE-GENE+. While SAIGE-GENE+ examines gene-based p-values, GAUSS adopts a subset-based approach, transforming these values into a test statistic to determine gene-set p-values and extract the most influential genes within each gene set.
Simulation studies demonstrate controlled type I error rates and increase in power. In scenarios where a phenotype is associated with several PPI networks, each having a varying number of associated genes, it achieved power gains as the number of associated genes in the network increased—from 0.282 in networks with a single associated gene to 0.973 in networks with five associated genes. Case studies of the UK-Biobank WES data analysis validate our approach, successfully identifying key genes not detected through SAIGE-GENE+. Specifically, our results identified *PCSK9* as associated with disorders of lipoid metabolism (Phecode: ICD.272), and *HSF4* with cataracts (Phecode: ICD.366), which had gene-based p-values of 3.3e-5 and 1.7e-4, respectively, when tested by SAIGE-GENE+. Further network visualization and phenotype clustering based on our results provided clearer insights into potential gene-phenotype connections, with *COL1A1* notably serving as a hub gene for cataracts, rheumatic disease of the heart valves, and osteoporosis. Our study highlights the benefits of integrating SAIGE-GENE+ and GAUSS with PPI data, providing a robust framework for elucidating complex genetic interactions and deepening our understanding of the biological mechanisms underlying various phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4090T: Linkage analyses identify novel loci for Alzheimer's disease in non-Hispanic white families

Authors:

N. Gomez[1], M. Jean-Francois[1], K. Hamilton[1], S. Slifer[1], B. Kunkle[1], M. Cuccaro[1], E. Martin[1], J. Haines[2], C. Reitz[3], M. Pericak-Vance[1], B. Vardarajan[3], G. Beecham[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Case Western Reserve Univ., Cleveland, OH, [3]Columbia Univ., New York, NY

Abstract:

Background: Alzheimer disease (AD) is a progressive dementia with high heritability. While genome-wide association studies have identified common variation associated with AD, most of these loci have effects too small to explain the segregation of disease within multiplex families. As such, these multiplex families likely harbor novel genetic variants with strong effects, and thus still play an important role in assessing the genetic etiology of AD.
Methods: Linkage analyses were conducted on 670 individuals across 70 non-Hispanic white (NHW) families multiplex for AD or mild cognitive impairment (MCI). A combination of genome wide array data and imputed data was used for the analyses. Rare variants, and low-quality variants were excluded; remaining variants were LD pruned. Linkage analyses included both multipoint (MPT) and two-point (2PT) approaches, with non-parametric, parametric (affected only; AO), and parametric (age-penetrance; AP) models. Primary analyses included (1) all AD+MCI individuals and (2) all AD+MCI individuals with age at onset 65 or lower (early-onset AD). Sensitivity analyses included the same models but excluding MCI.
Results: We identified three primary linkage regions (HLOD > 3). For the MPT-parametric-AO model (AD+MCI) we identified chromosome 5q33 (HLOD=3.017; 149MB-156MB). A region on 8p11-p21 (20MB-37MB) was identified using the MPT-parametric-AO model (AD and early-onset models; HLOD=3.09, 3.27). Finally, chromosome 14q22-24 (57MB-70MB) was identified using the MPT-parametric-AP model (AD+MCI; HLOD=3.56). The 5q33 region contains IL12B and ADAM19 genes. IL12B is an inflammatory cytokine that has been associated with amyloid burden and AD affection status. ADAM19 has been associated with AD progression and has functional links to APP. The 8p locus houses the CLU locus, first identified in large AD GWAS, while the 14q locus is near the PSEN1 gene. Further analyses of whole genome sequence data in these regions are ongoing.
Conclusion: These analyses highlighted three regions and several genes connected with Alzheimer's disease. These new loci, along with the established loci identified, may give more insight into the causes of AD for multiplex families.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4091T: localQTL: A GPU-Accelerated Software for Accurate Molecular QTL Mapping in Admixed Populations ⭐

Authors:

K. J. Benjamin[1,2]; [1]Lieber Inst. for Brain Dev., Baltimore, MD, [2]Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Abstract:

Accurately mapping diverse molecular quantitative trait loci (molQTLs) like expression QTLs (eQTLs), methylation QTLs, and protein QTLs in admixed populations is essential for achieving equitable understanding of the genetic underpinnings of complex traits and diseases across all populations. However, existing methods often struggle to account for local population structure, leading to spurious associations and reduced power. Local ancestry information improves molQTL mapping, but existing approaches are computationally expensive. Here, we introduce localQTL, a Python software program that leverages tensorQTL and integrates local ancestry information using the memory-efficient rfmix-reader (similar to pandas-plink) for accurate and efficient molQTL mapping. localQTL uses GPU acceleration, achieving significant speedups compared to CPU-based methods. We demonstrate its effectiveness on gene expression data from admixed Black Americans in the African Ancestry Neuroscience Research Initiative, demonstrating superior performance in eQTL discovery and computational efficiency for admixed populations. By enabling researchers to dissect the complex interplay between genetic variation and molecular traits in admixed populations, localQTL unlocks insights into the underlying genetic mechanisms of diseases in diverse populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4092T: Longitudinal changes in epigenetic age acceleration prior to T1D onset in the Diabetes Autoimmunity Study in the Young (DAISY)

Authors:

K. Bergstrom, H. Smith, R. K. Johnson, L. Vanderlinden, J. Seifert, K. Hohsfield, I. V. Yang, S. Slack, F. Dong, K. Kechris, M. Rewers, J. M. Norris, P. Carry; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

The etiology of type 1 diabetes (T1D) remains unknown but is believed to be influenced by genetic and environmental factors. Epigenetic age, a biomarker of aging, has been associated

with the development and prognosis of many chronic diseases, especially when it is increased relative to chronological age. This difference between epigenetic age and chronological age is termed epigenetic age acceleration (EAA). This study aims to explore longitudinal changes in EAA among individuals at high genetic risk for T1D. We identified 170 high-risk children (85 T1D cases and 85 matched controls) from the Diabetes Autoimmunity Study in the Young (DAISY). DNA methylation (DNAm) from peripheral blood samples was quantified with either the 450k or EPIC Infinium Methylation Beadchips. Longitudinal samples (age range 0.58 to 22.79) were obtained from up to 4 visits per participant, including two visits prior to islet autoimmunity (IA) seroconversion (Median age: 1.25), a pre-clinical phase defined by the appearance of autoantibodies, and two visits after IA seroconversion (Median age: 5.47). The visits were defined as an early visit in infancy, the final visit preceding IA seroconversion, the first visit with IA autoantibodies, and the final visit preceding a diagnosis. EAA was estimated for each visit point from the DNAm data based on Horvath pan-tissue clock, which is appropriate across tissue types and age ranges (including pediatrics). EAA calculations were implemented with the Methylclock R package. Linear mixed models were used to compare average changes in age-adjusted EAA before and after IA seroconversion. We tested whether pre- vs post-IA changes in EAA differed between cases and controls (case*time interaction) after adjusting for sex, HLA-DR3/4 genotype, first two principal components representing genetic ancestry, cell proportions and the type of array used (either 450k or 850k) for methylation quantification. The change in EAA before and after IA seroconversion significantly differed between T1D cases and controls (p=0.02). Among T1D cases, EAA significantly decreased from pre- to post-IA by 0.361 residual years (95% CI: -0.64 to -0.08), p=0.01). Among controls, there was no change in EAA pre- vs post-IA (mean difference: 0.045 residual years, 95% CI: -0.23 to 0.32), p=0.75). EAA changed differently over time in the cases, characterized by decreasing EAA in T1D cases, and no change in EAA among controls. There is a strong need for future work to identify early environmental stressors that are correlated with increased EAA and T1D risk. Evaluation of EAA prior to IA could serve as an indicator of future T1D risk and inform efforts to delay or prevent T1D.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4093T: Mapping aging and disease pathways through integrating single-cell gene expression and plasma proteins

Authors:

Y. Yang[1], C. Wu[2], B. Zhao[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]UT MD Anderson Cancer Ctr., Houston, TX

Abstract:

Proteomics has been widely utilized to understand the structures and functions of various organs within the human body. Numerous studies have explored the organ-level connections between plasma proteins and key biomarkers associated with aging and human diseases. However, the fine-grained associations between proteins and individual cells within each organ remain unclear. In this study, we trained a set of single-cell level models based on the gene expressions within various cell types, including brain cells, lung cells, and blood cells. We identified a variety of plasma proteins that are specific to multiple cell types, such as organ-specific plasma proteins enriched in multiple cell types and multi-organ proteins that share similar cell types across different organs. We demonstrated the utility of these single-cell proteomics findings through two applications. First, we developed cell-type specific protein aging models, which refined organ-level aging signatures to a single-cell level. Second, we identified cell-type specific proteins related to major disease categories, such as brain disorders and cardiovascular diseases. Subsequent enrichment and causality analyses provided more accurate insights into the roles of disease-relevant proteomes. Overall, our work extends protein expressions to the single-cell resolution and highlights its potential in studying organ aging and human diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4094T: Mendelian randomization borrowing strength via shared pleiotropic pathways across populations

Authors:

B. Kang[1], Y. Lu[1], K. Xu[2], L. Chen[1]; [1]The Univ. of Chicago, Chicago, IL, [2]Univ. of Notre Dame, Notre Dame, IN

Abstract:

Two-sample Mendelian randomization (MR) assesses the causal effect of exposure on outcome by utilizing genetic variants as instrumental variables (IVs) and integrating summary statistics from genome-wide association studies (GWASs). Classic MR assumptions are violated when IVs are associated with unmeasured confounders, creating correlated horizontal pleiotropy (CHP). Identifying invalid IVs with CHP is challenging when sample sizes are limited. In this work, we propose a multi-population MR method for jointly estimating the causal effect across two or more populations and identifying IVs with CHP, leveraging at least partially shared pleiotropic pathways across populations. Our approach improves causal estimation and inference in populations with limited sample sizes by borrowing information from populations with larger samples. Using the IVs identified with CHP, we map the cis-genes, which in turn inform

pathways and etiology shared across populations. We applied our method to identify risk factors for complex diseases across multiple ethnic groups. Additionally, we used our method to identify risk factors for AD in the oldest-old, defined here as age above 95, by borrowing information from samples of all ages. Our method can also be applied to analyze multiple correlated outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4095T: Mendelian randomization identifies known and novel regulation between enzymes and metabolites

Authors:

A. van der Graaf, Z. Kutalik; Univ. of Lausanne, Lausanne, Switzerland

Abstract:

Background and motivation
Human metabolic pathway regulation remains poorly understood. Guided by recent improvements in i) Mendelian randomization (MR) methodology, ii) power for metabolite/protein quantitative trait locus (m/pQTL) detection, we perform bidirectional MR between enzymes and their catalyzed molecules to determine the usefulness of MR analyses in understanding pathway regulation.
Methods
Using 4 mQTL studies of up to 54,219 individuals and 3 pQTL studies of up to 118,461 individuals, we apply 3 MR methods: MR-IVW, MR-PCA and MR-link-2, to estimate causality between 566 enzymes and 214 metabolites involved in 1,070 reactions derived from 2 independent pathway databases. We assess enrichment for pathway reactions and interpret metabolite-enzyme pairs linked through catalytic reactions and pathway regulation.
Results
Bonferroni significant MR results ($P < 3x10^{-7}$) are enriched for true reactions (odds ratio 15.4, P: $5x10^{-10}$). Here, we report MR-link-2 results in depth as precision is highest (11%, max 3% for other methods).
Out of the 66 significant enzyme-to-metabolite links 11 are true positives according to pathway databases. 5 estimates are known reactions not in the reference. MR correctly identifies a rate limiting step in glycolysis between GADPH and the glycolysis product: pyruvate (α: 0.60, P: $1.5x10^{-7}$). ALPI protein abundance influences 7 metabolites that are strongly enriched for ammonia recycling in humans (P: $3x10^{-13}$), suggesting a novel regulatory role for ALPI abundance.

Out of the 38 significant metabolite-to-enzyme MR estimates, 5 metabolites influence their catalyzing enzyme. For instance: n-acetyl-aspartyl-glutamate and its catalyzing enzyme glutamate carboxylase II (FOLH1) (α: -0.09, $P < 4.3x10^{-39}$), implying a sensing mechanism that maintains reaction steady state. We also identify (α: -0.1, P: $2.9x10^{-7}$) a causal relationship between bilirubin and cholinesterase (BCHE) suggesting a mechanism that corroborates how choline supplementation reduces neuronal damage from hyperbilirubinemia.

Conclusion

Significant MR results are strongly enriched for metabolic reactions and highlight known and novel regulation, indicating that MR is a promising orthogonal analysis technique to understand metabolite pathway regulation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4096T: Mendelian randomization of sex-dependent traits

Authors:

E. Sanders, L. Strug; Univ. of Toronto, Toronto, ON, Canada

Abstract:

Mendelian Randomization (MR) is a method that leverages genetic instruments to identify causal exposure-outcome relationships while overcoming common issues of confounding. This can guide investigations of biological mechanisms driving a disease. Two-sample MR has been implemented frequently in the recent literature, conducted using statistics from two independent studies such as genome-wide association studies (GWASs), one for the exposure and one for the outcome. However, sex-specificity in exposure-outcome relationships can pose challenges when applying MR, as state-of-the-art methods produce a single effect estimate per genetic locus. This leaves researchers with the option of either performing MR marginally by sex or disregarding the possibility of sex-specific associations, both of which can reduce power. To address this, we propose novel extensions to both single-locus and multi-locus MR techniques such as MR-Egger and the median- and mode-based estimates. Our methods are based on constructing underlying models incorporating sex-specific genotype-outcome/genotype-exposure relationships, and jointly estimating more than one effect parameter of interest. We investigate via simulation how these novel MR implementations perform relative to typical approaches, investigating situations with and without sexual dimorphism and situations with typical assumption violations such as horizontal pleiotropy or incorrectly chosen genetic instruments. We observe that the novel approach introduces some unmeasured confounding and heteroskedasticity, but that no bias is introduced to the main effects of interest and that

power is comparable to sex-stratified analyses. Our novel methods offer researchers a convenient implementation of MR that can identify sex-specificity, with no need to stratify analyses by sex and without sacrificing power.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4097T: Meta-SAIGE: Scalable and accurate meta-analysis for rare variants

Authors:

E. Park[1,2], K. Nam[2], S. Jeong[2,3], K. Keat[3], D. Kim[3], V. Bansal[4], W. Zhou[5], S. Lee[2]; [1]Univ. of California San Diego, La Jolla, CA, [2]Seoul Natl. Univ., Seoul, Korea, Republic of, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Univ California, San Diego, La Jolla, CA, [5]Massachusetts Gen. Hosp., Chestnut Hill, MA

Abstract:

The growing quantity of sequencing data from large biobanks necessitates effective meta-analysis methods to enhance the statistical power of rare variant association studies. However, existing methods, such as MetaSTAAR, struggle to control type I error rates when analyzing low-prevalence binary traits and are computationally intensive to analyze many phenotypes due to the requirement of separate linkage disequilibrium (LD) matrices for each phenotype. To address these challenges, we introduce Meta-SAIGE, a meta-analysis method built on SAIGE-GENE+, a set-based rare variant test. To control type I error rates, Meta-SAIGE employs two-level saddlepoint approximation (SPA): SPA on the score statistics of each cohort and a genotype count-based SPA for combined score statistics from multiple cohorts. Additionally, Meta-SAIGE allows the use of a single sparse LD matrix for the covariance matrix of score statistics across all phenotypes, significantly reducing computational costs when conducting phenome-wide analyses involving hundreds of phenotypes. Simulation studies using UK Biobank whole exome sequencing data have shown that Meta-SAIGE can effectively control type I errors for imbalanced binary phenotypes with a prevalence of 1%, whereas the unadjusted method exhibited 100-fold higher type I error rates. Power simulations illustrate that Meta-SAIGE outperforms the weighted Fisher method, which simply combines SAIGE-GENE+ p-values from each cohort. A computational cost analysis using the TTN gene demonstrated that Meta-SAIGE significantly outperforms Meta-STAAR in both CPU time and memory usage. When applied to a whole exome meta-analysis of UK Biobank and All of Us data for 83 low-prevalence disease phenotypes in European samples, Meta-SAIGE identified 231 associations at the exome-wide significance level of 2.5E-6. Among these, 94 were novel associations that were not significant in

either the UK Biobank or the All of Us data alone. These results underscore Meta-SAIGE as a valuable tool for researchers exploring the role of rare variants in complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4098T: mixTWAS: a framework to identify genes from GWAS using mixed effects to account for relatedness and polygenicity-driven inflation

Authors:

S. Mi[1], A. Sanchez Roige[2], A. Palmer[3], H. Im[4]; [1]UC San Diego, San Diego, CA, [2]Univ. of California, San Diego, La Jolla, CA, [3]Univ. of California San Diego, La Jolla, CA, [4]Univ. of Chicago, Chicago, IL

Abstract:

Genome-wide association studies (GWAS) have implicated specific alleles and genes as risk factors for numerous complex traits. However, translating GWAS results into biologically and therapeutically meaningful discoveries remains extremely challenging. Most GWAS results identify noncoding regions of the genome, suggesting that differences in gene regulation are the major driver of trait variability. To better integrate GWAS results with gene regulatory polymorphisms, we previously developed PrediXcan (also known as “transcriptome-wide association studies” or TWAS), which maps SNPs to predicted gene expression using GWAS data. However, current implementations of TWAS do not correct for relatedness among individuals nor inflation caused by polygenicity of the complex traits (Liang et al BioRxiv 2023). To address this limitation, we developed mixTWAS, a mixed-effects model that accounts for high relatedness and polygenicity with a random effect term. To address the computational challenges of fitting mixed models, we transform the phenotype and predicted expression so that the error terms are uncorrelated, then estimate effect size using linear regression. We decorrelate the non-independent error term by multiplying the trait and predicted gene expression with the inverse square root of the covariance matrix. To demonstrate that our approach yields calibrated type I error, we performed a simulation study on 100 null phenotypes and performed the traditional and mixed effects model association. When association between gene expression to simulated phenotypes is performed with mixed effects modeling, we observe uniformly distributed p-values, correcting the inflation resulting in traditional methods. Thus, mixTWAS is a computationally feasible method for inferring the relationship between gene expression and phenotypes while accounting for relatedness and polygencity-driven inflation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4099T: Multi-ancestry gene-based association tests identify novel loci and drug targets in minority populations

Authors:

N. Lorincz-Comi[1], Y. Hou[1], F. Cheng[2]; [1]Cleveland Clinic Fndn., Cleveland, OH, [2]Cleveland Clinic, Cleveland, OH

Abstract:

Background Over-representation of European populations in GWAS has led to the delayed discovery of disease-associated genes in minority populations. We present the first gene-based association test which leverages shared association evidence from multiple populations to identify disease-associated genes and demonstrate that our method is more powerful for gene discovery in minority populations than existing approaches.
Methods The Multiple-ancestry Gene-based Test (MuGenT) provides a general statistical framework to perform a weighted joint test of association between a gene-specific set of genotypes and a disease in multiple populations. As a case study, we used MuGenT to study type 2 diabetes (T2D) in African American (n=153K), European (n=1.8M), Hispanic (n=88K), and East (n=427K) and South (n=2.5K) Asian cohorts using public GWAS summary data. We tested 18,257 genes and used ancestry-specific LD reference panels from 1000 Genomes Phase 3. Unlike existing gene-based tests, the MuGenT test has a closed-form null distribution and so requires less than one second per gene. When available, MuGenT can also integrate population-specific xQTL information to perform a multivariate causal assessment using the same principles as Mendelian Randomization (MR).
Results We identified 49 new genes associated with T2D at the FDR-corrected level of significance that were not within 50kb of any SNPs associated with T2D at genome-wide significance in any ancestry-specific GWAS. Of the 49 new genes, 48 (98%) were replicated (P &lt 5E-8) in a multi-ancestry GWAS which combined individuals from all cohorts. In total, we identified 3 new genes whose protein products are actionable by existing pharmacotherapies (ITPR3, TNIK, PPP3CA), but which were not detected in any population-specific GWAS for T2D. TNIK is associated with body mass index and regulates lipid and glucose metabolism in mice, suggesting it may be a therapeutic target for humans.
Conclusions We used MuGenT to detect novel disease-associated genes in minority populations. While we used T2D as a case study here, we will present additional results with similar findings for gene associations with the plasma abundance of 2,923 proteins and highlight those encoded by known Alzheimer's disease risk genes such as APOE, BIN1, and CLU in African,

European, American, East and South Asian, and Middle Eastern cohorts from the UKB-PPP. We conclude by advocating for the use of multi-population methods as the new analytical standard in gene discovery and propose the MuGentT framework as a new strategy for discovering population-shared and population-distinct risk genes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4100T: Multi-ancestry meta-analysis identifies genetic modifiers of age-at-onset of Alzheimer's disease

Authors:

E. Blue[1], J. Broome[1], D. Xue[1], S. Gogarten[1], A. Naj[2], E. Wijsman[3]; [1]Univ. of Washington, Seattle, WA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Univ Washington Sch Med, Seattle, WA

Abstract:

Background: While Alzheimer's disease (AD) is highly heritable, a significant proportion of AD risk is explained by age, ancestry, *APOE* genotype, and sex. Genome-wide association study (GWAS) signals can vary dramatically in strength, direction of effect, or both when studies vary by these factors. GWAS of age-at-onset of AD (AAO) can offer increased statistical power relative to a typical case/control design. Despite their smaller numbers and sample sizes, AAO GWAS have replicated 20 AD risk loci and nominated three novel signals near genes associated with differential expression or neuropathology in AD. We sought to investigate the relationship between sex, *APOE*, and genetic modifiers of AAO in data sets representing diverse ancestry. **Methods:** We cleaned and imputed to TOPMed a discovery cohort of 16 publicly available AD GWAS datasets, while the independent Alzheimer's Disease Genetics Consortium (ADGC) data underwent parallel quality control and TOPMed imputation. Within each data set, we performed relatedness estimation, principal components (PCs) analysis, local ancestry inference, and GWAS. GWAS with and without adjustment for sex and APOE genotype were contrasted and meta-analyzed, with genome-wide significance set to $P<5\times10^{-5}$.**Results:** Subjects with >20% non-European ancestry were well represented in both the discovery (44% of 21,736 subjects) and ADGC data (36% of 19,483 subjects). Covariate adjustment for sex and sparsifying the genetic relatedness matrix had minimal effects on GWAS effect sizes (discovery $r^2>0.96$, ADGC $r^2>0.98$), whereas *APOE* adjustment influenced effect sizes beyond chr19 (discovery $r^2>0.71$, ADGC $r^2>0.77$). Meta-analysis of GWAS adjusted for sex, PCs, and relatedness revealed 13 significant loci, including one novel signal (9q21.32) and 12 at known AD risk loci (*APOE, BIN1, CASS4, CD2AP, CR1, EPHA1, MS4A6A, PICALM, PILRA, SHARPIN, WDR81, WWOX*). Meta-analysis of GWAS adjusting for APOE revealed 17 significant loci, including four novel signals

(9p22.3, 9q21.33, 11q13.1, 18p11.21) and 13 at AD risk loci (*APOE, BIN1, CASS4, CD2AP, CR1, EPHA1, HAVCR2, MS4A6A, MYO15A, PICALM, PILRA, SHARPIN, UMAD1*). **Conclusion:** Our results reinforce the power of an AAO approach by identifying novel signals and replicating several signals only recently discovered in GWAS representing hundreds of thousands of cases and controls with predominantly European ancestry. GWAS signals like ours that are robust to diverse ancestry are likely to be more precise given population differences in linkage disequilibrium, and therefore better able to infer individual-level risk regardless of ancestry.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4101T: Multi-group causal-TWAS provides a unified framework to integrate multi-omic QTL data from multiple tissues with GWAS ⭐

Authors:

S. Qian[1], K. Luo[1], X. Sun[1], W. Crouse[1], S. Zhao[2], M. Stephens[1], X. He[1]; [1]Univ. of Chicago, Chicago, IL, [2]Dartmouth Coll., Lebanon, NH

Abstract:

One common strategy to annotate variants from GWAS and to prioritize risk genes is the use of molecular QTL, especially eQTL data. However, commonly used methods such as TWAS and colocalization are susceptible to pleiotropic effects of variants and have high false positive rates. Our recently developed method, causal-TWAS (cTWAS, Zhao et al. Nature Genetics, 2024), uses an extended fine-mapping framework to jointly model variants and gene expression in a GWAS region, leading to better control of false discoveries. Nevertheless, cTWAS was designed to analyze a single eQTL dataset. Growing evidence suggested limited power of eQTLs from a single context to explain GWAS signals. As such, it is important to extend the analysis to QTLs from multiple modalities (e.g. expression, splicing) and contexts (tissues/cell types). Importantly, the joint analysis may allow one to determine the molecular mechanisms and contexts through which genetic variations act on phenotypes.

We introduce a new method, multi-group cTWAS, extending causal-TWAS to this multi-omic and multi-tissue setting. Our model allows us to share important parameters across contexts, improving accuracy of inference. It combines information of the same genes across modalities and contexts to increase power. Additionally, we improved the computational efficiency of the software by more than 10-fold, allowing it to handle large numbers of molecular traits.

Our simulations demonstrated the accuracy of the multi-group cTWAS in identifying causal genes and causal contexts. We applied it to GWAS of multiple common phenotypes, using several QTLs (expression, splicing, alternative polyadenylation - APA) across multiple tissues

from GTEx. We found that while molecular QTLs from single tissues explain approximately 15-20% of the heritability of traits, using cross-tissue analysis increases this to 30-50%. Overall, the proportions of heritability explained by splicing and APA-QTLs are comparable to eQTLs. We discovered 30-200 candidate genes, at least doubling the number of candidate genes from single-tissue eQTL analysis. In over 80% of cases, we were able to assign a single modality to the causal genes, and in over 50% of cases, the most likely causal tissues. Our analysis found many genes that would have been missed by single-tissue eQTL analysis, including some known risk genes, such as APOB, LDLR, and HMGCR for the LDL phenotype.
In conclusion, the multi-group cTWAS method solves a fundamental challenge of integrative QTL-GWAS analysis, greatly reducing false positives. It also provides a powerful framework to incorporate complex molecular QTL data with GWAS to reveal the genetic basis of complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4102T: Multi-omics integrative risk model for Alzheimer's Disease in a large-scale biobank

Authors:

R. Venkatesh, A. Moore, K. Cardone, R. Kumar, Y. Bradford, M. Ritchie; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Alzheimer’s Disease (AD) is the most prevalent condition that impacts the aging population, with no effective treatment or singular underlying causal factor identified. As a complex disease, characterizing the genetic risk of developing AD has proven to be difficult; most predictive risk models exclusively utilize genetic variants identified from genome-wide association studies (GWAS) and achieve poor performance. However, tissue-specific gene expression and protein expression data are shown to provide orthogonal and complementary information to genetics alone, suggesting that a multi-omics approach may yield more comprehensive risk prediction. As such, this study constructed comprehensive integrative risk models (IRM) for AD by incorporating imputed transcriptomic and proteomic associations. A GWAS of AD was conducted using SAIGE on individuals from the Alzheimer’s Disease Sequencing Project v11 (ADSP), a dataset of 10,167 individuals with AD (cases), and 13,967 controls. Transcriptome-wide and proteome-wide association studies (TWAS/PWAS) were performed using PrediXCan to investigate genes and proteins associated with AD. Genetically regulated gene expression and protein expression were imputed using individual-level genotypes from ADSP, transcript values

(eQTLs) across 49 tissues from the Genotype-Tissue Expression (GTEx) v8, and plasma protein quantitative trait loci (pQTLs) from TOPMed MESA and the Atherosclerosis Risk in Communities study (ARIC). Preliminary GWAS and TWAS results in multi-ancestry ADSP, and a subset of 1,196 ADSP participants of European ancestry from the Religious Orders Study/Memory and Aging Project (ROSMAP) implicated variants in *APOE* and *APOC1* as genome-wide significant at a p-value threshold of 5E-8, and *DAPK2, MVK,* and *NXT1* at a suggestive transcriptome-wide Bonferroni p-value threshold of 1.5E-5 in ROSMAP. Initial PWAS in ROSMAP and ADSP identified proteins VEGFA, WFIKKN2, and CTSB at a suggestive proteome-wide Bonferroni p-value threshold of 3.3E-3. As such, known and putatively novel genomic, transcriptomic, and proteomic associations were identified via G/T/P-WAS for AD risk from ADSP, demonstrating the utility of imputed multi-omics data. Using summary statistics, imputed individual-level gene and protein expression, and clinical covariates of age, sex, and principal components for ancestry, we construct and train elastic-net logistic regression models to integrate imputed multi-omics data to predict genetic risk of AD. We further characterize precision and accuracy of multiple iterations of such models and compare performance of the multi-omics IRMs against a baseline risk model.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4103T: Multi-trait inference of full genome-wide associations of type 2 diabetes subtypes informs distinctive biology and putatively causal genes ⭐

Authors:

S. Yoshiji[1], O. Ruebenacker[1], P. Smadbeck[1], T. Nguyen[1], A. Sayici[1], C-Y. Su[2], **E. Llamas**[3], B. Richards[4], J. Florez[5], M. Udler[1], J. Flannick[1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]McGill Univ., Montreal, QC, Canada, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada, [5]Massachusetts Gen Hosp, Boston, MA

Abstract:

Background: Type 2 diabetes (T2D) is a heterogeneous disease, and clustering of T2D genome-wide association study (GWAS) variants has suggested multiple genetic clusters of T2D with distinctive characteristics (hereafter referred to as T2D subtypes). However, the number of known variants for each subtype is limited, and their GWAS cannot be directly conducted because they are latent endophenotypes. We propose a novel method to generate full GWAS for latent subtypes of complex traits through multi-trait Bayesian inference. Applying it to T2D, we significantly expand our understanding of T2D subtypes.

Methods and Results: We developed a Bayesian method to infer full genetic associations of latent T2D subtypes using observed summary statistics for 41 T2D-related metabolic traits. The method consists of training and classification. During training, we used genetic variants previously associated with three diabetes subtypes identified through clustering of T2D GWAS variants after excluding outliers: 83 for obesity, 42 for beta-cell, and 30 for lipodystrophy. Conditional on observed associations of these variants with 41 T2D-related traits, the model learned a latent "obesity" pattern (e.g., positive effect for BMI and waist-hip ratio (WHR)), a "lipodystrophy" pattern (negative for BMI but positive for WHR), and a "beta-cell" pattern (negative for fasting insulin). In classification, we used the trained model to infer genetic associations of 2.3 million variants with each subtype. We identified 387 subtype-associated variants at $P < 5 \times 10^{-8}$. Variant-to-gene mapping with Open Target V2G and cS2G scores, with further filtering, identified 375 subtype-specific causal gene candidates. For instance, for the lipodystrophy subtype, we captured known genes (*IRS1*, *COBLL1*, and *PPARG*), as well as novel ones such as *TGFB2*, which codes for an adipokine hormone. Pathway analyses showed enrichment of lipid-related pathways for the obesity and lipodystrophy subtypes, and insulin secretion-related pathways for the beta-cell subtype. Moreover, plasma proteome-wide Mendelian randomization found 20 subtype-specific causal proteins, including BCL2 for lipodystrophy subtype ($P = 8.8 \times 10^{-6}$), offering potential circulating therapeutic targets.
Conclusions: Using a novel multi-trait Bayesian inference method, we inferred full GWAS of T2D subtypes and identified putatively causal genes and plasma proteins. We also found enrichment of relevant pathways. The method is applicable to any complex trait in any genetic ancestry with a training set of variants and GWAS of relevant traits, opening avenues to disentangle genetic heterogeneity of complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4104T: MV-DeLIVR: deep learning based multivariable instrumental variable regression for nonlinear causal inference

Authors:

R. He[1], C. Li[2], Z. Lin[3], X. Shen[4], W. Pan[5]; [1]Univ. of Minnesota, Twin Cities, Minneapolis, MN, [2]Iowa State Univ., Ames, IA, [3]Florida State Univ., Tallahassee, FL, [4]Univ. of Minnesota Twin Cities, Minneapolis, MN, [5]Univ. of Minnesota, MINNEAPOLIS, MN

Abstract:

Instrumental Variable (IV) regression is a foundational approach to causal inference, particularly in contexts where randomized experiments are not feasible. An important and novel application

of IV regression is to Transcriptome-Wise Association Studies (TWAS) to identify causal genes (as exposures) for a trait (as an outcome) using genetic variants as IVs. However, current TWAS applications are exclusively based on either linear models or univariable methods, overlooking potential nonlinear gene-trait relationships or pleiotropic/confounding effects of genetic variants due to linkage disequilibrium. In this paper, we introduce MV-DeLIVR, a robust multivariable IV regression method based on a neural network, accounting for invalid IVs (e.g. due to horizontal pleiotropy of genetic variants) while estimating possibly more complex exposure-outcome (e.g. gene-trait) associations. We offer theoretical support for the proposed method, demonstrating estimation consistency of the neural network and selection consistency for invalid IVs. We apply the new method to the GTEx gene expression data and UK Biobank GWAS data to conduct simulations and perform real data analyses, focusing on high-density Lipoprotein cholesterol (HDL) as the target trait. MV-DeLIVR significantly reduced false positives and improved statistical power over those from univariable models. For example, gene *CDK2AP1*, *C12orf65*, and *RP11-282O18.3* were all deemed significantly associated with HDL by univariable models; however, these genes share many SNPs as IVs, resulting in their highly correlated expression levels. Consequently, univariable models may incorrectly identify all of them as significant. Additionally, some IVs were invalid in univariable analysis due to LD. In contrast, after adjusting for multiple genes, MV-DeLIVR identified only *RP11-282O18.3* as the significant gene. MV-DeLIVR also uniquely identified additional genes like *MT2A*. This increased statistical power arises because including other significant genes in the multivariate model reduces the residual variance, amplifying subtle signals overlooked by univariable models. Our results demonstrate that MV-DeLIVR substantially reduces false positives and possesses high power across various scenarios, offering significant improvements over existing univariable methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4105T: Network Construction using Sparse Gaussian Graphical Model based on GWAS Summary Statistics

Authors:

M. Subedi[1], X. Cao[2], B-J. Kim[1], Q. Sha[1]; [1]Michigan Technological Univ., Houghton, MI, [2]Columbia Univ., New York, NY

Abstract:

In genome-wide association studies (GWAS) thousands of genetic variants are tested to identify the association of genetic variants with the phenotype. GWAS have identified many strongly

associated genetic variants with a phenotype and have greatly enhanced our understanding of the genetic architecture of complex phenotypes and diseases. Joint analysis of multiple phenotypes can increase the overall statistical power to detect genetic associations and allow for the identification of pleiotropic loci. A phenotype-phenotype network (PPN) represents phenotypes as nodes and the relationship between them as edges, which allows for the visualization of complex relationships between phenotypes, making it easier to identify clusters and help us intuitively grasp how different phenotypes are related. In this study, we propose a new method to construct a PPN using the sparse Gaussian Graphical Model (sGGM) based on GWAS summary statistics. This approach will isolate the direct relationship between phenotypes, making it easier to identify clusters of phenotypes conditional on other phenotypes that reflect the more meaningful biological or functional connections. Then we apply the community detection method to partition phenotypes into disjoint modules based on the partial correlation matrix of phenotypes. For each module, various multiple phenotype test methods can be employed to test the association between a SNP and phenotypes in that module. We conducted a comprehensive simulation study to compare the performance of several multiple phenotype association tests using sGGM, network modules obtained from the correlation matrix of phenotypes, and those based on all phenotypes, respectively. The simulation results show that all multiple phenotype test methods can control type I error rate well. Additionally, the result indicates that most multiple phenotype tests based on sGGM exhibit higher power than those derived from the correlation matrix, as well as those considering all phenotypes without modular segmentation. The proposed method has been currently applied to a set of diseases from the UK Biobank. We anticipate that applying this method to the real data will substantiate these findings and further illustrate the method's practical utility.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4106T: NeuroID: Target discovery engine for Parkinson's disease

Authors:

M. Chao, Y. Li, F. Rapino, F. Rapaport, J. Bonner, S. Dujardin, J. Huang, G. Gaglia, S. Lessard, C. Fulco, H. Wesseling, T. Ahn-Horst, D. Liu, H. He, T. Iyyanki, Z. Du, E. de Rinaldis, S. Khader, R. Krishnan, C. Kayatekin, C. Chatelain; Sanofi, Cambridge, MA

Abstract:

We developed and evaluated an advanced modular target discovery platform, the NeuroID engine, for prioritizing therapeutic targets of Parkinson's disease (PD). The NeuroID engine

centralizes multi-omics data including human genetics of PD risk and progression (53,858 patients), brain-specific epigenetics (ABC maps, Hi-C), bulk transcriptomics (from 51 selected studies), single-cell RNA-seq, CRISPR screens (5 studies), proteomics (CSF/plasma), pathway/network biology and text mining. Integrative analysis of multi-omics data was performed by combining variant annotation approaches, MR/coloc of GWAS signals with molQTLs/ABC maps, Hi-C-coupled multi-marker analysis of genomic annotation (H-MAGMA), differential gene/protein expression to identify putative causal genes. A specific machine learning model was created to prioritize PD targets, trained using this vast integrated data repository and benchmarked against a list of prioritized targets by the Michael J. Fox Foundation (MJFF). We demonstrated superior performance in predicting MJFF-prioritized PD targets over other target discovery systems, such as Open Targets. This heightened performance is the consequence of comprehensive integration of various omics types, quality of the data integrated and innovative analytical methods. The NeuroID engine represents a major advancement in data driven target discovery of PD. With its proven ability to predict relevant therapeutic targets, we expect it to uncover novel therapeutic opportunities, marking a significant leap in the fight against PD as well as other neurological diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4107T: New insights into kidney function: GWAS analysis of MRI-derived kidney volumes

Authors:

S. Monteiro-Martins[1], O. Borisov[1], Y. Li[1], E. Kellner[2], P. Sekula[1], M. F. Russe[2], W. Reichardt[2], M. Reisert[2], A. Köttgen[1,3]; [1]Inst. of Genetic Epidemiology, Faculty of Med. and Med. Ctr. – Univ. of Freiburg, Freiburg, Germany, [2]Dept. of Radiology, Faculty of Med. and Med. Ctr. – Univ. of Freiburg, Freiburg, Germany, [3]Johns Hopkins Univ., Baltimore, MD

Abstract:

Genome-wide association studies (GWAS) of kidney function rely on biochemical biomarkers like serum creatinine or cystatin C to estimate the glomerular filtration rate (eGFR). However, these biomarkers have limitations, including different eGFR equations, biological variation, and difficulty distinguishing loci affecting kidney function from those influencing biomarker metabolism. This study explores the genetic architecture of imaging-derived kidney volumes as potential alternative markers of kidney function. We utilized Dixon abdominal MRIs from 38,816 UK Biobank (UKB) participants of European ancestry. Total kidney volume (TKV), cortex, medulla, and sinus volumes were calculated from automatically segmented kidneys using

Patchwork, a hierarchical, multiscale U-net based CNN framework trained with binary cross-entropy loss, with ground truth established by experienced radiologists. The volumes were normalized to body surface area and inverse normal transformed. GWAS was performed using linear regression implemented in regenie (v3.2.9), testing imputed UKB genotypes with a minor allele frequency > 1%. Analyses were adjusted for age, age-squared, sex, assessment centre, and the first 10 genetic principal components. The most probable causal gene for each locus was prioritized using a developed annotation pipeline incorporating gene proximity, VEP, genetic colocalization with cis-eQTL and -pQTLs, and variant linkage disequilibrium. Locus overlap between volumes and marker-based kidney function was assessed with colocalization analyses (genepicoloc). GWAS followed by conditional analysis identified 36 significant ($p < 5e-8$) independent loci for TKV, 27 for medulla, 27 for cortex, and 71 for sinus volumes. Pairwise colocalization between volumes (PPH4 > 0.8) revealed 7 loci specific to medulla and 57 to sinus. Additionally, we observed colocalization with biomarker-based kidney traits and related traits like hypertension and diabetes. We found that 21 of 36 TKV loci colocalized with eGFR and 10 TKV loci colocalized with the urine albumin-creatinine ratio. For example, the eGFR-decreasing G allele at rs77924615 (*UMOD)* was also associated with lower TKV ($p = 1.5e-21$). Gene Ontology enrichment analysis indicated that the prioritized genes are involved in growth and development, particularly kidney development. These findings enhance our understanding of the genetic architecture of kidney volumes and suggest that developmental genes influence organ volume and related functions among adults. Notably, our study revealed novel genetic loci unreported in former association studies of biochemical markers of kidney function

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4108T: Optimal Polygenic Risk Score Ancestry Adjustment Occurs Within Individualized Neighborhoods

Authors:

S. Kulm[1], G. Busby[1], J. Kintzle[1], A. Bolli[2], P. Di Domenico[1], G. Botta[1]; [1]Allelica, New York City, NY, [2]Allelica, New York, NY

Abstract:

Humanity's geographic dispersal led to genetic differences between groups that are related to these historical movements. Some of these differences are adaptive but many of the genetic differences observed today are the result of genetic drift and neutral mutational processes. One consequence of these differences relates to the distribution of polygenic risk scores (PRSs) across groups, and the now well established observation that PRS values in one group are likely

to differ from another group due to these underlying genetic differences.
Various methods have been proposed to remove the influence of geographic stratification on PRSs. One commonly used method groups individuals by their similarity to a predefined genetically homogenous group (for example one of the 1000 Genome Project superpopulations). Then within this group principal components (PCs) mean-standardize the PRS values, which are in turn used to assess risk within a genetically similar group with clinical outcome information. Alternatively, the group used in these two primary steps - PC-based adjustment and risk estimation - can comprise a larger set of individuals, for example based on a global population. While effective, this method requires assigning individuals to discrete continental-sized groups, which may not reflect the genetic variation of many individuals.
We hypothesized that a better approach to identifying a group on which to adjust and assess risk might be better defined by proximity in genetic PC-space. We tested this hypothesis by assessing the predicted relative risk from 10 PRSs of common diseases in a cohort of 100,000 ancestrally diverse individuals in the UK Biobank. Calibration of the resulting models was tested using Brier Skill Scores (BSS) and was on average positive - and therefore better calibrated - when comparing models based both on the PC-neighborhood to the continental approach (median BSS= 1.245e-03) and the global approach (BSS = 3.450e-04). We further observed that considering both the variance and the mean of the PCs during the standardizations step further improved accuracy of the PC-neighborhood approach, (median BSS = 6.826e-06). While not necessarily easy to interpret, these BSS values represent that with just a small modification to the PRS modeling process, over 100 individuals can be prevented from risk misclassification. These results indicate that by eliminating discrete population descriptors, which don't represent humanity's true continuous diversity, the full utility of PRSs can be realized.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4109T: Pairwise Likelihood with Latent Variables for Detecting Genetic Variants Associated with Human Complex Traits Using Multiple Types of Phenotypes

Authors:

M. Zhou[1], K. Zhang[2]; [1]Michigan Technological Univ., Houghton, MI, [2]MICHIGAN Technological Univ, Houghton, MI

Abstract:

With the availability of multiple types of phenotypes in genome wide association studies (GWAS), many methods utilizing multiple phenotypes have been developed to identify genetic

variants associated with human complex diseases and have been shown to be more powerful than single phenotype methods. However, most of existing methods cannot be easily to be extended to handle multiple types of phenotypes that contain both continuous and ordinal phenotypes, multiple types of covariates, and multiple genetic variants. For example, some existing methods have limitations in either treating all ordinal phenotypes as continuous phenotypes, or imposing linearity assumptions between phenotypes and genotypes, thereby sacrificing ordinal nature and efficiency.

In this paper, we propose a method that allows to jointly model multiple genotypes, multiple types of phenotypes, and multiple types of covariates efficiently. Our method is based on the pairwise likelihood and the latent variable model with several advantages over existing methods. First, our method can preserve the ordinal nature of ordinal phenotypes, leading to a more accurate depiction, as opposed to treating them as continuous phenotypes. This preservation has the potential to significantly increase statistical power. Second, our approach eliminates the need for continuity assumptions on genotype values and avoids imposing linearity assumptions between phenotypes and genotypes, departing from traditional models. This independence from regression assumptions enhances flexibility. Third, the method explores the covariance matrix concept, comprehensively examining relationships among genotypes, phenotypes, and covariates, a departure from prior methods with limited evaluations of this complex interaction. Fourth, the method can be used to handle a large number of phenotypes and genetic variants without much loss of efficiency.

To evaluate the proposed method, we conducted comprehensive simulations to assess its type I error and power with different sample sizes, minor allele frequencies, and number of phenotypes and covariates. Simulation results have shown that the type I error rates of the proposed methods are well controlled. Comparing with several existing methods, including genetic association tests between SNPs one at a time and multiple phenotypes (Multiphen), Cauchy combination test (CLM and Cauchy), and multivariate ordinal regression models (Mvord), our method exhibits superior power in most situations. We also applied our method to the Genetics of Kidneys in Diabetes (GoKinD) study and have identified novel genetic variants that are associated with type 1 diabetes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4110T: pecotmr: A unified framework for integration of genome wide association studies with molecular quantitative trait loci

Authors:

D. Nachun[1], T. Edwards[2], A. Wang[3], O. Olayanka[4], H. Sun[5], R. Feng[5], C. Liu[6], D. R. Lamb[5], E. Marcora[2], A. Goate[2], X. Zhang[4], F. Morgante[6], t. ADSP FunGen-AD Consortium[7], S. B. Montgomery[1], G. Wang[5]; [1]Stanford Univ., Stanford, CA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Univ. of Hong Kong, Hong Kong, Hong Kong, [4]Boston Univ., Boston, MA, [5]Columbia Univ., New York, NY, [6]Clemson Univ., Greenwood, SC, [7]Natl. Inst. of Aging, Bethesda, MD

Abstract:

Integrating genome wide association studies (GWAS) with molecular quantitative trait loci (QTL) is a widely used approach to identify candidate molecular targets and mechanisms that underlie complex traits and diseases. These integration approaches rely on the relationship between GWAS and QTL - while a GWAS associates common variants with a complex trait, a molecular QTL associates these variants with a molecular trait. In the *pecotmr* R package, we have provided a unified probabilistic framework with a single, easy-to-use interface for four types of integration analyses: heritability enrichment, colocalization, transcriptome wide association studies (TWAS) and Mendelian randomization (MR). Each method requires appropriate adjustment for linkage disequilibrium, which can greatly inflate test statistics and posterior probabilities. Within *pecotmr*, we extend the sum of single effects regression (SuSiE) for fine-mapping to obtain posterior effect sizes, standard errors and inclusion probabilities from individual or summary-level molecular QTL data, to perform heritability enrichment informed colocalization analysis and to refine instrumental variables for MR. For TWAS, we further implemented 11 individual level and 11 summary level penalized regression methods for estimating molecular QTL weights as alternatives to fine mapping. We have also integrated novel approaches for adjusting SuSiE output when variants in the QTL have not been tested in the GWAS, imputing missing effects in GWAS summary statistics, aligning fine mapping windows between GWAS and QTL, and identifying and correcting for discrepancy in summary statistics and LD reference panels. We applied *pecotmr* to pseudobulked scRNA-seq eQTL datasets for 6 brain cell types integrated with a GWAS for Alzheimer's disease (AD). Using 4 individual-level TWAS methods, two of which are novel to *pecotmr,* we identified 97 gene-context pairs associated with AD risk. Notably, 26 of these pairs were identified only with our novel TWAS methods using weights estimated by SuSiE and mr.ash and not with the widely used elastic net and LASSO methods. Among these novel gene-context pairs were six genes associated with microglial expression and AD risk (*AP4E1*, *CASS4*, *CLASRP*, *NCK2*,*SORL1*, *ZYX*), several of which have not previously been identified by other TWAS analyses or have not been specifically implicated in microglia. The *pecotmr* R package makes the application of GWAS integration with QTL significantly easier and more accessible to the broader genomics community, while also providing several novel innovations to improve power to detect associations between molecular traits and complex diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4111T: PeReS:personalized reference sets for polygenic risk scores negates need for arbitrary groupings

Authors:

A. Sartori[1], J. Sharma[1], E. Herrera-Luis[1], Y. Ding[2], S. Buyske[3], C. Gignoux[4], C. Haiman[5], E. Kenny[6], C. Kooperberg[7], K. North[8], B. Pasaniuc[2], G. Wojcik[1]; [1]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]UCLA, Los Angeles, CA, [3]Rutgers Univ., Piscataway, NJ, [4]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Fred Hutchinson Cancer Ctr., Seattle, WA, [8]Univ. of North Carolina, Chapel Hill, NC

Abstract:

Polygenic scores (PGS) have helped translate GWAS results into clinical practice but continue to have limited utility across individuals of diverse genetic backgrounds, especially recently admixed individuals. PGS performance estimation often uses a crude population reference set that relies on subjective groupings, typically by race/ethnicity, which may obfuscate significant substructure within populations. Here we introduce a novel approach using a personalized reference set (PeReS) for PGS, which addresses these challenges by tailoring PGS to a single individual's genetic background. Focusing on a cohort of 43,674 individuals with diverse self-identified race/ethnicity (Hispanic/Latino=19,379, African American=16,242, Asian=4,194, Native Hawaiian=2,258, Native American=629, Other=972) from the PAGE Study, we employed PeReS for lipid and anthropometric traits. A k-nearest neighbors (KNN) model (k=1,000) was trained on principal components (PC)-based genetic distance. For each index individual, we defined new reference sets reflecting individual-specific genetic similarities, resulting in raw PGS scores re-scaled within each set. Comparative analysis revealed similar PGS accuracy between our approach and normalization by race/ethnicity (incremental R^2=0.069 to R^2=0.070 for PGS_{LDL}, R^2=0.091 to R^2=0.090 for PGS_{HDL}, R^2=0.063 to R^2=0.060 for PGS_{Height}, respectively). Next, we evaluated changes in individual's risk before and after PGS scaling with PeReS. Individuals were assigned to 3 risk categories based on quintiles (1st=high; 2nd,3rd,4th=normal; 5th=low). We observed significant changes across risk categories for PGS_{LDL} (P(ANOVA)<0.001). We then examined individuals at the edges of trait and PGS distributions, focusing solely on LDL, signifying notable deviation, and identified 134 outliers for PGS_{LDL} and 121 outliers for PGS_{LDL_scaled} (z-score>3). Our method yielded 9.70% fewer outliers compared to traditional approaches, thus allowing the inclusion of more individuals and reducing the risk of exclusion based on subjective categorizations. Finally, we applied PeReS to PGS_{LDL} in 87,175 diverse

individuals from the All of Us Research Program (AoURP), and while the results were consistent across both methods, akin to PAGE, the accuracy was overall lower (incremental R^2=0.058 to R^2=0.057). By avoiding discrete categorizations, PeReS can achieve the predictive accuracy of conventional methods and reflects the genetic variability present in diverse populations, while eliminating the need of artificial boundaries in precision medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4112T: PhenoSS: Phenotype semantic similarity-based approach for rare disease prediction and patient clustering

Authors:

S. Chen[1], Y. Hu[2], C. Liu[3], C. Weng[3], K. Wang[4]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children s Hosp. of Philadelphia, Philadelphia, PA, [3]Columbia Univ., New York, NY, [4]CHOP, Philadelphia, PA

Abstract:

Human Phenotype Ontology (HPO)-based characterization of individual patients with rare diseases has been widely adopted in recent years. Such analysis typically starts from a list of HPO terms manually curated or computationally extracted via Natural Language Processing (NLP) approaches, followed by statistical methods to prioritize a list of possible disease-causing genes or likely disease diagnoses for the patient. However, several challenges remain for efficiently utilizing phenotype information in such analyses. First, existing methods do not consider the hierarchical structure of HPO terminology. Second, existing methods do not take into account the correlated structure between multiple similar HPO terms output by NLP methods; third, batch effects that influence the precision and completeness of HPO terms for patients are commonly observed but are not typically corrected. To address these limitations, here we present PhenoSS, an effective algorithm that makes disease prediction and performs patient clustering based on HPO concepts. PhenoSS uses the Gaussian copula technique by modeling the marginal prevalence of each HPO term for each disease and utilizes a multivariate normal distribution to link them together to account for term correlations. We utilized the OARD (open annotations for rare diseases) API for inferring the frequency of HPO terms in a diverse range of rare diseases. PhenoSS can calculate the phenotype similarity between any two patients for finding similar patients or clustering purposes, or between one patient and any candidate diseases for diagnosis support. Another desired feature is its ability to account for potential batch effects. Extensive simulation studies and real-world data analysis demonstrated the effectiveness of PhenoSS in the face of varying scenarios of data quality. PhenoSS can be extended beyond HPO to other types of phenotype descriptors such as SNOMED-CT, CPT, or

icd9/icd10 codes. As the scale of EHR studies continues to grow, PhenoSS represents a much-needed and valuable tool for biomedical researchers to disentangle the complex clinical heterogeneity in rare disease studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4113T: Powerful, robust and fast whole-genome regression for Time-to-Event Data in large-scale biobanks

Authors:

Y. Zou, J. Marchini, J. Mbatchou; Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

Large-scale biobanks often have extensive genetic data and a broad range of phenotypes including time-to-event data on study subjects. The Cox proportional hazard model, a commonly used model for time-to-event data, has been shown to enhance the detection of genetic signals compared to case-control analysis, as it takes account of the time until events occur. However, the existing methods can be limited by their computational cost, robustness to rare variants and rare events (eg. wald test), population structure/relatedness (eg. SPACox[1]), or the lack of accurate effect size estimates (eg. score test). To address these limitations, we have expanded the capabilities of REGENIE[2] by incorporating the Cox model to analyze time-to-event data, REGENIE-COX. We incorporated Firth correction[3] to achieve a more robust test when testing rare variants association with rare events. To reduce the computational burden incurred by using Firth correction in the Cox model, we employed a fast approximation test and leveraged multi-threading. Through numerical simulations, we highlighted that REGENIE-COX can have enhanced power over REGENIE-Logistic, ***especially for high event rates phenotypes, where we observed up to 5 fold increase in empirical power***. Furthermore, the Firth correction provides more accurate effect size estimates and standard errors than the score test and test using SPA correction. Through the application of REGENIE-COX to large-scale time-to-event phenotypes in UK Biobank, we highlighted several examples where REGENIE-COX provided stronger signals compared to a standard case-control analysis, demonstrating its potential for more powerful, robust and efficient analysis of time-to-event data.
[1] Bi, W. et al. (2020). A Fast and Accurate Method for Genome-Wide Time-to-Event Data Analysis and Its Application to UK Biobank. American Journal of Human Genetics. https://doi.org/10.1016/j.ajhg.2020.06.003[2] Mbatchou, J. et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nature Genetics, 53, 1097-1103. https://doi.org/10.1038/s41588-021-00870-7[3] Heinze, G., & Schemper, M. (2001). A

Solution to the Problem of Monotone Likelihood in Cox Regression. Biometrics, 57(1), 114-119. https://doi.org/10.1111/j.0006-341X.2001.00114.x

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4114T: PRS-GRID: A Cross and Within Ancestry Polygenic Risk Prediction Method Based on Individual Genetic Distance

Authors:

X. Chen[1], L. Tang[2], V. Napolioni[3], J. Huang[2]; [1]The George Washington Univ., Washington, DC, [2]Southern Univ. of Sci. and Technology, Shenzhen, China, [3]Stanford Univ. Sch. of Med., Palo Alto, CA

Abstract:

Background Two decades of genome-wide association studies (GWAS) have led to the fast-growing application of polygenic risk prediction (PRS). However, due to population structure and evolutionary path differences, the PRS substrate derived mostly from studies of European ancestry does not work equally well for other ancestries. There is an association between prediction accuracy decay and individual genetic distance (GD) to the genetic centers (GC) of various populations. **Methods** We hypothesize that adding a GD-based weight into PRS methods would enhance its risk prediction performance, particularly for populations of multiple ancestry. We explore the GD first by principal components (PC) and then by phylogenetic tree structures. Building on top of an emerging software (PRS-CSx) that achieves high prediction accuracy across multiple-ancestries, we present PGS-GRID, where "GRID" stands for "**G**enetic **R**eference based on **I**ndividual **D**istance". **Results** We developed a preliminary version of PRS-GRID and pilot tested its prediction performance for a classic quantitative trait (*e.g.*, height) and a disease trait (*e.g.*, type-2 diabetes, T2DM). We found slight but noticeable improvement of risk prediction in minority populations. For the four populations (European, African, South Asian, Eastern Asian), the AUC for T2DM increased from 0.787, 0.681, 0.702, 0.813 to 0.790, 0.700, 0.716, 0.819, respectively; The *R2* for height increased from 0.748, 0.598, 0.615, 0.632 to 0.755, 0.612, 0.636, 0.691, respectively. **Conclusions** Our work provides a foundation for future research and clinical applications aimed at reducing health disparities and improving population health through personalized medicine. ***Acknowledgement*** This research was conducted using the UK Biobank resources under application 66137. We thank the participants of UK Biobank for making this work possible.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4115T: *Quantile IV*: A nonparametric Mendelian randomization method to estimate heterogeneous and nonlinear causal effects

Authors:

M-A. Legault[1,2], J. Hartford[3], B. J. Arsenault[4], A. Y. Yang[1], J. Pineau[1,2]; [1]McGill Univ., Montreal, QC, Canada, [2]Mila, Montreal, QC, Canada, [3]Valence Labs, Montreal, QC, Canada, [4]Université Laval, Quebec, QC, Canada

Abstract:

Mendelian Randomization (MR) enables the estimation of causal effects while controlling for unmeasured confounders. However, traditional MR relies on strong parametric assumptions and can be biased if these assumptions are violated. We recently introduced Quantile Instrumental Variable (IV), a new machine learning MR estimator. Quantile IV is a two-stage nonparametric estimator that uses neural networks to estimate the IV—exposure and the exposure—outcome relationships.

Our estimator is distinctive in its ability to estimate nonlinear and heterogeneous causal effects and offers a flexible approach for subgroup analysis. For example, Quantile IV can estimate conditional treatment effects without pre-specifying an interaction model. Confidence intervals for treatment effects are estimated using bootstrap aggregation or frequentist inference for continuous outcomes.

We show that Quantile IV performs competitively in realistic MR simulations and apply it in the drug target MR and conventional MR contexts. First, we investigated the impact of circulating sclerostin levels on heel bone mineral density, osteoporosis, and cardiovascular outcomes in the UK Biobank. Sclerostin is the drug target of romosozumab, an anti-sclerostin monoclonal antibody developed for fracture prevention in osteoporosis. The cardiovascular safety of sclerostin inhibition is debated. Using MR and colocalization techniques, our analysis reveals that a genetically predicted reduction in sclerostin levels significantly increases heel bone mineral density and reduces the risk of osteoporosis, with no increase in the risk of ischemic cardiovascular diseases.

We then estimated the causal effect of changes in waist-to-hip ratio (WHR) and low-density lipoprotein-cholesterol (LDL-c) on the same outcomes. Since LDL-c and WHR are established cardiovascular risk factors, they were used as positive controls. We observed deviations from linearity for the effect of both LDL-c and WHR on cardiovascular outcomes especially at higher values of the exposures where the effect was increased. Furthermore, the causal effect of a genetically predicted increase in LDL-c appeared to be modified by statin use, as constant genetically predicted increase in LDL-c led to a smaller increase in cardiovascular outcome risk

in statin users than non-users, possibly due to the drug offsetting the effect of the LDL-c increasing genetic variants.
Quantile IV contributes to the advancement of MR methodology, and the case studies demonstrate its applicability to estimate conditional average treatment effects in drug target and traditional MR settings.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4116T: Rare and common genetic associations with internalising and externalising behaviours across childhood and adolescence

Authors:

E. Wade, D. Malawsky, O. Wootton, M. Koko, Q. Huang, W. Huang, M. Hurles, H. Martin; Human Genetics, Wellcome Sanger Inst., Hinxton, United Kingdom

Abstract:

Genetic factors are known to influence vulnerability to common mental conditions; however, their contribution to related early life behaviours is unclear. Here, we investigated the rare and common genetic contribution to childhood and adolescent internalising and externalising symptoms in two UK birth cohorts, the Millenium Cohort Study and Avon Longitudinal Study of Parents and Children, with SNP array and whole exome sequencing data (~5000 trios). We performed mixed effects modelling with measures of common and rare genetic variation to evaluate associations with emotional and behavioural symptoms repeatedly measured by the Strengths and Difficulties Questionnaire. Across development, we found positive associations between polygenic indices (PGIs) for attention deficit hyperactivity disorder with externalising symptoms (p<10-9), major depressive disorder with externalising (p<10-5) and internalising symptoms (p<10-3), and neuroticism with internalising symptoms (p<10-5). Conversely, higher PGIs for educational attainment were associated with fewer externalising symptoms (p<10-5). Having a higher exome-wide burden of deleterious rare variants was also associated with more externalising (p<10-6) and internalising symptoms (p<10-3). Controlling for parental genetic measures, we saw little evidence for an association between non-transmitted alleles in the parents and children's behaviours. However, associations between the children's genetic scores and the phenotypes were attenuated after controlling for parental scores, leaving only a positive association between externalising symptoms with the ADHD PGI (p<10-6) and a negative association with the EA PGI (p<10-3). Effects of transmitted rare variant burden were inconsistent between MCS (positive association with externalising, p<10-4) and ALSPAC (positive association with internalising, p<10-4). Additionally, we tested for an interaction between the

genetic score and the age at which the phenotype was measured but found none, suggesting the influence of genetics was consistent across development. Our results elucidate genetic risk factors for externalising and internalising behavioural symptoms and may inform future research into early identification and prevention of mental disorders.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4117T: Rare variant analysis pipelines: a systematic review

Authors:

A. Ziegler, **C. Riccio**, M. L. Jansen, H. Solleder; MedizinCampus Davos, Davos, Switzerland

Abstract:

The sequencing of increasingly larger cohorts has unveiled a large number of rare genetic variants, presenting an opportunity for unravelling the genetic basis of complex traits. Rare variants may have strong phenotypic effects, but their low frequency necessitates specialized statistical analyses. Many software packages have been proposed for analyzing rare variants. However, to be used in practice, rare variant analyses need to be automated in user-friendly pipelines that support various trait types and statistical tests. In this work, we provide an overview of the functionality of currently available rare variant analysis pipelines. We searched MEDLINE and Google Scholar up to November 27, 2023, and included open-source rare variant pipelines that accept genotype data from cohort and case-control studies and group variants into testing units. Eligible pipelines were assessed based on a pre-defined list of 45 functionality and usability criteria and introduced a new score for evaluating the quality of the pipeline documentation. We identified 15 rare variant pipelines that collectively support many types of traits, association tests, testing units, and variant weighting schemes. However, no pipeline meets all analytical needs and desirable usability criteria. No pipeline supports variants occurring in short tandem repeats or mitochondria. The documentation scoring revealed areas for improvement in some pipelines. This review revealed that several automated scalable pipelines are available for analyzing rare variants in large whole genome sequencing studies. Most pipelines provide p-values from statistical association tests only, and they generally fall short of providing effect estimates and biological interpretation. The functionality and usability of rare variant pipelines should thus be expanded to fully exploit the potential of rare variant analyses in uncovering the genetic basis of complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4118T: Rare variant intensity estimation for genetic mapping of complex traits

Authors:

J-R. Jhuang[1], W-Y. Wei[1,2], Y-C. Lin[1], C-Y. Guo[2], H-C. Yang[1,2]; [1]Inst. of Statistical Sci., Academia Sinica, Taipei, Taiwan, [2]Inst. of Publ. Hlth., Natl. Yang-Ming Chiao-Tung Univ., Taipei, Taiwan

Abstract:

Rare causal variants often underlie complex disorders within the general population. With the advent of next-generation sequencing technologies, there's been a significant stride in assessing the impact of these rare variants on complex traits at a large scale. However, analyzing such copious sequencing data poses both opportunities and challenges, primarily revolving around the limited statistical power of rare-variant association analyses. This study introduces a novel approach for genetic mapping in complex traits through rare variant intensity estimation. The methodology involves two key steps. Firstly, employing the nearest neighbor method to generate sliding windows along each chromosome. Secondly, utilizing local polynomial regression within each window to estimate the rare variant intensity. This regression incorporates local weights, reflecting the proximity of genetic markers to an anchor point, and locus weights, accounting for the pronounced effect of a rare variant on the complex trait. Subsequently, the generalized linear model is employed to ascertain associations between the rare variant intensity estimates and the complex trait. Through simulation studies, the performance of the proposed method is compared against conventional approaches such as the burden test, the variance component tests, and the combination of burden and variance component tests. Results indicate that the proposed method exhibits superior statistical power and better type I error control. Additionally, analysis of a dataset from Genetic Analysis Workshop 19 using the proposed method identifies several genes associated with hypertension and systolic and diastolic blood pressure.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4119T: Relationships Between Physical Activity And Overall Brain And Body Health Demonstrated By Multi-Organ Imaging Genetics

Authors:

Y. Guo[1], E. Yang[2], Z. Fan[2], J. Shu[1], P. Paschou[1], B. Zhao[2]; [1]Purdue Univ., West Lafayette, IN, [2]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

While physical activity is vital for physical and mental well-being, our understanding of its interplay with multiple organs and associated dysfunctions from a whole-body perspective remains limited. Based on evidence suggesting a moderate heritability of physical activity, we investigated genetic connections between 79 self-reported and device-measured physical activity traits and the structure and function of multiple organs derived from imaging data of over 40,000 individuals in the UK Biobank. We found widespread genetic links to imaging phenotypes throughout the body, including total brain volume, average cortical thickness, brain functional connectivity, white matter microstructure, heart chamber volumes, retinal layer thickness, abdominal organ volumes and fat composition. Furthermore, we identified genetic correlations between physical activity and various diseases involving these organs and explored the three-way connections among physical activity, imaging phenotypes, and relevant diseases. In summary, our findings uncovered structural and functional imaging indices for physical activity and the underlying genetic mechanisms of potential beneficial effects of physical activity on brain and body health.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4120T: Removing array-specific batch effects in GWAS mega-analyses by applying a two-step imputation workflow

Authors:

M. Nasr[1], E. König[2], C. Fuchsberger[3], S. Ghasemi[4], U. Völker[5], H. Völzke[1], H. Grabe[6], **A. Teumer**[7]; [1]Univ. of Med. Greifswald / DZHK, Greifswald, Germany, [2]Inst. for Biomedicine, Eurac Res., Bolzano, Italy, [3]Eurac Res., Bolzano, Italy, [4]Inst. of Genetic Epidemiology, Med. Ctr. – Univ. of Freiburg, Freiburg, Germany, [5]Interfaculty Inst. for Genetics and Functional Genomics, Greifswald, Germany, [6]Univ. of Med. Greifswald / DZNE, Greifswald, Germany, [7]Univ. Med. Greifswald, Greifswald, Germany

Abstract:

Background: Combining individual-level data in genetic association studies (mega-analyses) increases statistical power for identifying genetic effects underlying diseases. However, batch effects arising from combining imputed genotypes from different genotyping arrays pose a significant challenge in mega-analysis. Here, we developed a two-step imputation workflow to overcome the array-type bias. We then conducted as a proof of concept a genome-wide association study (GWAS) to test for genetic associations with goiter risk and thyroid gland volume to compare results from conventional imputation and our two-step method.

Methods: We included genotype data from 10,647 individuals from five different array types (Illumina Omni 2.5, Illumina GSA, Illumina PsychArray, Affymetrix Axiom, and Affymetrix SNP 6.0). In the two-step imputation, array-specific reference panels were designed for the first imputation step and subsequently imputed against the 1000GenomesPhase3 reference panel. We compared three approaches: conventional array-specific imputation (reference), one-step imputation using the intersection of variants across arrays, and our novel two-step imputation. Batch effects were assessed using the first twenty genetic principal components of the combined sample after the final imputation. The performance of the different approaches was evaluated by comparing the deviation of allele frequencies.
Results: The one-step approach included 58,091 genotyped variants for imputation. In the two-step design, the intermediate imputation resulted in 1,866,951 autosomal variants used as input for the 1000Genomes imputation. The proposed two-step imputation eliminated the batch effect from the first twenty principal component of the genetic data. The two-step imputation showed strong correlation with the conventional approach in the allele frequencies (squared correlation > 0.99). GWAS using data from the two-step method confirmed the four known associations on these traits and revealed four novel loci for thyroid volume (*TG, PAX8, IGFBP5, NRG1*), and a novel one for goiter (*XKR6*), which was not statistically significant following the meta-analysis of the summary statistics with conventional imputation. All associations were also associated with thyrotropin supporting true positive associations results by biological plausibility.
Conclusion: We developed and tested a two-step imputation workflow that provides high quality imputation results without technical batch effects, fostering mega-analysis of genotype data obtained from multiple array types for GWAS and other potential applications like risk score analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4121T: Removing confounding and refining mendelian randomization genetic instruments of dietary intake traits

Authors:

K. Sutton, J. Cole; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Background: Genome-wide association studies of dietary intake have opened the door for testing Mendelian randomization (MR) on myriad dietary traits and health outcomes. However, it is unclear whether the genetic instruments associated with diet act via direct biological

mechanisms impacting intake or via indirect mechanisms due to heritable or non-heritable confounding. The ease of MR analyses coupled with non-specific genetic instruments have led to hundreds of false causal claims between diet and markers of human health.
Method: We performed a battery of MR approaches to better understand the behavior of dietary intake genetic instruments. To address instrument misspecificity due to correlated and uncorrelated horizontal pleiotropy and incorrect directionality, we performed both traditional and robust MR approaches (inverse variance weighted (IVW), weighted median (WM), MR-Egger intercept, Steiger) and newer more comprehensive methods including MR-CAUSE (MRC) and PheWAS-Cluster (PWC). We focused on dietary exposure-outcome relationships with causal evidence supported by randomized controlled trials (RCT). These "positive control" pairs included (1) adding salt to food, (2) fresh fruit intake, and (3) raw vegetable intake on blood pressure, (4) a diet pattern driven by whole grain consumption on apolipoprotein B, and (5) overall oily fish intake on triglycerides. Heterogeneity was evaluated by Cochran's Q and instrument strength was evaluated by the F-statistic.
Results: On non-filtered instruments, the Egger intercept test only indicated horizontal pleiotropy for oily fish on triglycerides (P-value = 0.04). However, for all diet traits the PheWAS-Cluster method identified and removed variants from the genetic instrument that had stronger associations with 400 phenome-wide traits in UK Biobank suggesting the presence of indirect pathways missed by Egger. Despite our focus on established causal relationships from clinical trials, most relationships were not consistently significant across MR methods although the effect direction often matched the expectation (4/5 of the directions of dietary trait-health outcome pairs matched their RCT). Importantly, filtering genetic instruments tended to improve heterogeneity (e.g. fresh fruit on systolic blood pressure: unfiltered instrument, Q=266; PC filtered, Q=152), but did not make a substantial impact on the results of the MR tests.
Conclusions: Although MR is an attractive tool for inferring causal relationships between dietary intake and health outcomes, inconsistent MR findings using positive controls warrants caution when performing and interpreting dietary MR.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4122T: Retrospective analysis frameworks to control for population admixture and sample relatedness in large-scale genome-wide association studies

Authors:

W. Bi, H. Xu, Y. Ma, P. Zhang; Peking Univ., Beijing, China

Abstract:

Population admixture and sample relatedness are two major confounders in large-scale GWAS and can lead to inflated results if not appropriately controlled. To address these issues, numerous approaches have been proposed, primarily for binary and quantitative traits. Recently, there is a growing interest in studying complex traits with intricate structures, such as time-to-event traits and longitudinal traits. However, applying existing strategies to control for population admixture and sample relatedness in these complex traits is technically challenging to. Therefore, there is an urgent need for an analysis framework that is applicable to a wide variety of traits, scalable to analyze biobanks containing millions of subjects, and statistically powerful.

We propose two scalable and accurate general-purpose analysis frameworks, SPA_{MIX} and SPA_{GRM}. These frameworks employ retrospective strategies that consider genotypes as random variables, making them applicable to statistical models charactering time-to-event and longitudinal traits. By incorporating saddlepoint approximation, the framework outperforms regular retrospective methods that rely solely on normal distribution approximation, especially when testing for low-frequency and rare variants.

SPA_{MIX} addresses population admixture by estimating individual-level allele frequency for each variant using raw genotype data and SNP-derived principal components (PCs). The strategy offers higher robustness compared to prospective methods. For example, when analyzing quantitative traits, regular methods based on linear regression become invalid in the case of variance heterogeneity, whereas SPA_{MIX} is still accurate. The SPA_{MIX} framework can also incorporate local ancestry to test for ancestry-specific genetic effects.

SPA_{GRM} adjusts for sample relatedness by utilizing a sparse genetic relationship matrix and identity-by-descent information to account for the implicit genetic correlation in score statistics. We evaluated SPA_{GRM} in longitudinal data analysis. Linear mixed effect models and generalized estimation equations, the two fundamental approaches, can both be used to fit a null model, and SPA_{GRM} can handle subsequent steps to conduct GWAS. Due to the distinct architectures of longitudinal traits, no single method can be uniformly most powerful. Thus, we propose $SPA_{GRM(CCT)}$ to aggregate results from different methods via Cauchy combination test. Both simulation studies and real data analysis in the UK Biobank demonstrated that $SPA_{GRM(CCT)}$ can serve as an optimal unified approach for longitudinal trait analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4123T: Retrospective set-based gene-environment interaction tests for biobank data

Authors:

G. Xu[1], A. Amei[2], L. Shen[2], H. Zhou[3], Z. Wang[3]; [1]Yale Univ., New Haven, CT, [2]Univ. of Nevada, Las Vegas, Las Vegas, NV, [3]Yale Univ., New Haven, CT

Abstract:

As human complex diseases are influenced by the interplay between genetic factors and the environment, the identification of gene-environment interactions is important for understanding disease mechanisms and predicting disease risk. Biobank data, with its large sample size and rich genetic and non-genetic information, offers unparalleled opportunities for identification of gene-environment interactions. Recently, set-based gene-environment interaction tests that aggregate information of genetic variants in a region have been developed. However, many of these methods are computationally intensive and challenging to apply to biobank data. In this study, we propose retrospective set-based gene-environmental interaction tests (RSITs) to identify the interaction between a set of genetic variants and an environmental factor on a binary trait. To reduce the computational burden, we first fit a null generalized linear model without genetic effects. We then use matrix projection to account for genetic main effects in a retrospective model of genetic variants given the phenotype, covariates, and environment. To aggregate a set of genetic variants, we employ the burden test (RSIT-B), variance component test (RSIT-S), and an efficient hybrid test combining the burden and variance component test (RSIT-E). Simulation studies demonstrated that RSITs had well-controlled type I error. Additionally, RSIT-E and RSIT-S were more powerful than other existing methods compared, with RSIT-E being the most powerful overall. We applied RSITs to analyze gene-physical activity interactions on bone fracture in 140,047 women in the UK Biobank. Among the top nine genes with p-values $< 10^{-4}$, RSIT-E had the smallest p-values for eight genes. Notably, two genes, *CREG22* and *MACC1*, had been previously reported to be associated with bone health. Our results demonstrate that RSITs can detect biologically relevant genes interacting with physical activity to influence bone fracture risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4124T: Revealing ancestry-specific genetic effects on Alzheimer Disease in African Americans

Authors:

K. Hamilton-Nelson[1], F. Rajabli[1,2], K. Scott[1], G. Tosto[3], S. Williams[4], B. Kunkle[1,2], J. M. Vance[1,2,5], M. Cuccaro[1,2], C. Reitz[3], S. Blanton[1,2], K. McInerney[5], J. Haines[4], G. Byrd[6], Alzheimer's Disease Genetics Consortium, M. Pericak-Vance[1,2,5]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Dept. of Neurology, Columbia Univ. Coll. of Physicians and Surgeons, New York, NY, [4]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [5]Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC

Abstract:

Background: There is known heterogeneity in effect sizes and genetic loci associated with Alzheimer disease (AD) across ancestral populations. African Americans (AA), with their admixed African and European genetic backgrounds, provide an opportunity to investigate these differences while controlling for non-genetic factors. By leveraging local ancestry information, this study examines 86 known AD loci in African (AF) and European (EU) backgrounds within the AA population to elucidate ancestry specific effects. These findings are essential for developing more effective and personalized approaches to AD prevention and treatment across all populations. Methods: This study used TopMed imputed 5866 self-reported AA individuals from the AD Genetics Consortium (1881 cases; 3985 cognitively unimpaired). We estimated principal components using EIGENSTRAT. Local ancestry (LA) was inferred using reference samples from the 1KG and HGDP panels and estimated using Eagle2.4 and FLARE. We utilized TRACTOR to investigate ancestry specific effects, adjusting for age, sex, principal components (PC1-3), and the number of African ancestry copies at each locus. Results were meta-analyzed using the RE2 approach in METASOFT. We followed up on 86 lead markers reported in Bellinguez et al and Kunkle et al genome wide association studies for EU and AF ancestries. Results: We observed comparable effect sizes for risk markers on the AF and EU backgrounds at the *BIN1* locus (AF: OR=1.1, p=0.01; EU: OR=1.1, p=0.2), the *EPHA1* locus (AF: OR=0.9, p=0.03; EU: OR=0.9, p=0.2), and for the protective *APOE-e2* allele (AF: OR=0.6, $p=5.7x10^{-12}$; EU: OR=0.6, $p=9.3x10^{-3}$). The risk associated with the *APOE-e4* allele (AF: OR=2.4, $p=1.7x10^{-54}$; EU: OR=2.8, $p=1.9x10^{-16}$) and *UMAD1* (AF: OR=1, p=0.73; EU: OR=1.4, p=1.3x10-*3*) were greater on the EU background. Additionally, we found ancestry specific signals that are significant on one background, but monomorphic on another: *ABCA7* marker (Kunkle et al, AA GWAS) on the AF background (OR=1.4, $p=4.9x10^{-4}$), and the *CR1* marker on the EU background (OR=1.4, p=0.03). Conclusion: Our study identifies ancestry specific AD genetic effects in AA. We observed similar effects in the *BIN1* and *EPHA1* risk loci, the *APOE-e2* protective allele, and heterogeneity in *APOE-e4* and *UMAD1* risk, with the increased risk on the EU background. Additionally, we detected ancestry specific signals in *ABCA7* (AF) and *CR1* (EU). These findings

highlight the importance of developing and implementing polygenic risk scores that account for ancestry specific genetic variations to improve the accuracy of risk prediction and the effectiveness of targeted interventions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4125T: scFM: An efficient statistical fine-mapping approach for eQTLs using large-scale single-cell data

Authors:

X. Rui, Z. Lu, Z. Zhang, S. Gazal, N. Mancuso; Univ. of Southern California, Los Angeles, CA

Abstract:

Identifying genetic variants that impact gene expression (eQTL) at the cellular level is crucial in understanding disease biology and developing therapeutic targets. Population-scale single-cell transcriptomic technologies (scRNA-seq), when combined with genetic data, allow the uncovering of candidate variants associated with gene expression in specific cell types. However, it is challenging to pinpoint causal variants for eQTLs due to linkage disequilibrium between adjacent variants. Current statistical fine-mapping approaches applied to eQTLs were mostly designed for bulk RNA seq while ignoring the plausible shared eQTL effects between cell types observed in scRNA-seq data.To address this, we present Single Cell Fine Mapping (scFM), a multi-cell-type fine-mapping framework, to identify putative causal variants for eQTLs using individual-level scRNA-seq data. scFM can capture eQTL effects that are shared or specific between cell types while considering the presence of multiple independent causal variants. Importantly, scFM explicitly models residual correlations across cell types to account for the similarity of gene expressions due to non-genetic factors within individuals. To perform inference, scFM leverages an ultra-efficient variational inference algorithm that is scalable to large-scale scRNA-seq data and runs seamlessly on CPUs, GPUs, or TPUs.We performed extensive simulations to evaluate scFM and compared it with CAFEH, a fine-mapping approach for eQTLs using bulk RNA-seq from multiple tissues. Overall, scFM consistently obtained higher mean posterior inclusion probability (PIP) at true causal eQTLs compared with CAFEH across varying sample sizes ($p = 8.46 \times 10\text{-}6$), number of simulated cell types ($p = 9.45 \times 10\text{-}4$), and eQTL heritability ($p = 2.26 \times 10\text{-}2$). Similarly, we discovered more credible sets containing causal eQTLs in the same simulated scenarios. Finally, we observed that scFM provides robust estimations of residual correlations ($R2 = 0.99$) and remains unbiased when the number of causal eQTLs is misspecified. To benchmark scFM and gain biological insights in real data, we will apply scFM on the OneK1K scRNA-seq dataset (N = 982, Yazar et al. 2022) to perform fine-

mapping for causal eQTLs across 14 PBMC blood cell types.Taken together, scFM provides a scalable and robust approach to perform fine-mapping for eQTLs by appropriately accounting for correlation between cell types in single-cell data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4126T: Sex-Specific Genetic Overlap Between Multisite Chronic Pain and Asthma by Age of Onset

Authors:

G. Koli Farhood[1,2], M. Parisien[1,3], L. Diatchenko[4,2,3], **A. Grant**[1,2,3]; [1]McGill Univ., Montreal, QC, Canada, [2]Alan Edwards Ctr. for Res. on Pain, McGill Univ., Montreal, QC, Canada, [3]Faculty of Dental Med. and Oral Hlth.Sci., McGill Univ., Montreal, QC, Canada, [4]Mc Gill Univ, Montreal, QC, Canada

Abstract:

Chronic pain (persistence of pain experience > 3 months) and asthma are common diseases, with a prevalence of at least 10% worldwide. Our previous Mendelian Randomization and biological function study showed a causal effect of Multisite Chronic Pain (MCP) on older adult-onset asthma. While the prevalence of chronic pain and asthma increases disproportionally among women with advancing age, the common genetic architecture underlying these traits has not been systematically explored by sex. We investigated the extent to which the shared genetics underlying MCP and asthma, considering age-of-onset (AO), might be differentially correlated by sex. We applied multilevel regression analysis on random samples of healthy controls (80%) and cases (90%) to identify phenotypic associations between MCP (number of chronic pain sites) and asthma-AO phenotypes (< 18 years (childhood-onset), 18-40 years (younger adult-onset), and > 40 years (older adult-onset)). Sex-stratified trait-specific genome-wide association scans were conducted across 9.2 million variants on European-descent female (N=248,103) and male (N=208,853) subjects at baseline visit from the UK Biobank. We estimated single trait narrow-sense heritability and cross-trait genetic correlation to quantify differential shared genetic variances explained by sex using Linkage Disequilibrium Score (LDSC) regression analysis. The multivariate regression analysis revealed statistically significant associations between asthma-AO and MCP by sex. The magnitude of MCP risk was highest among females with older adult-onset asthma (OR = 1.36; 95% CI: 1.31-1.42), whereas being female acted as a protective factor against childhood-onset asthma (OR = 0.59; 95% CI: 0.56-0.61). The heritability of childhood-onset asthma was higher in males (4.1%) compared to females (3.4%). Conversely, females had a higher estimated heritability than males in younger

adult-onset (2.2% vs. 1.6%), older adult-onset asthma (2.2% vs. 1.1%), and MCP (7.8% vs. 6.2%). The cross-trait genetic correlations between asthma-AO and MCP consistently increased from childhood-onset to older adult-onset asthma, with no correlation for childhood-onset asthma in either sex, and low to moderate correlations in younger and older adult-onset asthma in males (0.09 ± 0.07 and 0.24 ± 0.07, respectively). In contrast, females exhibited the highest significant genetic correlation in older adult-onset asthma (0.48 ± 0.04). A common polygenic component was shown to underlie older adult-onset asthma and chronic pain among females. Further investigation is planned to identify the shared biological mechanisms across these traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4127T: simGO: A Comprehensive Simulation Framework for Modeling Relationships Among Genotypes, Omics, and Phenotypes

Authors:

Q. Mei, C-T. Liu; Boston Univ. SPH, Boston, MA

Abstract:

Introduction Rapid advances in sequencing technologies and the increasing complexity of analysis have driven the development of novel statistical methods to tackle challenges in complex disease research using genetics and omics data. A comprehensive simulation framework is essential to evaluate the performance of these methods, yet no existing software fully accommodates the diverse range of required analysis. Consequently, researchers often create custom simulations, which are time-consuming, error-prone, and hinder reproducibility and comparability across different studies.**Methods and Results** We present *simGO*, a user-friendly tool designed to generate synthetic datasets for genotypes, omics, and phenotypes. *simGO*features an R wrapper that facilitates the genotype generation across different ancestries using the HAPGEN2, and allows the import of genotype data in Plink format or data generated from other software such as G2P and Sim1000G. It integrates various genetic architecture scenarios for omics and phenotypes generation. These scenarios include drawing the effect size of single nucleotide polymorphisms (SNPs) from normal, point-normal, or mixed normal distributions as used in the BSLMM and BayesR models. The tool can also model the relationship between SNP effect size and minor allele frequency to reflect the strength of negative selection. *simGO* can simulate gene expression data based on linkage disequilibrium (LD) block and gene annotations, and also accepts user-defined LD blocks and gene definitions for customized research needs. The effect size and number of causal genes can be specified or sampled from a user-specified distribution, which is useful for fine-mapping analysis.Under the

transcriptome-wide association studies (TWAS) framework, *simGO* supports multiple models to train weights between omics and SNPs, including least absolute shrinkage and selection operator (LASSO), ridge regression, elastic net, BSLMM, and BayesR. Moreover, *simGO* supports simulation across multiple traits, ancestries, and tissues with flexible parameters (e.g., genetic correlation, residual correlation, and proportions of genetic variance explained by SNPs and omics). Different scenarios across statistical methods have been validated with *simGO*, such as TWAS, and polygenic risk score. Furthermore, our developed Shiny web application simplifies the analysis process.**Conclusion** *simGO* offers a comprehensive framework for generating synthetic datasets, enhancing the reproducibility and comparability of genetic studies. Its wide range of features and user-friendly interface make it a valuable tool for genetics researchers.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4128T: SPLENDID improves polygenic risk prediction across the genetic ancestry continuum using diverse biobank data ⭐

Authors:

T. Chen[1], H. Zhang[2], R. Mazumder[3], X. Lin[1]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Natl. Cancer Inst., Cambridge, MA, [3]Massachusetts Inst. of Technology, Cambridge, MA

Abstract:

Background: Polygenic risk scores (PRS) are powerful clinical tools for risk stratification of complex traits and diseases, but their accuracy often varies across diverse populations. Current methods attempt to improve PRS accuracy in under-studied populations by using multi-ancestry GWAS summary statistics. Despite these efforts, these methods don't address the continuous nature of genetic ancestry, potentially limiting their clinical effectiveness in diverse patient populations. **Methods**: We introduce SPLENDID, an ensembled group-L0L1 penalized regression framework designed for biobank-scale data. SPLENDID enhances PRS accuracy across all ancestries without the need for discrete ancestry labels during model training. Our method captures interactions between genetic variants and ancestry principal components (PCs), allowing for the simultaneous modeling of shared and heterogenous effects to provide more individualized genetic risk assessment. We evaluated SPLENDID against several cutting-edge approaches using simulated genotype data with realistic linkage disequilibrium (LD) for 250,000 subjects of diverse ancestry. Additionally, we analyzed data from the All of Us Research Program (AoU) and UK Biobank (UKB) encompassing over 550,000 individuals for 9 continuous traits. **Results**: In large-scale simulation studies and analysis of data from AoU and UKB, SPLENDID demonstrates significant improvements in prediction accuracy over recently

proposed methods across diverse populations, including >100% over PRSCSX and 68% over iPGS in African ancestry. SPLENDID also selects an extremely sparse set of SNPs for enhanced risk interpretation. Notably, our model identifies variants in blood cell traits that show differential effects across ancestries in line with existing literature. Additionally, SPLENDID's framework allows for incorporation of external genomic data and other genetic interactions, further improving precision medicine efforts. **Conclusion**: SPLENDID offers a novel approach to developing polygenic risk scores by directly incorporating the continuum of genetic ancestry into genetic risk modeling. Our method significantly improves prediction accuracy across diverse populations, suggesting its potential to advance personalized medicine. By integrating external genomic data and clinical risk factors, SPLENDID provides a framework for more accurate and individualized genetic risk assessment, which can help improve health outcomes for patients from all backgrounds.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4129T: Spatial multi-modal atlas of bone tissue from human femur

Authors:

W. Lin, Y. Li, C. Qiu, B. Zou, Y. Gong, X. Zhang, D. Tian, W. Sherman, F. Sanchez, D. Wu, K-j. Su, Z. Luo, Q. Tian, Y. Chen, H. Shen, H-W. Deng; Tulane Univ., New Orleans, LA

Abstract:

The skeletal system is indispensable for structural integrity and physiological functions like mineral homeostasis and hematopoiesis in the human body. Maintaining equilibrium within the intricate bone microenvironment is critical, as imbalances can precipitate bone diseases. Elucidating the cellular constituencies, signaling cascades, and spatial regulatory mechanisms that orchestrate bone microenvironmental homeostasis is pivotal for comprehending the tissue-specific dynamics of gene expression and cellular behavior, paving the way for innovative therapeutic modalities against osteoporosis, osteoarthritis, and other bone disorders. In this pioneering study, we integrated spatial transcriptomics and single-cell transcriptomics to construct a multi-modal map of the human bone microenvironment within the femoral head. Our study unveiled two major niches within this microenvironment, with niche 2 enriched for osteoblastic lineage cells and fibroblasts exhibiting characteristics of maintaining bone homeostasis and facilitating regeneration, while niche 1 was mainly enriched for immune cells potentially involved in immunomodulatory processes, underscoring their specialized functional roles in bone tissue biology. Remarkably, the communication patterns and dependencies among

osteoblastic lineage cells and fibroblasts mirror the spatial distribution of the niche they comprise, highlighting the intricate interplay between cellular crosstalk and spatial organization. Moreover, we found an inverse correlation between the distance from the bone trabeculae and the abundance of osteoblastic lineage cells, the expression levels of key bone-related genes such as RUNX2, as well as the activities of pathways associated with bone formation and remodeling like PI3K, TGFβ, and VEGF. In contrast, the abundance of immune cells and the activities of inflammatory pathways such as NFκB and TNFα related to bone resorption generally increased with increasing distance from the bone trabeculae. In sum, our study contributes to a deeper understanding of the fundamental processes regulating bone homeostasis and regeneration and lays the foundation for future investigations aimed at developing innovative therapeutic strategies for bone-related disorders.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4130T: Splicing-specific transcriptome-wide association uncovers genetic mechanisms for Schizophrenia

Authors:

J. Hervoso[1], K. Amoah[1], J. Dodson[1], M. Choudhury[1], A. Bhattacharya[2], G. Quinones-Valdez[1], B. Pasaniuc[1], X. Xiao[3]; [1]UCLA, Los Angeles, CA, [2]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [3]Univ. of California Los Angeles, Los Angeles, CA

Abstract:

Recent studies have highlighted the essential role of RNA splicing, a key mechanism of alternative RNA processing, in establishing connections between genetic variations and disease. Genetic loci that influence RNA splicing variations show considerable influence on complex traits, possibly surpassing those affecting total gene expression. Dysregulated RNA splicing has emerged as a major potential contributor to neurological and psychiatric disorders, likely due to the exceptionally high prevalence of alternatively spliced genes in the human brain. Nevertheless, establishing direct associations between genetically altered splicing and complex traits has remained an enduring challenge. To address this, we introduce Spliced-Transcriptome-Wide Associations (SpliTWAS), a framework to integrate alternative splicing information with GWAS. SpliTWAS pinpoints genes and specific exons linked to traits through exon splicing events, providing a more granular view of genetic associations. We applied SpliTWAS to two schizophrenia (SCZ) RNA-seq datasets from dorsolateral prefrontal cortex (DLPFC) brain tissues, BrainGVEX and CommonMind (CMC), revealing 137 and 88 trait-associated exons (in 84 and 67 genes), respectively. This approach revealed a greater number of significant gene associations

compared to traditional TWAS, and more gene associations at GWAS risk loci for SCZ, demonstrating the enhanced sensitivity of SpliTWAS. Enrichment analysis of the associated gene sets highlighted biological functions converging on neuronal function and development, immune cell activation, cellular transport, and immune response, all of which are highly relevant to SCZ pathophysiology. Additionally, SpliTWAS variants were found to impact RNA-binding protein (RBP) binding sites, suggesting potential disruptions in RNA-protein interactions that affect splicing. We extended the probabilistic fine-mapping method FOCUS to the exon level, identifying 36 genes and 48 exons as putatively causal for SCZ. Notably, specific exons in genes such as *VPS45* and *APOPT1*, were associated with disease risk, eluding detection by conventional gene expression analysis. Our comprehensive analysis supports the substantial role of alternative splicing in shaping the genetic landscape of SCZ, providing a valuable approach for future investigations in this area.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4131T: Symptom-level characterization of Gulf War Illness in deployed and non-deployed participants enrolled in the Million Veteran Program

Authors:

J. He[1], G. A. Pathak[1], B. Cabrera-Mendoza[1], M. Aslan[2], D. A. Helmer[3], E. Hauser[4], R. Polimanti[1], Department of Veteran Affairs Cooperative Study Program #2006; [1]Yale Univ., New Haven, CT, [2]VA Clinical Epidemiology Res. Ctr. (CERC), VA Connecticut Hlthcare System, West Haven, CT, [3]Baylor Coll. of Med., Houston, TX, [4]Duke Univ Med Ctr, Durham, NC

Abstract:

Background: Gulf War Illness (GWI) has been diagnosed in a substantial portion of Veterans reporting multiple chronic symptoms following military service in the 1990-1991 Gulf War (GW). Unfortunately, after several decades, GWI pathogenesis remains unclear and diagnostic instruments are still very limited. Methods: Leveraging observational and genome-wide data available from the Million Veteran Program (MVP), we investigated GWI research definitions and performed a symptoms-level analysis. After performing a genome-wide association study (GWAS) of GWI research definitions based on Kansas and the Centers for Disease Control and Prevention (CDC) criteria, we investigated 14 GWI symptoms (i.e., fatigue, musculoskeletal, mood/cognition, and gastrointestinal symptoms) in 34,179 MVP participants who served between year 1990 and 1991. Results: Our GWI GWAS highlighted differences between Kansas vs. CDC research definitions, showing higher heritability and stronger polygenic overlap with

known GWI comorbidities. Considering the CDC criteria, the overall prevalence rates of 14 GWI symptoms ranged from 23.6% to 81.7%, with the highest for joint pain in both deployed and non-deployed Veterans. However, in those reporting at least one GWI symptom, sleep problems were the most severe symptom. In line with the CDC definition, cluster analysis distinguished gastrointestinal symptoms from the other symptoms. Being female, young age, non-European ancestry, low education attainment, GW-deployed status, and enlisted rank (vs. officer) were associated with increased GWI symptoms. In GW deployed Veterans, land deployment was associated with a higher GWI symptoms prevalence rate than sea deployment. With respect to GWI age of onset, Veterans with GWI onset before year 2000 showed stronger associations than Veterans with later disease onset. Conclusion: The present findings confirm the contribution of genetic variation to GWI and highlight the need for genetically informed symptom-level analyses to model the genetic architecture of this chronic condition, also exploring the contribution of gene-environment interactions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4132T: Synthetic surrogates improve power for genome-wide association studies of partially missing phenotypes in population biobanks

Authors:

J. Gao[1], Z. McCaw[2], X. Lin[2], J. Gronsbell[1]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Harvard T.H. Chan Sch Pub Hlth, Boston, MA

Abstract:

Large population biobanks (e.g., UK Biobank; UKBB) now contain data from hundreds of thousands of individuals, enabling numerous genetic discoveries. However, the statistical power of genome-wide association studies (GWAS) for certain traits remain limited due to missing phenotype information. For example, phenotypes that rely on advanced imaging techniques, such as dual-energy X-ray absorptiometry (DEXA) and magnetic resonance imaging (MRI), are often available only for a subset of the cohort. While machine learning (ML) models have the potential to predict missing values of target phenotypes by leveraging other data in the biobank, directly replacing missing values in target phenotypes with ML-imputed values in GWAS can lead to spurious associations. To address this issue, we propose a new method, synthetic surrogate (SynSurr) analysis, which jointly analyzes an incompletely observed target phenotype along with its predicted value from an ML model. SynSurr provides valid statistical inference and estimates the same effect size as a standard GWAS of the target phenotype, without requiring

the ML model to be correctly specified. Additionally, SynSurr improves power compared to standard GWAS as long as the imputed phenotypes are correlated with the target phenotypes. To assess the performance of SynSurr, we conducted an ablation analysis by randomly removing part of the observed data of two well-studied traits with minimal missingness from the UKBB: height and FEV1. SynSurr consistently recovered a higher proportion of associations present in the full sample than standard GWAS using observed target phenotypes ablated at different missingness levels. When applied to six incompletely measured body composition phenotypes from DEXA in the UKBB, SynSurr uncovered an average of 15.6 times more genome-wide significant associations. These findings were enriched for biologically relevant gene sets and substantially overlapped with known body composition associations from the GWAS catalog.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4133T: Systematic Comparison of Type I Error and Power in EMMAX and GENESIS

Authors:

G. E. Mitrick[1], J. M. Chernus[1], M. Lee[2], N. Mukhopadhyay[2], S. Curtis[3], J. C. Carlson[1,4], A. Butali[5], C. J. Buxó-Martinez[6], J. T. Hecht[7], L. M. Moreno Uribe[8], C. D. Padilla[9,10], F. A. Poletta[11,12], G. L. Wehby[13], E. Feingold[1,2,4], J. R. Shaffer[1,2], E. J. Leslie[3], J. C. Murray[14], S. M. Weinberg[1,2,15], M. L. Marazita[1,2,16]; [1]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral Biology, Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA, [3]Dept. of Human Genetics, Emory Univ. Sch. of Med., Emory Univ., Atlanta, GA, [4]Dept. of Biostatistics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [5]Dept. of Oral Pathology, Radiology and Med., Dows Inst. for Dental Res., Coll. of Dentistry, Univ. of Iowa, Iowa City, IA, [6]Dental and Craniofacial Genomics Core, Sch. of Dental Med., Univ. of Puerto Rico, San Juan, PR, [7]Dept. of Pediatrics, McGovern Med. Sch. and Sch. of Dentistry UT Hlth.at Houston, Houston, TX, [8]Dept. of Orthodontics, Coll. of Dentistry, Univ. of Iowa, Iowa City, IA, [9]Dept. of Pediatrics, Coll. of Med.; and Inst. of Human Genetics, NIH, Univ. of the Philippines Manila, Manila, Philippines, [10]Philippine Genome Ctr., Univ. of the Philippines System, Quezon City, Philippines, [11]ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP (Natl. Inst. of Population Med. Genetics), Porto Alegre, Brazil, [12]CEMIC: Ctr. for Med. Ed. and Clinical Res., Buenos Aires, Argentina, [13]Dept. of Hlth.Management and Policy, Coll. of Publ. Hlth., Univ. of Iowa, Iowa City, IA, [14]Dept. of Pediatrics, Carver Coll. of Med., Univ. of Iowa, Iowa City, IA, [15]Dept. of Anthropology, Univ. of Pittsburgh, Pittsburgh, PA, [16]Clinical and Translational Sci., Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Genome-wide association studies (GWAS) must balance type I and II errors to enable the identification of genetic variants associated with complex traits. Among the statistical methods used in GWAS, EMMAX (Efficient Mixed-Model Association eXpedited) and GENESIS (GENetic EStimation and Inference in Structured samples) are commonly used due to their ability to account for population structure and relatedness, therefore reducing false positives while retaining power. This study aims to systematically compare the performance and results of EMMAX and GENESIS.
We performed related association analyses of randomized phenotypes and genotypes in a dataset comprising subjects recruited for a study of orofacial clefting. Subjects include closely related individuals and considerable ancestral diversity. To compare type I errors, selected chromosomes were tested using EMMAX and GENESIS for association with a null phenotype. To compare type II errors, simulated associations were “spiked-in” by simulating genotypes conditional on phenotypes.
Preliminary results indicate that while both methods identify overlapping significant loci, there are notable differences. For a binary phenotype, EMMAX produced a higher number of associations. GENESIS, which uses a more sophisticated model to account for population substructure and relatedness, demonstrated more control over type I error rates.
The results suggest that while EMMAX is advantageous for its speed and ability to handle large datasets efficiently, GENESIS offers enhanced precision in controlling type I error rates, making it more reliable for identifying true genetic associations. The loci consistently identified by both methods indicate the robustness and reliability of these findings, highlighting critical genetic variants that are less likely to be false positives. This approach will help determine the sensitivity and specificity of each method in detecting true genetic signals, further informing their optimal use in genetic studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4134T: The dual influence of reference panel size on genotype imputation accuracy and marker selection

Authors:

A. Mittal[1], Y. Shi[2], N. D. Duro[2], **N. Khazanov**[2], P. Devineni[2], A. Fowler[2], S. Wu[1], S. Pyarajan[2], J. Gollub[1]; [1]Thermo Fisher Scientific, Santa Clara, CA, [2]on behalf of the Million Veteran Program, Boston, MA

Abstract:

Genotype imputation increases the power of GWAS studies, and the accuracy of imputation directly influences the results of subsequent analyses. Increasing the size of reference panels has the potential to improve both the accuracy and efficiency of genotype imputation. In our preliminary study using our previously published marker selection algorithm and a custom reference panel, we observed that increasing the reference panel size from 100 to 1000 samples led to improved imputation accuracy for both low-frequency and common variants. Additionally, the markers selected based on the larger panel proved to be more efficient than those based on smaller panels. When fixing the number of markers used for imputation, the larger panel provided superior imputation accuracy, indicating that a larger reference panel facilitates the selection of more informative markers.
Building on our preliminary findings that larger reference panels improve imputation accuracy and allow for more effective selection of genetic markers, we extend the analysis to a broader range of reference panel to profile the benefits of further expansion. By incorporating whole-genome sequencing data from the Million Veteran Program (MVP) cohort, we aim to systematically establish the relationship between the sizes of reference panel—ranging up to 50,000 samples—and their corresponding imputation performance across different MAF bins and ethnicities. Moreover, our expanded analysis seeks to evaluate how the size, diversity, and composition of the reference panel enable more efficient marker selection for population genetics microarrays. By conducting this comprehensive analysis, our research aims to provide insights into the return on increased panel size in terms of imputation accuracy and marker selection efficiency.
Disclaimer: The findings from this study are provided for research use only.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4135T: The effect of population differences on two-sample Mendelian randomization.

Authors:

J. Li, J. Morrison; Univ. of Michigan, Ann Arbor, MI

Abstract:

Mendelian randomization (MR) uses genetic variants as instrumental variables to infer causal links between exposures and outcomes. In two sample MR (2SMR), different study samples are used for each trait. To obtain valid inference from 2SMR, both studies must be performed in individuals from the same population. Using studies from different populations may bias the causal estimate of 2SMR due to differences in linkage disequilibrium (LD) or instrument-

exposure associations between populations.
2SMR studies usually match populations based on continental ancestry, but the possibility of mismatch remains due to differences in subpopulations. Moreover, matching population studies may not always be available for understudied populations. Thus, it is important to characterize the validity of 2SMR under different degrees of population mismatch.
We conduct a broad survey of replicability of 2SMR estimates using matching and mismatching populations, in addition to simulation experiments to isolate different sources of bias. We perform 2SMR across 675 trait pairs (27 exposures, 25 outcomes). For each trait pair, we identify 2-6 publicly available GWAS studies for both traits across a range of ancestries, and compute 2SMR estimates for all study pairs. To augment these results, we conduct 2SMR in data simulated with different LD but matching instrument-exposure associations, allowing us to isolate the effect of mismatching LD.
In our empirical survey, we find that population mismatch often biases the 2SMR estimate towards zero. We observed this bias even between populations with the same continental ancestry, such as between Finnish and other European populations. In our simulations, we found that mismatching LD alone also biases 2SMR estimation towards zero, but to a much smaller extent than the bias observed in our empirical studies. This suggests that the bias in 2SMR from population mismatch cannot be resolved solely by accounting for differences in LD, and may be largely driven by differences in genetic associations. Our work suggests that mild mismatch can bias the estimate of 2SMR; thus, researchers should acknowledge this possibility when mismatch is unavoidable.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4136T: The longitudinal effect of *APOL1* risk alleles on sickle cell disease-associated kidney damage

Authors:

S. Rashkin[1], G. Kang[2], C. Takemoto[1], M. Weiss[1], K. Ataga[3], J. Lebensburger[4], R. Zahr[5]; [1]Dept. of Hematology, St. Jude Children's Res. Hosp., Memphis, TN, [2]Dept. of Biostatistics, St. Jude Children’s Res. Hosp., Memphis, TN, [3]Ctr. for Sickle Cell Disease, Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [4]Div. of Pediatric Hematology and Oncology, Univ. of Alabama at Birmingham, Birmingham, AL, [5]Div. of Pediatric Nephrology and Hypertension, Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

Abstract:

Kidney damage in sickle cell disease (SCD) results in progressive chronic kidney disease (CKD), characterized by childhood hyperfiltration, albuminuria, and subsequent decline in kidney function. Prior to development of late CKD, a rapid decline in estimated glomerular filtration rate (eGFR) portends early CKD and early mortality. Several genetic modifiers of SCD-related kidney damage have been reported, the most notable being *APOL1* G1/G2 variants. While it is known that these *APOL1* variants are associated with poorer outcomes, few studies have examined the longitudinal effect, and none have explored the effect as children progress into adulthood. Previously, we examined the effect of *APOL1* G1/G2 on albuminuria in children with SCD, but the later risk for CKD is unknown. We address this via longitudinal analysis in the same cohort. Understanding the effect of *APOL1* on kidney disease progression in SCD is vital for defining disease trajectory, identifying at-risk individuals at all ages, and determining the optimal timing for prospective renoprotection, including the use of new *APOL1* channel drug inhibitors.

We examined 492 individuals enrolled in the Sickle Cell Clinical Research and Intervention Program (SCCRIP) longitudinal study (mean ± standard deviation age at end of follow-up: 15 years ± 9). We conducted time-to-event analyses at age <13 for albuminuria (albumin-to-creatinine ratio [ACR] > 30 mg/g; N=274; N_{cases}=36), kidney injury (CKiD U25 formula eGFR < 90 ml/min/1.73m^2; N=414; N_{cases}=111), and hyperfiltration (eGFR > 135 ml/min/1.73m^2; N=414; N_{cases}=267). We evaluated the effect of *APOL1* G1/G2 using Cox regression, adjusted for sex, sickle genotype (SS/Sβ^0 vs other), α-thalassemia, hydroxyurea and chronic transfusion use, and principal components 1-5. The effect of *APOL1* G1/G2 on the yearly median of ACR (square root log transformed; N=253; measurements [M]=787) and eGFR (log transformed; N=269; M=1176) at ages 13-25 was evaluated via linear mixed models with random intercept and slope, adjusted for the same covariates, age, and, if significant, the age-*APOL1* interaction.

APOL1 was associated with earlier incidence of albuminuria (hazard ratio [HR]=3.5, P=0.002) but not kidney injury (HR=0.5, P=0.07) or hyperfiltration (HR=0.8, P=0.3) in age <13. In age >13, *APOL1* was associated with an increase in ACR (β=0.3; P=3x10^{-10}) and eGFR (β=0.3; P=0.004) and a greater rate of eGFR decline with age ($\beta_{interaction}$=-0.02; P=5x10^{-5}), independent of albuminuria, kidney injury, or hyperfiltration incidence at age <13. Thus, *APOL1* is associated with development of early albuminuria and later decline in kidney function in patients with SCD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4137T: The polygenic risk score of biological age in nine human organ systems

Authors:

J. Wen[1], Y. Tian[2], Z. Yang[3], L. Shen[3], A. Zaleskey[2], C. Davatzikos[3]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of Melbourne, Melbourne, Australia, [3]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: Human biological age is interconnected with biological processes encompassing multiple organ systems. The biological age gap (BAG) quantifies the difference between AI-predicted and chronological age. Our previous studies linked the BAG of nine organ systems with clinical and genetic factors. Here, we developed polygenic risk scores (PRS) for the nine BAGs and assessed their predictive capabilities, leveraging multi-omics data comprising 30,108 to 111,543 participants of European ancestry from the UK Biobank.
Methods: We calculated the organ-specific PRS using GWAS summary statistics in an age- and sex-matched, split-sample setting. Specifically, the weights of the PRS were defined based on split1 GWAS data (base/training data); the split2 GWAS data were used as the target/test data for PRS calculation. The quality check steps include i) removal of duplicated and ambiguous SNPs for the base data; ii) clumping of the base genotype data; iii) pruning of highly correlated SNPs in the target data; iv) removal of high heterozygosity samples in the target data; v) removal of duplicated, mismatching and ambiguous SNPs in the target data. After these, we used PRS-CS to generate PRS for the split2 population using the classic C+T approach. To determine the optimal PRS, we performed a linear regression using the PRS calculated at different P-value thresholds (i.e., 0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5) to predict the BAG, controlling for age, sex, and the first forty genetic principal components. For each BAG PRS, we present the incremental R^2 for all five P-value thresholds.
Results: In the split2 target population, all nine PRSs significantly improved the predictive performance for the nine BAGs (P-value<10^{-30}). Notably, the renal PRS showed the highest additional explained variance, contributing 5.92% (P-value threshold=0.1) to the renal BAG. This was followed by the metabolic PRS (R^2=5.67%, P-value threshold=0.05), pulmonary PRS (R^2=3.78%, P-value threshold=0.3), hepatic PRS (R^2=2.58%, P-value threshold=0.1), cardiovascular PRS (R^2=2.41%, P-value threshold=0.3), brain PRS (R^2=2.41%, P-value threshold=0.3), immune PRS (R^2=2.40%, P-value threshold=0.05), musculoskeletal PRS (R^2=2.21%, P-value threshold=0.4), and eye PRS (R^2=2.02%, P-value threshold=0.5).
Conclusions: The nine organ-specific PRSs demonstrate significant additional predictive capability. Subsequent studies will explore their combined potential with the nine BAGs to predict systemic chronic diseases in the general population.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4139T: The PRS-RV Causal Pivot: A Segmentation Approach to Complex Disease Genetics

Authors:

C. Shaw[1], N. Di[2], T. Tan[1], C. J. Williams[3], J. Shulman[4], J. Belmont[5]; [1]Baylor Coll. of Med., Houston, TX, [2]Rice Univ., Houston, TX, [3]Genetics and Genomics Services, Houston, TX, [4]Duncan Neurological Res Inst, Houston, TX, [5]Genetics & Genomics Services, Inc., Houston, TX

Abstract:

Genetic contributions to complex disease can be modeled as the collective influence of common variation alongside individual influences of rare variations (RV) with strong effects. We propose a structural approach to segment complex disease into four causal types: 1) polygenic - driven by collective influences of common variation, 2) monogenic - driven by RV, 3) those that are blended, and 4) disease individuals who lack either type and include environmental phenocopies. We develop a causal graph with a polygenic risk score (PRS) factor for common variation and a binary factor for pathogenic RV. Disease outcome is modeled as their combined effect. Under this model, RV is independent of PRS in the general population but conditioning on disease induces PRS-RV correlation. The unconditional independence assumption can be tested, and violations can be overcome using methods from causal analysis. We consider the potential impact of our model for complex disease genetics. We propose the phenomena of outcome-induced PRS-RV correlation can be exploited for discovery. The PRS is a causal pivot - a known factor against which candidate causes can be opposed. RV that unconditionally lack association with PRS but which have induced association conditional on disease are candidate causes. Mathematical analyses using a likelihood ratio test demonstrate this approach has good statistical power when RV effects are approximately five-fold that of a PRS. The model also has study design implications. The model predicts that in cases-only cohorts those individuals with the lowest PRS have the highest likelihood of causal RV. We demonstrated the impact of this phenomena for improved discovery in cases-only designs by simulation studies. To test the relevance of this framework in human disease we use Europeans from the UK Biobank (UKBB). We focus on positive control complex diseases of hypercholesterolemia (HC, LDL-c direct ≥ 4.9 mmol/L) (n=27,083) breast cancer (BC, female, ICD10 C50) (n=12,913) and Parkinson's Disease (PD, ICD10 G20) (n=2,950). We use exomes and enumerate ClinVar pathogenic RVs; for HC we focus on *LDLR* (125 RV), for BC we focus on *BRCA1* (106 RV), for PD we use *GBA* (61 RV). We obtain these results: 1) no correlation between RVs and PRS in the general population; 2) significant negative correlation between RV and PRS among affected individuals within diseases and 3) no significant associations in cross-disease analyses of RV and PRS as a negative control. Taken together our results demonstrate the potency of the PRS-RV Causal Pivot to drive discovery, to inform research design and to improve reasoning in human genetics.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4140T: Transcriptome-wide association analysis identifies novel genes underlying coronary heart disease and stroke using kidney tissue and brain cell type transcriptome in individuals with chronic kidney disease

Authors:

W. Guan, J. Wen, B. Lin, Q. Sun, M. Jiang, S. Huang, L. Zhou, Z. Lan, Y. Li, N. Franceschini, CRIC Study Investigators; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

This study aims to identify genes associated with coronary heart disease (CHD) and stroke through integrating genome-wide association study (GWAS) summary statistics and kidney/brain eQTLs in high-risk individuals with chronic kidney diseases (CKD) using transcriptome-wide association studies (TWAS). Through summary statistics-based TWAS, we identified 11 genes (42 gene-trait associations) using predicted gene expression in kidney, 190 genes (244 associations) in astrocytes, 247 genes (320 associations) in endothelial cells, 284 genes (346 associations) in excitatory neurons, 252 genes (328 associations) in inhibitory neurons, 169 genes (215 associations) in microglia, 223 genes (274 associations) in oligodendrocytes, 249 genes (293 associations) in oligodendrocyte precursor cells/committed oligodendrocyte precursors (OPCs/COPs), and 200 genes (241 associations) in pericytes, after Bonferroni correction. Among the 11 genes revealed from kidney eQTL-based TWAS, 8 (72%) have been reported as cardiovascular disease (CVD) risk genes in either the GWAS Catalog or external TWAS databases (including TWAS Hub, TWAS Atlas, webTWAS), within relevant tissues. For brain eQTL-based TWAS, we observed 72-83% of the cell-type-specific TWAS genes (mean = 227 genes per cell type) were similarly supported by literature. Among the 306 remaining genes identified as novel for CVD risk, 59% are lncRNA and 39% are protein-coding. One protein-coding gene *YY1*, identified by brain eQTL TWAS for fatal myocardial infarction and congestive heart failure in neurons, is a ubiquitous transcription factor involved in central nervous system development and highly expressed in neurons, cerebellum, and white matter according to multiple data sources. Based on UK Biobank data, ExPheWas links *YY1* to multiple CKD and CVD risk factors, including coronary atherosclerosis, creatinine, glycated hemoglobin, cystatin C, and high light scatter reticulocyte percentage (q-value < 0.05). In summary, using TWAS, our study identified novel genes associated with CVD in individuals with CKD who are at high-risk for CHD and stroke, which offers insights for disease biology and potential therapeutic approaches.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4141T: Transferability of a European-derived Alzheimer's Disease Genetic Risk Score across Multi-Ancestry Populations.

Authors:

A. Nicolas, B. Grenier-Boley, J-C. Lambert; Inst. Pasteur de Lille, Lille, France

Abstract:

The most complete genetic landscape of Alzheimer's disease (AD), in European-ancestry populations, includes 75 loci. The resulting Genetic Risk Score (GRS) is associated with an increased risk of developing future AD/dementia and with an increased risk of conversion from mild cognitive impairment to AD/dementia. This study aimed to evaluate the association of this GRS with AD and more broadly with dementias, in the context of different genetic ancestries. We observed a consistent association of this GRS with AD risk in 15 European countries, age at onset and cerebrospinal fluid (CSF) AD biomarker levels, regardless of the Apolipoprotein E (APOE) genotype. Then, we extended this AD risk association to multi-ancestry populations and assessed a decreasing order of magnitude in European-American, North-African, East-Asian, Latin-American, African-American, sub-Saharan African and Indian genetic backgrounds, respectively. This GRS captured information more specifically related to AD as its association decreased as the diagnosis broadened to AD and related dementias and to dementias, in European, Hispanic admixed and African Admixed genetic ancestries. To improve its performance in African American and Latino American genetic ancestry groups, we developed a Polygenic Risk Score, using PRSCSx and multi-ancestry summary statistics, tested in African American, Latino-American and East Asian ancestry groups. We also added variants to the initial GRS, surrounding the European hits, but independent and associated with AD (P-value < 0.001) in diverse ancestry GWASes. In conclusion, a larger population diversity is needed to better understand the AD genetic complexity across genetic ancestries and to include this genetic diversity in a GRS/PRS. This GRS/PRS, combined with other information such as APOE status, CSF pTau and CSF Ab1-42, is a first step towards AD risk prediction and personalized medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4142T: Triangulation of the causal relationship between plasma proteins and systemic lupus erythematosus

Authors:

K. Yu, D. Baird, Y. Xu, D. Chasioti, C-Y. Chen, T. Reynolds, D. Malhotra, E. Tsai; Biogen, Cambridge, MA

Abstract:

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by diverse manifestations, including arthritis, skin disease, and nephritis. Genome-wide association studies (GWAS) have identified some genetic risk loci associated with SLE, yet the disease etiology remains unclear. Recent advances in proteomics have opened new avenues for understanding disease mechanisms. In this study, we performed a differential protein expression analysis on 2,941 proteins among individuals of European ancestry in the UK Biobank Plasma Proteomics Project (UKB-PPP), comparing 300 prevalent SLE cases to 2,437 healthy controls. We identified 271 proteins with significant different levels between SLE cases and controls (p-value < 1.7e-5; Bonferroni correction N = 2,941). To further investigate the causal relationship between plasma proteins and SLE, we performed Mendelian Randomization (MR) using protein quantitative trait loci (pQTLs) from the UKB-PPP as instruments. Prior to MR, we conducted a meta-analysis of SLE risk incorporating data from published GWAS studies and FinnGen, including a total of 12,997 SLE cases and 392,272 controls of European ancestry. We identified 27 proteins with causal effect on SLE risk (p-value < 5.2e-5; Bonferroni correction for the number of proteins tested N = 955). Five of them overlapped significant signals identified in the case-control analysis with consistent direction of effect, including tumor necrosis factor (*TNF*; MR b=3.63, p=9e-10; case-control b=1.05, p=7e-23), a well-known pathological cytokine in SLE and several other autoimmune disorders. Future work will focus on evaluating the impact of SLE disease status on plasma protein level via reverse MR and incorporating expression QTLs to identify cell and tissue types enriched for SLE genetic signals.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4143T: TRUHiC: A transformer model incorporating the U-2 Net to enhance Hi-C data for 3D chromatin structure delineation

Authors:

C. Li, M. Mowlaei, HGSVC Functional Analysis Working Group, V. Carnevale, X. Shi; Temple Univ., Philadelphia, PA

Abstract:

High-throughput chromosome conformation capture (Hi-C) technology has been extensively employed to investigate the three-dimensional (3D) structure of genomes and chromatin folding. Key structures like topologically associating domains (TADs), TAD boundaries, and chromatin loops help reveal the relationships between functional regulatory elements and disease-associated genes. However, accurately identifying these regions depends on Hi-C data of high resolution, which is often costly and requires much deeper sequencing depths than traditional sequencing experiments.

To address this challenge computationally, deep learning (DL) convolutional neural networks (CNNs) have been proposed to enhance the resolution of Hi-C data. However, CNNs cannot capture long-range dependencies among data features in Hi-C, a shortcoming that can be addressed by using the attention mechanism in transformers. However, transformers themselves are limited in that they have a quadratic memory cost with regard to the feature size, making it prohibitive to use them for Hi-C data directly. Thus, we propose TRUHiC, a DL model utilizing a U2-Net architecture with an integrated transformer block. A U-2 Net is a nested U-Net, where a U-Net is an autoencoder with skip connections from the encoder to the decoder to mitigate information loss due to the compression at the bridge block. We replace the bridge block in a U2-Net with a transformer to leverage the benefits of a transformer, seeking a trade-off between feature granularity and memory consumption.

Additionally, we introduced a directionality index loss function to better constrain the target region during training and pre-trained TRUHiC on published Hi-C data from human lymphoblastoid cell lines. We conducted comprehensive experiments to verify the generalization of our model, demonstrating its ability to significantly enhance the low-resolution Hi-C data from different downsampled ratios across various cell types and human and mouse cell lines (K562, IMR90, NHEK, and CH12-LX). Experimental results showed that our model excels in both technical and biological metrics, outperforming other DL-based Hi-C data enhancement models (like DFHiC), resulting in higher overall structural similarity and consistency scores for characterizing TADs and loops.

In summary, this study presents a novel method utilizing the transformer's attention mechanism for capturing long-range genomic interactions and demonstrates its application for 3D genome research. It provides a new generative model trained on extensive Hi-C data to alleviate the challenges posed by insufficient Hi-C data resolutions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4144T: Uncovering disease-critical pathways for diabetes through global and local embeddings of CRISPR perturbation data

Authors:

T. Fabiha, D. Liu, D. Yang, D. Huangfu, K. Dey; Sloan-Kettering Inst. of Mem. Sloan-Kettering Cancer Ctr., New York, NY

Abstract:

Decoding the cellular programs and pathways of disease genes is critical to understand the functional architecture of disease and inform drug targets; CRISPR Perturb-seq assays provide a promising avenue for uncovering these programs. Here, we propose global and local embedding methods to extract gene programs capturing gene regulatory alterations in response to a large or small number of perturbations. As a case study, we applied our methods to a CRISPR-ko Perturb-seq assay (N=88,355 cells) targeting 36 genes associated with Type 2 Diabetes and Maturity Onset Diabetes of the Young (MODY) Mendelian disorder along pancreatic development from ESC stage to SC-beta cell stage; omics readouts were observed at 4 different days (Days 3, 7, 11 and 18) thereby shedding light on different cell types represented at different development stages. We first extracted gene programs representing shared perturbation effects across many perturbations using the Penalized Matrix Decomposition (PMD) approach; this method was performed on data from each developmental stage as well as across all stages. 16 and 25 out of 40 PMD programs showed significant associations (FDR < 0.05) with MAGMA GWAS genes for T2D genes and MODY Mendelian genes respectively. We also observed 1.6x and 1.7x higher associations with T2D and brain-related traits for late pancreatic development gene programs (Days 11, 18) compared to early development (Day 3). Many of these late development gene programs showed synergistic disease information for T2D and MODY. To detect gene programs downstream of one or a small set of perturbations, we performed a random walk based embedding approach, called PerturbWalk, on a knowledge graph depicting perturbations, differentially regulated genes and underlying cell type and developmental stage of action. PerturbWalk applied to Day 18 was able to successfully distinguish perturbations and differentially regulated genes in endocrine and non-endocrine celltypes, while also capturing shared and distinct profiles of Sc-alpha, Sc-beta and SC-delta celltypes among the endocrine cells. We performed k-means clustering of the embeddings to identify hubs of co-regulated perturbations and genes, and evaluated these hubs for disease signal. We observed 24 hubs with significant enrichment or depletion for T2D and MODY genes; notably, the hubs with significant depletion in T2D tended to be shared by multiple perturbations (average 70% more perturbations than non-disease-related hubs). In summary, our study provides a new framework for identifying gene programs for CRISPR-KO Perturb-seq assays as well as benchmarking to understand the role of these programs for complex diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4145T: Unique Signatures of Highly Constrained Genes Across Publicly Available Genomic Databases

Authors:

K. Schmitz-Abe, Q. Li, M. Borrelli, M. Chinnathevar Ramesh, R. Agrawal, S. Luo, P. Agrawal; Univ. of Miami, Miami, FL

Abstract:

Publicly available genomic databases play a crucial role in understanding human population genetic variation. The Genome Aggregation Database (gnomAD) is the most extensive and widely used, publicly available collection of population variants from harmonized sequencing data. It is crucial in analyzing and interpreting candidate variants and novel genes in rare disease. This study delves into the top 2% of genes, analyzing high-constraint missense and loss-of-function Z-scores in gnomAD to explore their potential relationship with human disorders. Gene ontology analysis allowed us the functional enrichment of those high-constraint genes. They were also evaluated along with HGMD and ClinVar, further emphasizing their relevance and potential for advancing rare disease knowledge. This study analyzed 236 extremely constrained genes for missense and/or loss-of-function variants and 132 non-constrained genes (N-C) for both missense and loss-of-function. These 236 genes were sorted into three groups: 1) 138 genes constrained for both loss-of-function and missense (LoF-C & ms-C) variants, 2) 63 constrained genes for only loss-of-function (LoF-C), and 3) 35 constrained genes for only missense (ms-C). Each group had a unique and distinctive genetic signature. Of the 236 high-constraint genes, 144 genes are known to be associated with human disease in OMIM, and over 70% of those are dominant genes. In comparison, only 25.9% of the N-C group's OMIM known genes are dominant ($p = 5.7e-07$). The remaining 92 genes (41 from LoF-C & ms-C, 36 from LoF-C, and 15 from ms-C) are not associated with human diseases. Using GO enrichment analysis, we found that each group was enriched for certain pathways. For example, genes in the LoF-C & ms-C group are found to be highly enriched in histone modification and ATP-dependent chromatin remodeler activity; genes in the LoF-C group are enriched for genes encoding for lysine N-methyltransferase activity and histone methyltransferase complex; genes in ms-C group are associated with a structural constituent of cytoskeleton. Furthermore, N-C group genes are highly enriched in sensory perception of taste and protein tyrosine kinase binding. By integrating constraint metrics with gene ontology and disease databases, this study highlights the importance of this approach in identifying significant genes and prioritizing candidates for

rare genetic disorders. The highly constrained biological pathways based on our study are also of great interest.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4147T: Univariate and Multivariate Proteome-Wide Association Studies to Identify Causal Proteins for Alzheimer's Disease in the Presence of Invalid Instruments with GWAS Summary Data

Authors:

L. Fang[1], H. Xue[2], Z. Lin[3], W. Pan[4]; [1]Univ. of Minnesota, Minneapolis, MN, [2]City Univ. of Hong Kong, Hong Kong, Hong Kong, [3]Florida State Univ., Tallahassee, FL, [4]Univ. of Minnesota, MINNEAPOLIS, MN

Abstract:

Alzheimer's disease (AD) is a complex and progressive neurodegenerative disorder that accounts for the majority of dementia cases. Here we aim to identify causal plasma proteins for AD, shedding light on physiological and pathological processes of AD. We utilized the latest large-scale plasma proteomic data from UK Biobank Pharma Proteomics Project (UKB-PPP) and AD GWAS summary data from the International Genomics of Alzheimer's Project (IGAP). Via a new and robust univariate instrumental variable (IV) regression method, we identified 15 causal proteins through both cis- and trans-pQTLs (called trans-pQTLs for simplicity here) and 7 causal proteins through cis-pQTLs. To further reduce potential false positives due to high LD of some pQTLs and high correlations among some proteins, we developed a novel multivariate IV regression method for fine mapping to distinguish direct and mediating (causal) effects of proteins; some key features of the method is its robustness to invalid IVs and applicability to GWAS summary data. Application of our method further reduced the total number of causal proteins to nine. Among those identified proteins, APOE, CR1, CD2AP, TREM2 were also validated based on previous studies. Our work highlights some key differences between trans-pQTL and cis-pQTL analyses, and critical values of multivariate IV analysis for fine-mapping causal proteins, providing new insights into plasma protein pathways to AD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4148T: Unravelling the gene regulatory networks driving the polygenetic risk of human complex diseases

Authors:

H. Takahashi[1,2], H. Hatano[2], M. Nakano[2], Y. Tsuchida[3], S. Sumitomo[3], A. Suzuki[4], Y. Kochi[5], K. Fujio[3], K. Yamamoto[4], K. Ishigaki[2,1,6]; [1]Dept. of Microbiol. and Immunology, Keio Univ. Sch. of Med., Tokyo, Japan, [2]Lab. for Human Immunogenetics, RIKEN Ctr. for Integrative Med. Sci., Kanagawa, Japan, [3]Dept. of Allergy and Rheumatology, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [4]Lab. for Autoimmune Diseases, RIKEN Ctr. for Integrative Med. Sci., Kanagawa, Japan, [5]Dept. of Genomic Function and Diversity, Div. of Biological Data Sci., Med. Res. Inst., Tokyo Med. and Dental Univ., Tokyo, Japan, [6]Keio Univ. Human Biology-Microbiome-Quantum Res. Ctr. (WPI-Bio2Q), Tokyo, Japan

Abstract:

Human common diseases have highly complex pathology. However, leveraging their causal variants as an analytical tool has accelerated our understanding of disease mechanisms. Previous studies suggested that causal variants contribute to pathology by disrupting transcription factor (TF) activity and subsequent transcriptome dysregulation, given their enrichment at TF binding sites. While various publicly available transcriptome databases hold the potential for elucidating disease mechanisms, there needs to be more pipelines to extract transcriptome components unbiasedly and integrate them with GWAS results. Therefore, we aimed to construct an unsupervised analytical pipeline to identify TF-mediated gene expression regulatory networks (TF-GRN) underlying multiple transcriptome datasets and leverage them to assess complex disease pathology. Our method comprises three steps: i) construction of the minimal units of TF-GRN based on the results of TF ChIP-seq; ii) unsupervised integration of transcriptomes with TF-GRN units using canonical correlation analysis (CCA) to create an unbiased and comprehensive series of TF-GRN annotations; and iii) heritability enrichment inference within the TF-GRN annotations using stratified linkage disequilibrium score regression (S-LDSC). As a benchmarking measure, our method outperformed LDSC-specifically expressed genes (SEG) regarding heritability enrichment for autoimmune diseases. Furthermore, we demonstrated the cross-species applicability by applying our method to a mouse transcriptome dataset. Finally, we addressed a challenge inadequately tackled by traditional research methods: applying to clinical transcriptome datasets obtained from a cohort with multiple autoimmune disease patients. Our method successfully identified autoimmunity heritability enrichment among disease-relevant TF-GRN inferred in autoimmunity clinical transcriptome. For instance, a component of memory CD4+ T cell transcriptome reflecting IFN-related pathway, segregating cases and controls, exhibited strong enrichment for the heritability of autoimmune diseases (systemic lupus erythematosus (SLE): -log10P=4.6, rheumatoid arthritis (RA): -

log10P=9.0). This effectiveness of our method was further reinforced by identifying components that exhibited specific enrichment in RA, SLE, and Crohn's disease. Notably, enriched components did not necessarily have the largest variance, indicating that the conventional supervised approach was inefficient in capturing such components. Our study provides insights into the disease pathology, which will be a solid foundation for future studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4150T: Whole genome sequence-based association study on the multiple system atrophy in the Japanese population

Authors:

K. Orimo[1], J. Mitsui[1], T. Matsukawa[1], H. Ishiura[2], M. Tanaka[3], J. Nomoto[4], K. Tokunaga[5], Y. Omae[5], Y. Kawai[5], Japan Multiple System Atrophy Registry Consortium, T. Toda[1], S. Tsuji[1]; [1]Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [2]Okayama Univ., Okayama, Japan, [3]Intl. Univ. of Hlth.and Welfare, Chiba, Japan, [4]Intl. Univ. of Hlth.and Welfare, Tokyo, Japan, [5]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Abstract:

Introduction: Multiple system atrophy (MSA) is a rare neurodegenerative disorder characterized by various combinations of parkinsonism, cerebellar ataxia, and autonomic dysfunction. We have demonstrated that biallelic functionally deleterious *COQ2* variants are involved in familial MSA and a monoallelic variant, *COQ2* p.V393A (NM_015697), is associated with sporadic MSA in Japanese as well as in East Asian populations. Although several genome-wide association studies (GWAS) have reported various loci as potential risk variants for MSA, including *SNCA, MAPT, RABGEF1, KCTD7, GAB1, TENM2*, *PLA2G4C*, *lnc-LRRC49-3*, *ZIC1*, *ZIC4*, and *GBA1*, few have been sufficiently replicated. This study aimed to identify novel susceptible variants/genes associated with MSA as well as to examine the replication of previously reported loci based on the whole-genome sequence (WGS) analysis in the Japanese population. **Methods:** A total of 561 Japanese patients with MSA and 9,850 healthy individuals or controls from the National Center Biobank Network (Kawai *et al.* PLoS Genet. 2023) were included in this study. WGS analysis was conducted using Novaseq 6000 sequencers, followed by mapping to GRCh38 and variant calling using Parabricks (v3.1.0), equivalent to bwa (v0.7.15) and GATK (v4.1.0). Joint-calling was conducted using the gVCFtyper of the Sentieon package. After quality control, association analysis was conducted using plink (v1.9) focusing on variants with minor allele frequencies (MAF) larger than 1.0%. **Results:** GWAS focusing on common variants (MAF > 0.01) only revealed *COQ2* p.V393A with a genome-wide significance ($P = 1.0 \times 10^{-9}$, OR = 2.50; 95% CI,

1.87 to 3.36). Of the previously reported candidate susceptible genes, GWAS further revealed *PLA2G4C* p.Ile143Val (NM_003706) with $P = 7.76 \times 10^{-5}$, OR = 1.28; 95%CI, 1.13 to 1.44. The association analysis did not sufficiently replicate other previously reported risk alleles. **Conclusions:** The GWAS based on the whole-genome sequence data from the Japanese population replicated the association of *COQ2* V393A with MSA. To achieve a sufficient power to test association of an enormous number of variants revealed by WGS, the sample size of the current study is relatively small. WGS-based association studies of much larger sample sizes along with data sharing with other WGS data on MSA are warranted.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Thursday Session

Board 4151T: XPAT2: A Quality Control Framework for Sequencing-Based Association Studies Using Biobank Scale Data ⭐

Authors:

Y. Yu[1], R. Bohlender[2], P. Lupo[3], A. Sabo[3], P. Sok[3], C. Huff[4]; [1]UT MD Anderson Cancer Ctr., Houston, TX, [2]U.T. MD Anderson Cancer Ctr, Houston, TX, [3]Baylor Coll. of Med., Houston, TX, [4]U.T. MD Anderson Cancer Ctr, Pearland, TX

Abstract:

The increasing number of high-throughput sequencing studies, supported by various fund sources and Biobank initiatives, now provides genetics data from millions of samples. This unprecedented access opens up new possibilities for large-scale association studies that were previously unachievable. However, the growing heterogeneity of such data presents significant challenges to mega-analysis efforts. Additionally, emerging variation prioritization methods, such as SpliceAI, could significantly enhance the genetic association study if properly incorporated. We previously published XPAT, a toolkit designed to conduct sequence-based case-control analyses using cross-platform data. XPAT includes functions for joint genotype calling, quality control, association testing, and effect size estimation. Here, we introduce XPAT2, which is specifically designed to facilitate computationally efficient and robust analyses for biobank scale WES and WGS data. XPAT2 interfaces with several tools to detect cryptic relationships with high precision and incorporates various variant prioritization methods, including CASM, REVEL, and SpliceAI, to provide comprehensive annotation for association tests. Additionally, XPAT2 includes methods for rare variant effect size estimation through both mega- and meta-analyses. XPAT2 is now deployed as a Docker image for easy execution and reproducibility. We applied XPAT2 to case-control studies of multiple cancers using WES data from over 200,000 samples in the UK Biobank. We also conducted joint genotype calling and

case-control analyses using WGS data from approximately 11,000 samples from the Kids First Data Resource Center. The Kids First contains data generated from different projects with various experimental batches and specimen types (e.g., blood and saliva), presenting a significant challenge in QC due to data heterogeneity. XPAT2 successfully captured noise due to batch effect bias and controlled type I error inflation in these analyses. We applied XPAT2 to a WES study of rhabdomyosarcoma (RMS) involving approximately 749 RMS case, 1977 ARIC controls, and 299 VIVA controls. The genotypes were pre-called using xAtlas at Baylor. We observed no inflation in p-values (lambda=1.01) and detected nominally significant associations with established cancer predisposition genes. XPAT2 offers an efficient and comprehensive solution to the statistical and computational challenges in WES and WGS studies, enabling more accurate and reliable identification of disease-associated variants in increasingly large and diverse sample populations.

Friday, November 8, 2024, 2:30 pm - 4:30 pm

Session Title: Cancer Poster Friday Session

Board 8001F: A deep learning-based precision oncology framework integrating tumor omic and functional characterization for personalized treatment selection

Authors:

C. Sederman[1], D. Tony[1], S. Scherer[2,3], C-H. Yang[2,3], E. Cortes-Sanchez[2,3], X. Huang[1], Y. Qiao[1], B. Welm[2,3], A. Welm[2,3], G. Marth[1]; [1]Univ. of Utah Dept. of Human Genetics, Salt Lake City, UT, [2]Univ. of Utah Huntsman Cancer Inst., Salt Lake City, UT, [3]Univ. of Utah Dept. of Oncological Sci., Salt Lake City, UT

Abstract:

Despite several successful applications of genome-guided cancer therapy, a lack of rational guidance for treatment selection remains a critical barrier to precision oncology practice. Most tumors lack actionable molecular alterations, limiting the scope of biomarker-informed treatment selection. Further, response to biomarker-informed treatment is highly variable, even among patients harboring theoretically targetable alterations. Given these limitations, nonspecific chemotherapy remains a mainstay of cancer treatment, yet few data-driven tools exist to guide chemotherapy selection. Collectively, these considerations inform the need to expand guidance for treatment selection to tumors lacking actionable molecular alterations and to drugs lacking validated predictive biomarkers. To this end, we present a hybrid

computational/experimental approach, leveraging a clinically oriented functional precision oncology pipeline in concert with a deep learning (DL)-based cancer drug response prediction framework, ScreenDL, engineered to capitalize on the patient-specific data generated by this clinical workflow. ScreenDL leverages a phased training schema, incorporating general purpose pretraining in cell lines and fine-tuning using a newly generated pharmaco-omic resource spanning >100 advanced/metastatic breast cancer (BC) patient-derived xenograft (PDX)-derived organoid (PDxO) models. As new patients enter our precision oncology pipeline, parallel collection of tumor omics and functional drug screening in patient-derived organoids (PDOs) enables patient-specific fine-tuning of ScreenDL with our ScreenAhead module, generating a personalized response prediction model optimized for the N-of-1 precision oncology setting. ScreenDL outperforms existing DL models in never-before-seen tumor samples, achieving a median Pearson correlation between observed and predicted response across drugs of 0.49 in cell lines and 0.26 in our BC PDxO cohort. In both contexts, ScreenAhead improves performance, achieving median correlations per drug of 0.68 and 0.54. Further, ScreenDL consistently outperforms approved single- and multi-gene biomarkers, the current clinical practice. Mirroring our proposed clinical application to patients and matched PDOs, we applied ScreenDL to select precision treatments for a subset of 7 PDX models with matched PDxO drug screening. With ScreenAhead, ScreenDL achieves a response rate of 77% amongst selected therapies compared to 44% for the next best tested DL model, demonstrating the vast potential of our hybrid computational/experimental strategy for clinical precision treatment selection.

Session Title: Cancer Poster Friday Session

Board 8002F: A Discovery of Novel Methylation Profiles and Race-Specific Methylated Sites for Breast Cancer Risk

Authors:

Y. Wu[1], C. Thompson[2], F. Schumacher[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Penn State Coll. of Med., Hershey, PA

Abstract:

Breast cancer (BrCa) is the most common cancer in women worldwide. Black women with BrCa, on average, have an earlier age of onset, more aggressive subtypes and experience higher mortality compared with White women. Familial studies estimate ~ 40% heritability of this malignancy, yet the underlying genetic mechanisms remain uncertain. Alterations in DNA methylation contribute to BrCa development by directly impacting changes in gene expression. This study aims to identify novel CpGs associated with BrCa risk as well as by self-reported race.

To contextualize our results and improve statistical power, we explore potential biological pathways of the identified CpGs. A case-control study of BrCa was conducted among women diagnosed at University Hospitals in Cleveland, Ohio. DNA methylation levels were measured in 191 blood samples using Illumina Infinium MethylationEPIC (850K) array. After quality control, M-values for 191 samples and 807,893 methylation probes were evaluated. Differentially methylated positions were identified to determine hypo- (fold change < 0.67) and hypermethylated (fold change > 1.5) sites between BrCa cases and disease-free controls, adjusting for age. Stratified analyses were performed based on self-reported races. Over-representation of CpG probes classified as KEGG and GO pathways were contrasted against background CpGs using hypergeometric tests in R *missMethyl* package. Significance level was set at a false discovery rate (FDR) < 5%. Among 191 female participants, 78% (N=149) were diagnosed with BrCa, 34% were self-reported Black women and the average age was 56.7 ± 11.7 years. At FDR < 5%, 989 (5.7%) hypo- and 16,379 hypermethylated sites were identified for overall BrCa risk. Probe cg25141069, located in the obesity-related metabolic gene *HSD17B8,* exhibited hypomethylation pattern in overall BrCa cases compared to controls. Additionally, 5021 CpGs associated with BrCa risk only in White women, while 7109 CpGs were identified in Black women. Probe cg12858166 in HLA-*DPA1* gene and cg09797202 in the 3' UTR in *SCD* gene were hypomethylated exclusively in Black women. Significant biological mechanism enriched among Black participants included viral carcinogenesis and spinocerebellar ataxia pathways. Our study identified several candidate genes (e.g. *HSD17B8*) may contributing to BrCa through multiple biological pathways and discovered self-reported race-specific methylation sites for BrCa risk. For example, *HSD17B8* functions in steroid metabolism, supporting the potential role of steroid hormones in BrCa development. Further studies will be warranted to confirm our findings and determine clinical impact in BrCa.

Session Title: Cancer Poster Friday Session

Board 8003F: A machine learning approach towards prediction of pancreatic cancer using gene expression and DNA methylation

Authors:

Z. Caterer[1], P. Keller[2], S. T. Morehead[3], P. Nijhum[4], R. Gomes[2], R. Jansen[5]; [1]Univ. of Colorado Boulder, Boulder, CO, [2]Univ. of Wisconsin Eau Claire, Eau Claire, WI, [3]Univ. of Wisconsin, Eau Claire, WI, [4]North Dakota State Univ., Fargo, ND, [5]Univ. of Minnesota, Minneapolis, MN

Abstract:

DNA methylation is a process that can affect gene accessibility and therefore gene expression. Methylation can affect genes that are associated with suppressing or contributing to tumor growth and progression. In this research, we examined the potential of developing a scalable feature selection and deep learning framework capable of processing high dimensional genomic datasets to identify methylation sites in the human genome responsible for pancreatic ductal adenocarcinoma (PDAC). The methylation data being analyzed consisted of almost 485,578 CPG markers. Feature selection using Random Forest and Anova was performed along with random undersampling. Selected markers and gene expression data was analyzed to identify enriched pathways. Genes identified in pathways were studied for tumor progression. When compared with a list of 339 genes that were identified from literature, our results returned 98 common tumor genes including “KRAS", "BRCA1", "BRCA2", "PALB2", "CDKN2A", "TP53", "SMAD4". This reveals that the stratified feature selection technique is indeed useful at identifying important features. Finally a deep learning classifier was developed for identifying patients with PDAC. Data for this research was obtained from TCGA-PAAD project. Current work is focused on analyzing gene expression data and combine outcomes with methylation dataset to get deeper insights into the disease mechanisms. PDAC, or pancreatic ductal adenocarcinoma, is the most prevalent kind of pancreatic cancer, making up over 90% of cases. Aggressive tumor growth, early metastasis, and resistance to therapy are characteristics of this highly fatal cancer. Improving diagnosis, prognosis, and treatment outcomes require a thorough understanding of the molecular pathways underlying PDAC development and progression. The proposed research outcomes could aid in the development of non-invasive diagnostic tests and personalized treatment strategies.

Session Title: Cancer Poster Friday Session

Board 8005F: A transcriptome-wide association study identifies *PHACTR4* as a potential mediator of prostate cancer survival.

Authors:

R. Klein[1], W. Fu[1], J. Rothstein[2], W. Sieh[3], X. Song[4]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [3]MD Anderson Cancer Ctr., Houston, TX, [4]Duke-NUS Med. Sch., SINGAPORE, Singapore

Abstract:

While numerous genetic variants and target genes associated with inherited risk of prostate cancer have been identified, these genetic alterations do not predict survival time after diagnosis. Instead, using a genome-wide association study (GWAS) approach, we identified

common polymorphisms associated with survival time in prostate cancer. To identify additional genomic loci associated with prostate cancer survival and their target genes, here we present the results of a transcriptome-wide association study (TWAS). We built models to predict genetically regulated gene expression (GReX) in normal prostate tissue, primary prostate tumors, and metastatic prostate tumors, using a Bayesian approach that incorporates data from numerous SNPs at each locus. Using these new models, we conducted a TWAS using summary statistics from a previous prostate cancer survival GWAS, which was conducted in the Malmo Diet and Cancer (MDC) study from Sweden. We identified 22 genes for which the GReX predicted for at least one of the 3 tissue models was Bonferroni-significantly associated with survival time. We then validated these findings in data from the Prostate, Lung, Colorectal and Ovarian (PLCO) trial. Of our initial findings, two validated at a nominal significance level of 0.05 in normal tissue and 5 were validated in metastatic tumors. Interestingly, all of these validated genes were located at the same locus on chromosome 1, between positions 28.3Mb and 29.2Mb. Of these genes, one was identified using the normal prostate GReX models. Increased levels of PHACTR4 (phosphatase and actin regulator 4) was associated with better survival ($p=1.1x10^{-7}$ in MDC and p=0.02 in PLCO). In gene expression data from prostate tumors in The Cancer Genome Atlas, levels of *PHACTR4* are significantly lower in tumors with Gleason score >= 8 compared with <= 7 ($p=3.2x10^{-4}$). PHACTR4 is known to inhibit protein phosphatase 1 activity. Protein phosphatase 1 is a known regulator of the retinoblastoma protein, a key protein in the cell cycle that is often mutated in metastatic prostate cancer. To begin to characterize the effects of *PHACTR4* on prostate tumors, we confirmed expression of *PHACTR4* mRNA and protein in a panel of prostate cancer cell lines.

Session Title: Cancer Poster Friday Session

Board 8006F: Adiposity distribution and risks of 12 obesity-related cancers: a Mendelian randomization analysis

Authors:

E. Hazelwood[1], L. Goudswaard[1], M. Lee[2], M. Vabistsevits[3], D. Pournaras[1], H. Brenner[4], D. Buchanan[5], S. Gruber[6], A. Gsur[7], L. Li[8], L. Vodickova[9], R. Grant[10], J. Samadder[11], N. Timpson[1], M. Gunter[12], B. Schuster-Böckler[13], J. Yarmolinsky[12], T. Richardson[1], N. Murphy[2], E. Vincent[1]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Intl. Agency for Res. on Cancer, Lyon, France, [3]Univ. of Exeter, Exeter, United Kingdom, [4]German Cancer Res. Ctr. (DKFZ), Heidelberg, Germany, [5]The Univ. of Melbourne, Parkville, Australia, [6]City of Hope Natl. Med. Ctr., Duarte, CA, [7]Med. Univ. of Vienna, Vienna, Austria, [8]Univ. of Virginia, Charlottesville, VA, [9]Inst. of Experimental Med. of the Czech Academy of Sci., Prague, Czech Republic, [10]Princess Margaret Cancer Ctr., Toronto, ON,

Canada, [11]Mayo Clinic, Phoenix, AZ, [12]Imperial Coll. London, London, United Kingdom, [13]Univ. of Oxford, Oxford, United Kingdom

Abstract:

Background
There is growing evidence that overall adiposity is causally associated with multiple types of cancer. However, the causal effects of adiposity distributed at different anatomical sites on cancer risk are uncertain.
Methods
We performed two-sample Mendelian randomization (MR), a genetic epidemiological approach, to examine the effects of five adiposity traits (abdominal subcutaneous adipose tissue (ASAT); visceral adipose tissue (VAT); gluteofemoral adipose tissue (GFAT); liver fat; and pancreas fat) on risks of 12 obesity-related cancers (endometrial, ovarian, breast, colorectal, pancreas, multiple myeloma, liver, kidney (renal cell), thyroid, gallbladder, oesophageal adenocarcinoma, and meningioma) and cancer subtypes/subsites, where available. Then, we identified molecular traits which could be mediating the causal effects identified. We used the largest available GWAS from European populations (sample size across all GWAS ranged from 8,407 to 728,896 (median: 57,382); cancer GWAS ranged from 279 to 133,384 cases (median: 4,532), 3,456 to 727,247 controls (median: 68,802)).
Findings
Our results showed variable effects of the adiposity traits on cancer risk. For instance, higher ASAT increased risks of endometrial cancer (odds ratio (OR) per standard deviation (SD) increase in ASAT = 1.79, 95% confidence interval (CI) = 1.18 to 2.71), liver cancer (OR per SD increase in ASAT = 3.83, 95% CI = 1.39 to 10.53) and oesophageal adenocarcinoma (OR per SD increase in ASAT = 2.34, 95% CI = 1.15 to 4.78), but reduced risk of some breast cancer subtypes (e.g. luminal B/HER2-negative-like OR per SD increase in ASAT = 0.54, 95% CI = 0.40 to 0.73). Similarly variable effects were seen in MR analyses evaluating causal effects of other adiposity traits on cancer risk and of adiposity distribution on levels of molecular traits. Further MR analyses suggested that total testosterone, fasting insulin, sex hormone-binding globulin and IGFBP-1 likely play important roles in mediating the effects of adiposity distribution on obesity-related cancers.
Conclusions
Our analyses provide novel insights into the variability of the effect of adiposity distribution on cancer risk, with respect to both adiposity trait and cancer type. Moreover, we extend these findings to demonstrate that adipose tissue at different anatomical locations may have distinct effects on adiposity-related molecular traits. These insights enhance our understanding of the complex relationship between adiposity and cancer risk and suggest that changes to adipose

tissue distribution may be an important consideration alongside the total reduction in adiposity for cancer prevention.

Session Title: Cancer Poster Friday Session

Board 8007F: An exploration of sex differences in the genetic architecture of pancreatic cancer using GXwasR

Authors:

B. Bose, C. Stanhope, L. Pesce, B. Stranger; Northwestern Univ., Chicago, IL

Abstract:

Pancreatic cancer, which constitutes 3% of all cancers and accounts for 7% of cancer-related deaths in the U.S., has a dire prognosis with only a 10% five-year survival rate. This disease often predominantly affects males at a ratio of 1.3:1 relative to females. Notably, it also exhibits significant sex differences in response to treatment. The heritability of pancreatic cancer has been estimated at ~21%, and we hypothesized that sex and genetics might interact to influence disease risk. To explore this, we developed GXwasR, a suite of computational tools enabling the evaluation of sex-differentiated genetic effects on complex traits. Utilizing GXwasR, we explored the impact of sex on the genetic basis of pancreatic cancer. We analyzed genotype and phenotype data from the Pancreatic Cancer Cohort Consortium (PanScan; N = 5,117 cases, N= 8,845 controls) and the Pancreatic Cancer Case-Control Consortium (PanC4; N = 4236 cases; N= 3862 controls). Following rigorous quality control and genotype imputation, we performed sex-stratified and sex-combined genome-wide and X chromosome-wide association studies (GWAS and XWAS, respectively) for each dataset, followed by fixed-effect meta-analysis. We estimated sex-combined and sex-specific heritability, and quantified genetic correlation between male-only and female-only associations. We have tested multiple mechanisms of sex differences, including sex-dependent liability thresholds and variants with sex-differentiated effects on risk. We confirmed thirteen previously reported pancreatic cancer-associated loci ($P < 5 \times 10^{-8}$), and discovered that for several previously reported loci, there are significant sex differences in the genetic effect on risk (SNP × sex interaction; $P < 0.05$). Furthermore, in the sex-stratified analyses, we discovered several suggestive novel loci that are not observed in the combined-sex GWAS. We are currently validating these results. We are developing and testing sex-combined and sex-specific polygenic risk scores and are testing for differences in performance for pancreatic cancer risk prediction. This initiative represents a transformative step in the understanding of pancreatic cancer risk and demonstrates the value of considering sex as a biological variable in biomedical research.

Session Title: Cancer Poster Friday Session

Board 8008F: An innovative approach for detecting cancer-associated germline variants using automated and multiplexed targeted HiFi read sequencing

Authors:

T. Risch[1], M. Sauer[1], S. Angerbauer[1], K. Lenhard[1], V. Steinke-Lange[1,2], A. Benet-Pages[1,3], A. Hallermayr[1,2], T. Neuhann[1], A. Abicht[1,4], E. Holinski-Feder[1,2]; [1]MGZ – Medizinisch Genetisches Zentrum, Munich, Germany, [2]Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Univ. München, Munich, Germany, [3]Inst. of Neurogenomics, Helmholtz Res. Ctr., Neuherberg, Germany, Munich, Germany, [4]Friedrich-Baur-Inst., Dept. of Neurology, LMU Univ. Hosp., Ludwig Maximilian Univ., Munich, Germany

Abstract:

Hereditary tumor syndromes (HTS), characterized by pathogenic germline variants significantly elevating the risk of developing cancer, are present in 5-10% of cancer patients. Identifying these genetic predispositions is crucial for tailoring screening and treatment to improve patient outcomes. DNA analysis based on short-read whole exome sequencing is a standard procedure for genetic HTS testing, missing pathomechanisms not primarily involving exonic sequences. To close this diagnostic gap, we developed an innovative workflow to enhance the diagnostic yield for HTSs using the power of long reads for targeted sequencing of whole gene bodies. Sequencing long DNA fragments promises to cover challenging genomic regions (such as repeats in introns), to improve variant calling in paralog regions via phasing (e.g., *PMS2*), and to enhance detection of indels > 20 bp and inversions. We developed a comprehensive workflow for genetic testing, encompassing every step from sample preparation to variant analysis, in collaboration with PacBio, TWIST, Beckman, and EnGenome. We designed a gene panel targeting 175 key HTS-associated genes (total target size of 14 Mb), enabling detailed genetic testing across twelve common hereditary tumor types. The automated laboratory workflow (Beckman Biomek i7 system) includes hybrid-capture-based target enrichment and a two-step library preparation using TWIST and PacBio reagents. HiFi read sequencing is subsequently conducted on the PacBio Revio platform. We can multiplex 24 samples per SMRT Cell, enabling the parallel sequencing of 96 samples across four SMRT cells on a Revio, which increases the maximum annual throughput to 31,200 samples, compared to the 1,300 samples possible with whole genome sequencing. Preliminary sequencing results showed a HiFi read length of ~4.5 kb and average target coverage of 58-65x which relates to 67-75x for exonic (coding and UTR), 57-

64x for intronic and 53-60x for 1kb up-/downstream gene-flanking regions. We implemented an API integration for EnGenome's eVai bioinformatics platform into our LIMS for seamless variant analysis. The whole workflow has a turnaround time of ~7.2 days. In addition to our efforts to advance genetic testing for hereditary tumor patients, this work provides a novel outlook on the utilization of long-read techniques in hereditary disease diagnostics. Here, we present our optimized workflow for long-read whole gene body sequencing, including laboratory workflow, detailed coverage analyses and lessons learned.

Session Title: Cancer Poster Friday Session

Board 8009F: Analysis of extensive data towards Genome in a Bottle benchmarks for a broadly-consented tumor normal pair

Authors:

J. McDaniel[1], J. Wagner[1], N. Olson[1], V. Patel[1], C. Xiao[2], A. Liss[3], J. Zook[1], Genome in a Bottle Consortium; [1]Natl. Inst. of Standards and Technology, Gaithersburg, MD, [2]NIH, Bethesda, MD, [3]Dept. of Surgery, Massachusetts Gen. Hosp., Boston, MA

Abstract:

Here, we detail in-progress genome-scale measurements from a variety of technologies of the first tumor-normal benchmark from the Genome in a Bottle (GIAB) consortium. We created the first broadly-consented tumor cell line (HG008-T) from a pancreatic ductal adenocarcinoma with matched normal pancreatic and duodenal tissue. Data is being collected from a large batch of the tumor cell line and paired normal tissues. As data are received they are made publicly available on the NCBI hosted GIAB FTP site. This dataset currently includes WGS measurements from Illumina, ONT, Hi-C, Bionano, single cell WGS data, PacBio HiFi, PacBio Onso, Element, and Ultima. Analysis and development of variant calling benchmarks using this data is ongoing in collaboration with an open working group of the GIAB consortium.
The tumor cell line has substantial aneuploidy from translocations that primarily cause large deletions. We found 16 chromosomes with extensive loss of heterozygosity due to missing greater than 30% of one copy, in addition to a couple hundred smaller structural variants. Low coverage single cell sequencing that provides ploidy estimates across chromosomes showed that most of the large deletions appear in all or nearly all cells with some variation in a fraction of cells. Bulk WGS analyses from two batches of cells from passages 23 and 36 both clearly showed evidence of deletions that occurred in all the single cells; however single cell and bulk WGS of passage 23 showed additional deletions were only present in some of the cells.. Additionally, many of the observed large deletions correspond to deletions seen in the

population of TCGA chromosome-scale Loss of Heterozygosity samples. Examining somatic SNVs in non-repetitive regions, we find that close to 60% occur in almost all cells in diploid regions, 30% occur in almost all cells in haploid regions, and 5% in only some cells in diploid and haploid regions. The tumor appears to be p53-deficient, and many cells appear to have wide-spread loss of heterozygosity followed by whole genome doubling, which was previously found to be common. Initial curations of chromosome 1 in the tumor genome shows evidence of whole genome doubling in a majority of the cells followed by deletion on the q-arm in a substantial minority of cells. Additional large structural changes on chromosome 1 are a translocation with chromosome 5, and a ~20kb deletion and ~95kb duplication. We are continuing curations of other chromosomes to establish a CNA/SV benchmark. We will also work towards curated small variant benchmarks, which can be used to evaluate accuracy of somatic variant calls, similar to traditional GIAB germline benchmarks.

Session Title: Cancer Poster Friday Session

Board 8010F: Approaches to returning germline and somatic test results across the Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network ⭐

Authors:

O. Bhattacharyya[1], A. L. Schuster[1], **E. B. Claus**[2,3], E. Linnenbringer[4], S. Ryan[5], U. Brown-Glaberman[5], D. W. Craig[6], J. O. Culver[7], K. A. Janeway[8], S. W. Gray[6], H. Hampel[6]; [1]Ohio State Univ., Columbus, OH, [2]Yale Univ, New Haven, CT, [3]Brigham and Women's Hosp., Boston, MA, [4]Washington Univ. in St. Louis, Saint Louis, MO, [5]Univ. of New Mexico, Albuquerque, NM, [6]City of Hope Natl. Med. Ctr., Duarte, CA, [7]Univ. of Southern California, USC Norris Comprehensive Cancer Ctr., Los Angeles, CA, [8]Harvard Med. Sch., Boston, MA

Abstract:

The NCI-funded Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network, comprising five research centers and a Coordinating Center, engages diverse populations in cancer genomics research. The Network prioritizes disseminating research results, including aggregate study updates and individual-level genomics findings, to foster trust and ongoing collaborations. This study examines diverse approaches by the five research centers (N=5) in returning germline and somatic test results. Research center members, familiar with the process, completed an online survey on result return strategies, conducted by the Coordinating Center. Germline testing protocols involve pre-test genetic counseling (n=3) via in-person, or video consultations. Participant preferences for receiving results are considered (n=5), with

return methods as direct lab-to-participant communication (n=1) and study team-mediated processes (n=4). Testing is initially conducted or confirmed in CLIA-approved labs (n=4), mainly using next-generation sequencing (NGS) panels (n=4), and whole exome sequencing (WES) (n=1). Results are returned during follow-up appointments (n=5), within 1-4 months, delivered via in-house genetic counselors (GCs) (n=2), oncologists (n=1), lab vendors (n=1), or national virtual GC service providers (n=1). Genetic counseling is universally provided for positive results, for most variants of uncertain significance, and upon request for negative results. While two centers do not return results to deceased participants or those lost to follow-up, three centers do so if requested by family members. Somatic testing prioritizes personalized result return strategies at all sites, with four using external and internal lab vendors ensuring CLIA compliance. Two centers exclusively use WES, and one uses an NGS panel. Others augment WES with other techniques, like whole genome sequencing (WGS) (n=1), and transcriptome analysis (n=1). Somatic results are returned to participants by either study staff (n=3) or oncologists via a study-team mediated process (n=2). Additionally, four centers actively disseminate participant-centered study updates, including study processes, enrollment details, participant spotlights, shared decision-making insights, and interpretation of genomic findings, with centers incorporating participants' feedback in developing result reports. The PE-CGS Network enhances the return of genomics findings through personalized delivery strategies and active participant engagement, addressing critical gaps in clinical actionability and participant preferences in cancer genetics research.

Session Title: Cancer Poster Friday Session

Board 8011F: Association of polygenic-based breast cancer risk prediction with patient management ⭐

Authors:

S. Ratzel[1], K. Johansen Taber[1], E. Hughes[1], A. Gutin[1], B. DeHart[2], L. Becker[2], J. Jasper[1], C. Miller[1], D. Chawla[1], P. Morin[2], J. Certa[2], A. Kurian[3]; [1]Myriad Genetics, Inc., Salt Lake City, UT, [2]Optum, Eden Prairie, MN, [3]Stanford Univ., Stanford, CA

Abstract:

Background: Guidelines recommend that individuals with ≥20% lifetime risk of breast cancer (BC) undergo enhanced management, including annual screening mammography (SM) as early as age 30, annual breast MRI, and genetic counseling (GC). Lifetime BC risk can be estimated using a validated risk model, such as Tyrer-Cuzick (TC). A risk predictor that combines TC with a polygenic risk score ("combined risk score" or CRS) significantly improves risk prediction over TC

alone. Little is known about how clinicians manage patients based on CRS. Here, we describe management following receipt of CRS results.
Methods: De-identified administrative claims data from the Optum Labs Data Warehouse were linked with de-identified TC and CRS results originally provided to ordering clinicians. Patients were divided into four subgroups based on their lifetime risk predicted by CRS and by TC alone (high risk: ≥20%, average risk: <20%): A (CRS-high/TC-high), B (CRS-high/TC-avg), C (CRS-avg/TC-high), and D (CRS-avg/TC-avg). Patient management claims were compared 360 days before and after the CRS test date and differences evaluated with McNemar's test. Differences in post-test management were analyzed using multivariable logistic regression with adjustment for clinical factors.
Results: The study cohort consisted of 8,662 patients. Based on CRS and TC results, 2443 (28.0%), 696 (8.0%), 856 (9.9%), and 4687 (54.1%) were placed in groups A, B, C, and D, respectively. Before receiving risk results, 11.9%, 7.7%, 8.8%, and 5.0% of patients in A, B, C, and D received a SM before age 40, respectively; 3.4%, 2.6%, 2.3%, and 0.9% received an MRI, respectively; and 4.3%, 4.7%, 6.0%, and 3.6% received GC, respectively. After receiving risk results, SM in those under age 40 significantly increased 1.6-2.2-fold in A, B, and C; MRI significantly increased 4.7-5.6-fold in A, B, and C; and GC significantly increased 1.9-2.3-fold in A and B. SM, MRI, and GC did not increase in D. Prophylactic mastectomy occurred in only 0.17% of all patients after receiving results. After adjustment for confounders, those in A, B, and C were 3.8-5.2 times more likely to undergo SM; 11.6-23.1 times more likely to undergo MRI; and 2.0-2.9 times more likely to undergo GC, compared to those in Group D (all $p<0.001$).
Conclusions: Patients with a ≥20% lifetime risk for BC were more likely to undergo enhanced management compared to those with a lifetime risk <20%, regardless of whether their risk was based on the CRS or on TC. These results suggest that clinicians recommended management aligned with guidelines for those with ≥20% lifetime risk, even when such risk was predicted by the CRS.

Session Title: Cancer Poster Friday Session

Board 8012F: Atavistic reversion in cancer through fusion genes and ancient protein structures

Authors:

S. Sun[1], C. H. Lineweaver[2], J. Broatch[3], K. J. Bussey[4]; [1]Arizona State Univ., Tempe, AZ, [2]Australian Natl. Univ., Canberra, Australia, [3]Arizona State Univ., Glendale, AZ, [4]Midwestern Univ., Glendale, AZ

Abstract:

Fusion genes, resulting from chromosomal translocations via nonallelic homologous recombination (NAHR), play a pivotal role in oncogenesis by producing fusion proteins that drive cancer aggressiveness. The atavism theory suggests cancer reverts to an ancient cellular state, reactivating primordial mechanisms. This study investigates if fusion proteins mirror ancient protein structures, as suggested by the atavism theory. Using the COSMIC database, we reconstructed fusion genes. We predicted the 3D structures of the resulting fusion proteins with AlphaFold and VAST analysis identified structurally homologous proteins. The evolutionary ages of the original and fusion proteins were inferred by mapping the original genes to homology groups from the Ensembl Compara database and assigning age based on divergence time using the TimeTree database. The findings revealed most fusion proteins have structures homologous to ancient proteins in humans, yeast, and bacteria, indicating a re-emergence of ancient protein structures in cancer cells (Permutation test, $p=0.0201$). Additionally, 68% of fusion genes include one gene predating the origin of multicellularity (~ 600 Mya) and another emerging with or after this milestone (One-sample proportions test, $X\text{-squared} = 13.291$, $df = 1$, $p = 0.00027$). This suggests fusion events during oncogenesis have ancient homologs, thus bridging evolutionary gaps between unicellular and multicellular life forms, explaining cancer's regression to primitive behaviors. Moreover, the preservation of kinase functionality in fusion genes, particularly those predating multicellularity, suggests a mechanism by which cancer cells adopt primitive behaviors. This aligns with the previous findings on the upregulation of ancient genes in cancer, emphasizing a shift to basal cellular functions. Kinases, due to their antiquity and roles in cell signaling and cycle control, appear instrumental in this regression. Our VAST analysis identified structures common to oncogenic fusion proteins like BCR-ABL1 and NPM1-ALK, representing active kinase conformations. This might explain apoptotic resistance in conditions like *BCR-ABL1*-mediated leukemogenesis. Conversely, structures homologous to the fusion protein CCDC6-RET possess tumor suppressor-like functions, suggesting some fusions confer survival advantages under stress, presenting new therapeutic strategies to slow disease progression.

Session Title: Cancer Poster Friday Session

Board 8013F: Breast Cancer Risk Polygenic Score optimisation through a novel SNP selection algorithm

Authors:

C. Debernardi[1], A. Savoca[1], E. Casalone[1], K. Zguro[2], G. Brunelli[2], Z. Yang[3], B. N. Wolford[4], A. Russo[1], S. Ripatti[5], G. Matullo[6]; [1]Univ. of Turin, Torino, Italy, [2]Univ. of Siena, Siena, Italy, [3]Univ. of Helsinki, Helsinki, Finland, [4]Univ. of Sci. and Technology, Trondheim, Norway, [5]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [6]Univ. of Torinio, Torino, Italy

Abstract:

Introduction: Many different Polygenic Scores (PGS) have been described for Breast Cancer (BC), differing by number and selection of SNPs. This variability leads to a diverse set of scores in the same individual, thus to an inconsistent risk stratification. The aim of this study is to optimise genetic variants selection from published BC PGS through additional prioritization steps and new weight estimation.
Methods: Starting from the 100 PGSs available for BC on PGS Catalog, we aligned them obtaining a dataset composed by 7,318,868 variants. SNPs selection approach based on the association with Breast Cancer and on the concordance between the different PGSs reducing the number to 1,106,069. Most variants appear to have a different effect on BC risk depending on the score taken into consideration. We therefore selected only SNPs with a weight always in the same direction. Standardization and embedding operations were performed on the effect weights to create a new score.
Results: The newly created PGS counts 4,606 SNPs. We tested this PGS on Genomic England data to identify the best set of weights and variants. The combined approach, which took the highest weight for dangerous SNPs and the lowest weight for protective ones, showed the best performance. Distributions between BC patients and controls were significantly different (p-value < 0.0001) and individuals in the highest 1% of the score had a three-fold increased risk of developing BC compared with those in the middle quintile. Lifetime risk of BC for women in the lowest and highest quintiles of the PRS were 10% and more than 25%, respectively. We further evaluated the performance of our new PGS in other biobanks involved in INTERnational consortium of integratiVE geNomics prEdiction (INTERVENE) project: UK Biobank, FinnGen and HUNT. In all the data sets tested the results were similar to those calculated in Genomics England. The comparison of this new PGS with those available in literature showed that this SNPs selection could work better than millions of variants used by other scores. The future goal would be to further reduce the number of SNPs in order to make this score more generalizable between different ancestries.
Conclusion: The optimization we proposed may allow a better generalizability of PRS, helping clinicians to improve BC patients' stratification.

Session Title: Cancer Poster Friday Session

Board 8014F: Case Report: Prostate Cancer Gene *HOXB13* Involved in Metastatic Omentum Carcinoma

Authors:

E. Eroglu Simsek[1], I. Akalin[1], I. Sunar[2], Z. Derin[3]; [1]Metagentech Ctr. for Genetic Diseases, Istanbul, Turkey, [2]Metagentech Ctr. for Genetic Diseases, ISTANBUL, Turkey, [3]Bilgi Univ., Istanbul, Turkey

Abstract:

Metastatic carcinoma in the omentum is a secondary cancer that typically arises from the spread of cancer cells originating from primary tumors such as those in the gastrointestinal system, ovaries, or pancreas.A 35-year-old female patient came with complaints such as nausea, vomiting, weakness and constipation. His parents are distant relatives.In the sections of omentum biopsy samples, malignant cellular infiltration with large, hyperchromatic, pleomorphic nuclei, narrow cytoplasm, and poorly differentiated morphology was observed under the meosthelial layer. Signet ring morphology was noted in some places in the cells. In the immunohistochemical study performed on the cell origin, PAX8(-), Ca125(-), CDX2(+), Melan A(-), calretinin(-), HBME1(-), cerbB2(-), pdl1(-) (TPS=0%) , msh2(+), msh6(+),mlh1(+),pms2(+) , BRAF V600E(-) were found in the tumor cells. It is thought that cytokeratin 7+ and cytokeratin 20+ findings can be interpreted in favor of metastasis of a malignancy that may have originated primarily in the upper gastrointestinal system such as the stomach or pancreas.The patient was recommended to undergo whole exome sequencing (WES). WES revealed *HOXB13* NM_006361.6 c.799C&gtT:p.Arg267Cys heterozygous pathogenic mutation which is associated with prostate cancer, hereditary, type 9 (MIM#610997).The *HOXB13* gene is a member of the HOX gene family in humans, located in the chromosomal region 17q21.32. This gene plays a role in the development of the posterior segment (tail region) of the body during embryonic development, particularly in the development of prostate, intestinal, and breast tissues. The *HOXB13* gene plays a significant role in regulating cell growth, division, and scaling. Alterations or mutations in the *HOXB13* gene can contribute to the development of certain cancer types. Specifically, certain mutations in the *HOXB13* gene can increase the risk of prostate cancer. These mutations are more commonly observed in individuals with a family history and can elevate the risk of prostate cancer. Research suggests that mutations in the *HOXB13* gene may increase the risk of prostate cancer and lead to a more aggressive form of the disease 9C&gtT:p.Arg267Cys heterozygous change was detected. The detected change was previously detected in gnomAD data.Their genes are highly conserved and essential for embryonic development. Changes in this geneIt has been seen in the literature to be associated with various cancers such as stomach, colon, breast and prostate (PMID: 33099213).

Session Title: Cancer Poster Friday Session

Board 8016F: Comprehensive somatic mutation analysis of oral leukoplakias

Authors:

A. Anne, K. Mohan; Birla Inst. of Technology and Sci. Pilani, Hyderabad Campus, Hyderabad, India

Abstract:

Background: Oral leukoplakias (OL) are clinically identifiable oral potentially malignant disorders (OPMDs) that progress from hyperplasia to mild, moderate, or severe dysplasia, eventually culminating in oral squamous cell carcinoma (OSCC). OL exhibits a 40-fold increased risk of progressing to cancer within one year of diagnosis. In contrast to OSCC, the genomic basis for the initiation of OL from normal tissue caused by risk factors such as tobacco chewing or smoking is much less studied. Therefore, we studied genome-wide mutational landscapes at hyperplastic or mild dysplastic stages in OL patients lacking signs of malignancy to understand the molecular events associated with cancer initiation. Methods: Exome sequencing and analyses were conducted on nine OL cases and paired peripheral blood monocytes to identify OL-specific somatic mutations and exon-level copy number alterations (CNAs). Various bioinformatics approaches were employed to determine their pathogenicity, mutation signatures, cancer driver mutations, and significantly enriched pathways. Results: The pair-wise exome comparisons revealed the affected genes carrying mutation signatures associated with deamination, tobacco smoking, homologous recombination deficiency, APOBEC, and base excision repair deficiency. Among various cancer-driver genes that were recurrently mutated, *FAT1* (44%), *MUC4* (33%), *CLTC1* (33%), and *NOTCH1* (33%) were more frequent. Clinically significant variants affecting *TP53*, *SMO*, *PRDM1*, *HRAS*, *BRCA1*, *ATM*, *RAF1*, and *PTEN* were observed. Pathogenic exon-level CNAs were identified in 40 cancer driver genes, including *CDKN2A*, *MSH2*, and *KRAS,* among others. Pathway enrichment and gene ontology analysis of the affected genes indicated the involvement of various cancer-associated pathways such as cell cycle, DNA repair, immune response, ECM receptor interactions, and focal adhesion. Pair-wise integrative analysis of exome, transcriptome, and methylome data will be performed to estimate the mutation and epigenetic (DNA methylation) burden in these affected tissues, and the data will be discussed. Conclusion: In summary, this first report on oral leukoplakias devoid of malignancy makes a beginning toward understanding the early molecular events associated with OSCC susceptibility. Financial support: Centre for Human Disease Research program, BITS Pilani Hyderabad Campus and Council of Scientific and Industrial Research (CSIR), Govt. of India.

Session Title: Cancer Poster Friday Session

Board 8018F: Comprehensive tumor mutational profiling and associated impact on overall survival outcomes in young women (diagnosed ≤ 40 years old) with breast cancer carrying a pathogenic germline *BRCA1* or *BRCA2* variant in the POSH cohort.

Authors:

M. Xia[1], M. Akamandisa[1], B. Wubbenhorst[1], W. Cheah[2], K. D'Andrea[1], M. Fan[1], J. Shilan[1], E. Copson[2], R. Cutress[2], W. Tapper[2], D. Eccles[2], S. Domchek[1], K. Nathanson[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Southampton, Southampton, United Kingdom

Abstract:

Background: Women with a germline *BRCA1* or *BRCA2* (*gBRCA*) pathogenic variant (PV) have an increased lifetime risk of developing breast cancer (BC). Previous research has shown no significant difference in overall survival (OS) between young female *gBRCA* PV carriers versus non-carriers diagnosed with BC ≤ 40 years old from the Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) cohort in the UK. Allele-specific loss of heterozygosity (asLOH), the loss of the wildtype allele at *BRCA* locus within the tumor, is associated with homologous recombination deficiency (HRD) in DNA repair. Previous studies unselected for age have reported asLOH in the breast tumor of *gBRCA* PV carriers at 90% in *BRCA1* and 54% in *BRCA2*. The rate of asLOH and its influence on OS in young female *gBRCA* PV carriers with BC remains unknown. **Objective:** We aimed to understand (1) the tumor mutational profile, including asLOH status at *BRCA* locus (LOH or nonLOH), and (2) the impact of asLOH status on OS in young female *gBRCA* PV carriers with BC. **Methods:** Whole-exome sequencing was performed on DNA from 138 tumors and matched normal tissue from the POSH cohort (87 [63%] with *gBRCA1* PV and 51 [37%] with *gBRCA2* PV). Tumor mutational profile, including pathogenic single nucleotide variants (SNV), copy number variants (CNV) in 22 genes, and asLOH status at *BRCA* locus, was curated using a custom computational pipeline. HRD scores and tumor mutational burden (TMB) were calculated. OS was evaluated by Kaplan-Meier curves. **Results:** LOH was observed in *gBRCA* PV carriers at high proportion (*BRCA1*: 92%, *BRCA2*: 94%). HRD scores were significantly higher in *gBRCA* PV carriers with LOH than those without (*BRCA1* mean: 57.4 vs. 25.2, $p < 0.0001$; *BRCA2* mean: 43.9 vs. 20.3, $p < 0.001$), whereas TMB did not differ by LOH status. In *gBRCA1* PV carriers, having nonLOH tumors was suggestive of higher OS compared to having LOH tumors (100% survival at 10 years vs. 80%), but the difference was not statistically significant. In *gBRCA1* PV carriers with estrogen receptor (ER) negative tumors, nonLOH also suggested better OS though not significantly (100% survival at 10

years vs. 77%), whereas no difference was observed in OS in the ER-positive subgroup. Our analysis also revealed that the top five most frequently altered genes in our 138 tumor samples were: *MYC* (66%), *AKT3* (51%), *TP53* (49%), *PIK3CA* (41%), and *FAT1* (28%). **Conclusion:** We did not find a significant impact of asLOH status at *BRCA* locus on OS in young female *gBRCA* PV carriers with BC. However, we observed a potential survival advantage in *gBRCA1* PV carriers with nonLOH breast tumors which necessitates further investigation in a larger population.

Session Title: Cancer Poster Friday Session

Board 8019F: Decoding *CDH1*: the hidden role of regulatory noncoding sequences in hereditary diffuse gastric cancer

Authors:

C. São José[1,2], M. Ferreira[1], L. Cordova[3], A. Pedro[1], A. André[1], J. Garcia-Pelaez[1], J. Senz[3], P. Kaurah[4,3], F. Puntieri[2], J. Glaser[2], D. Huntsman[3,5,6], S. Mundlos[2,7,8], K. Schrader[4,3], C. Oliveira[9,10]; [1]i3S: Inst. de Investigação e Inovação em Saúde, Porto, Portugal, [2]Max Planck Inst. for Molecular Genetics, Berlin, Germany, [3]Dept. of Pathology and Lab. Med., Univ. of British Columbia, Vancouver, BC, Canada, [4]BC Cancer Agency, Vancouver, BC, Canada, [5]Dept. of Molecular Oncology, British Columbia Res. Inst., Vancouver, BC, Canada, [6]Genetic Pathology Evaluation Ctr., Univ. of British Columbia and Vancouver Gen., Vancouver, BC, Canada, [7]Inst. of Med. and Human Genetics, Charité Univ.smedizin Berlin, Berlin, Germany, [8]Berlin-Brandenburg Ctr. for Regenerative Therapies (BCRT), Charité Univ.smedizin Berlin, Berlin, Germany, [9]i3S: Inst. de Investigação e Inovação em Saúde, Univ. of Porto, Porto, Portugal, [10]Faculty of Med. of the Univ. of Porto, Porto, Portugal

Abstract:

Background: *CDH1* or *CTNNA1* germline inactivation explains 10-40% of families with hereditary diffuse gastric cancer (HDGC). Missing heritability prevents proper management and disease prevention in HDGC-suspected families. We hypothesized that defects in *CDH1*-regulatory elements contribute to HDGC missing heritability.
Methods: We called single-nucleotide/copy-number variants (SNV/CNV) from 19 HDGC probands from whole-genome sequencing data, performed gene-ontology analysis, 4C-seq and ATAC-seq in normal stomach epithelia, CRISPR-Cas9, RT-PCR and flow cytometry in cell lines, RT-PCR, immunohistochemistry and microsatellite instability (MSI) analysis in tumours and enhancer reporter assays in mouse embryos.
Results: We found a *MLH1* 2.7Kb germline deletion in a family presenting DGC alone overlapping a stomach-specific regulatory element in 1/19 HDGC-suspected families. Gastric

tumour displayed *MLH1* mRNA reduced expression, MLH1 protein loss and MSI. CRISPR-Cas9 mimicking the family deletion triggered reduced expression of *MLH1* and *CDH1*/E-cadherin mRNA and protein. We crossed ATAC-seq (46.249 peaks) and 4C-seq (370 *CDH1* interactor-regions) data from normal stomach with WGS data (1.882 rare CNVs) from 19 HDGC-suspected families. A *CDH3* 20kb-deletion and a 39bp-intergenic deletion downstream of *CDH1* were mimicked with CRISPR-Cas9 (homozygous) triggering 100% or 50% *CDH1* mRNA/protein downregulation, respectively, similarly to a *CDH1* coding deletion. Two sequences within the *CDH3* CNV triggered 50% *CDH1* mRNA/protein downregulation and enhancer activity in mouse embryos. Combined analysis of deleted accessible chromatin regions revealed a pattern of impaired immune-related pathways in additional 9 families.
Conclusion: We identified regulatory regions within a *MLH1* stomach-specific regulatory sequence, and within the *CDH1* topologically-associated domain that harbour CNVs potentially explaining the missing heritability in three HDGC-suspected families. Germline CNVs in stomach-specific regulatory regions favouring an immune-deficient phenotype may predispose 9 additional families to HDGC.
Funding: FCT: PTDC/BTM-TEC/30164/2017, PTDC/BTM-TEC/6706/2020, SFRH/BD/140796/2018; EU: SOLVE-RD_Ref 22184, HORIZON_HORIZON-HLTH-2022-CARE-08.

Session Title: Cancer Poster Friday Session

Board 8020F: Defining actionable co-mutations and mutually exclusive genes in*NF1*-altered sporadic breast cancer

Authors:

M. Faraji Zonooz[1], P. S. Dischinger[1], J. L. Grit[1], C. R. Graveel[1], M. R. Steensma[1,2,3]; [1]Van Andel Res. Inst., Grand Rapids, MI, [2]Corwell Hlth.System, Grand Rapids, MI, [3]Coll. of Human Med., Michigan State Univ., Grand Rapids, MI

Abstract:

Background: Neurofibromatosis type 1 (*NF1*) gene alterations are increasingly recognized in various sporadic cancers, including all types of breast carcinomas, suggesting a role in endocrine resistance through the suppression of Ras and estrogen receptor signaling. The interaction between *NF1* somatic variants and other genetic alterations in aggressive breast cancer and drug resistance is, however, not well understood.
Objective: This study aims to define actionable co-mutated and mutually exclusive genes in *NF1*-altered sporadic breast cancer and assess their impact on cancer development, molecular subtypes, and patient outcomes.Methods: We utilized DNA sequencing data and clinical

information from the TCGA-BRCA and METABRIC databases, encompassing 3,593 breast cancer samples. A targeted gene panel was developed to identify and analyze genomic alterations' co-occurrence and mutual exclusivity with *NF1*.Results: In our cohorts, 38% of patients exhibited *NF1* alterations. Significant co-occurrence of *NF1* shallow deletions with *MYC* amplification and *TP53* deletions was observed. *MYC* was the most amplified gene in cases with *NF1* deletions, noted in 36% of METABRIC and 22% of TCGA-BRCA cases ($p < 0.01$). *TP53* was the most prevalent shallow deletion, found in 88% of TCGA-BRCA and 85% of METABRIC cases with *NF1* deletions ($p < 0.01$). Kaplan-Meier analysis indicated that patients with *NF1* and *TP53* shallow deletions and *MYC* amplifications had the poorest survival rates. Additionally, samples were categorized into *NF1*-carriers (with *NF1* SNVs) and *NF1*-non carriers (without *NF1* SNVs). Comparative analysis revealed significant disparities, notably the increased mutation frequency of the *ARID1B* gene in *NF1* carriers. This significant somatic interaction between *NF1* and *ARID1B* highlights its potential role in tumorigenesis in breast cancer.Conclusion: This study highlights the critical impact of *NF1* somatic variants and their association with *TP53*, *MYC* and *ARID1B* genomic alterations in sporadic breast cancer. The identification of co-mutated and mutually exclusive genes in *NF1*-altered cases provides new insights into the molecular mechanisms of tumorigenesis and potential targets for precision treatment strategies, promising to enhance patient care and improve clinical outcomes.

Session Title: Cancer Poster Friday Session

Board 8021F: Diverse Genetic Architecture of Neuroblastoma: Insights from a Multi-Ancestry Genome-Wide Association Study ⭐

Authors:

X. Chang, H. Qu, Y. Liu, J. Glessner, J. Maris, H. Hakonarson, S. Diskin; CHOP, Philadelphia, PA

Abstract:

Neuroblastoma, a malignancy of the sympathetic nervous system, has been linked to multiple genome-wide association loci in European cohorts that have been replicated in African American, Italian and Chinese populations. Here, we present a comprehensive multi-ancestry genome-wide association study of neuroblastoma, encompassing 6,179 cases and 28,973 controls from European, African American, Hispanic/Latin American, and East Asian populations. We identified 22 independent GWAS signals, including three novel loci specific to individuals of African ancestry. Subtype analyses revealed 17 novel loci, emphasizing the diverse genetic architecture of neuroblastoma susceptibility across different ancestral backgrounds and subtypes.

Among the significant findings, we identified a novel locus at 11p13.3 near the *TPCN2* gene, which encodes a two-pore channel involved in cancer progression. Additionally, a missense variant in the *EXO1* gene, involved in DNA repair was identified. Both *TPCN2* and *EXO1* have been implicated in other cancers, highlighting the relevance of these genes in oncogenesis. Another notable locus was found near the *MYCN* gene, frequently amplified in aggressive forms of neuroblastoma. These findings highlight the interplay between somatic alterations and genetic predisposition in neuroblastoma development.Cell-type-specific enrichment of SNP heritability using stratified linkage disequilibrium score regression revealed distinct patterns across neuroblastoma subtypes. Significant heritability enrichment was observed in musculoskeletal/connective tissue and digestive cell types for *MYCN*-amplified neuroblastoma, while intermediate-risk neuroblastoma showed enrichment in blood/immune categories. Additionally, we found significant heritability enrichment within kidney tissues across several chromatin marks, aligning with clinical observations. Polygenic risk score analysis demonstrated strong predictive performance in European populations, explaining up to 1.4% of the variance in neuroblastoma risk. However, the predictive accuracy in non-European populations was lower, likely due to smaller sample sizes. This comprehensive analysis provides valuable insights into the genetic underpinnings of neuroblastoma, informing future research and potential therapeutic strategies.

Session Title: Cancer Poster Friday Session

Board 8023F: Finding cancer misdiagnosis windows through electronic health records

Authors:

C. Krieger, E. Bang, B. Kim, M. H. Bailey; Brigham Young Univ., Provo, UT

Abstract:

Electronic Health Records (EHRs) chart the path of a patient's condition and can provide insight into patterns of disease onset, treatment strategies, and the rigors of an illness. The All of Us Research Program, recently launched in the United States, aims to broaden our understanding of disease phenotypes from diverse and understudied populations. To accomplish this objective, the All of Us Workbench plans to democratize health data from EHRs from 1 million individuals. Version 7 of the All of Us database contains 254,487 consented adult patients with EHRs. Preliminary analysis of V6 revealed that nearly 30% of participants had two or more instances of neoplasms, suggesting that the All of Us Research Program is a valuable resource for studying cancer phenotypes. Furthermore, the average record length per participant before a positive

cancer diagnosis is 4.5 years, with some patient's data being available 40 years before diagnosis. Early cancer detection can improve cancer outcomes. Many tools and algorithms have been developed to predict whether a patient will develop cancer, but few have attempted to generate predictions on how long someone has lived with a disease without diagnosis. Here, we are developing an algorithm that leverages EHRs from All of Us to identify patterns of cancer misdiagnosis to estimate the time someone has lived with a malignant neoplasm without a diagnosis. Our objective, while similar to predicting who will and will not develop cancer, may provide critical estimates to evaluate effective cancer diagnosis by region, socioeconomic status, and indicators of improved outcomes.

Our algorithm uses a dynamic programming approach, similar to genetic sequencing alignment, to identify the 'best' alignment between any two patient's EHRs. After analyzing pairwise comparisons of multiple 2000 patient sample sets, we observed high record alignment in EHRs up to 3 years apart between patients before a positive cancer diagnosis. These preliminary findings suggest that tracking cancer symptoms before a known diagnosis can lead to accurate estimates of misdiagnosis.

Our study demonstrates a capacity to capture patient-specific windows of cancer misdiagnosis spanning from 1-6 years before a patient is positively diagnosed. The event similarity between patients' records will capture patterns of cancer symptoms and promote earlier cancer diagnoses. Additionally, when successful, it will provide metrics to evaluate current medical practices and help to illuminate key indicators of diagnostic disparities across the United States.

Session Title: Cancer Poster Friday Session

Board 8024F: From desert flora to cancer therapy: systematic exploration of multi-pathway mechanisms using network pharmacology and molecular modeling approaches

Authors:

A. Alblihy; KFSC, Riyadh, Saudi Arabia

Abstract:

The absence of effective diagnostic methods, increased resistance to chemotherapy, and the heterogeneous nature of ovarian cancer collectively contribute to the unfavorable prognosis observed in the majority of cases. Thus, there is a pressing need to explore therapeutic interventions that offer superior efficacy and safety, thereby enhancing the survival prospects for ovarian cancer patients. our research synergizes bioinformatics with a network pharmacology approach to investigate the underlying molecular interactions of Saudi Arabian

flora (*Onopordum heteracanthum*, *Acacia ehrenbergiana*, *Osteospermum vaillantii*, *Cyperus rotundus*, *Carissa carandas*, *Carissa spinarum*, and *Camellia sinensis*) in ovarian cancer treatment. Phytoconstituents of indigenous flora and their associated gene targets, particularly those pertinent to ovarian cancer, were obtained from open-access databases. The shared targets of plants and diseases were compared to identify common targets. A protein-protein interaction (PPI) network of predicted targets was then constructed for the identification of key genes having the highest degree of connectivity among networks. A compound-target protein-pathway network was constructed, which uncovered that, namely, hispidulin, stigmasterol, ascorbic acid, octopamine, cyperene, kaempferol, pungenin, citric acid, d-tartaric acid, beta-sitosterol, (−)-epicatechin gallate, and (+)-catechin demonstrably influence cell proliferation and growth by impacting the AKT1 and VEGFA proteins. Molecular docking, complemented by a 20-ns molecular dynamic (MD) simulation, was used, and the binding affinity of the compound was further validated. Molecular docking, complemented by a 20-ns MD simulation, confirmed the binding affinity of these compounds. Specifically, for AKT1, ascorbic acid showed a docking score of −11.1227 kcal/mol, interacting with residues Ser A:240, Leu A:239, Arg A:243, Arg C:2, and Glu A:341. For VEGFA, hispidulin exhibited a docking score of −17.3714 kcal/mol, interacting with Asn A:158, Val A:190, Gln B:160, Ser A:179, and Ser B:176. To sum up, both a theoretical and empirical framework were established by this study, directing more comprehensive research and laying out a roadmap for the potential utilization of active compounds in the formulation of anti-cancer treatments.

Session Title: Cancer Poster Friday Session

Board 8025F: FusionRank: A machine learning framework for annotating and prioritizing oncogenic gene fusions

Authors:

K. Rybacki[1], M. Ahsan[2], K. Wang[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Gene fusions (GFs), arising from chromosomal rearrangements, play a critical role in the development and progression of various cancers. GFs can lead to dysregulated oncogenes and tumor suppressors by creating fusion transcripts and proteins with new functions, making them key biomarkers and potential therapeutic targets in cancer. Identifying and prioritizing oncogenic GFs are essential for advancing cancer diagnosis and treatment. Long-read sequencing techniques, such as those from PacBio and Oxford Nanopore, now enable highly

sensitive detection of GFs from long-read RNA-seq data. However, GFs identified through long-read sequencing face challenges in functional interpretation due to the abundance of reported GFs from typical bioinformatics algorithms. While some may be clinically relevant or driver GFs, many are technical artifacts or passenger GFs without functional significance. This underscores the importance to differentiate between them to create an effective prioritization strategy. To address this challenge, we developed FusionRank, an integrated computational pipeline that combines comprehensive GF annotation with machine learning techniques for prioritization based on the predicted oncogenicity of the GFs. We first annotated GFs based on 15 attributes encompassing sequence-based, functional annotations, pathway involvement, and clinical information. Then, we incorporated these features into a weighted logistic regression model to predict the pathogenicity of the GFs and rank them accordingly. The model was trained on a diverse set of GFs reported in fusion databases, comprising a total of 3,138 known pathogenic GFs from Mitelman, Cosmic Fusion, ChimerDB, and Archer Quiver, with over 4,540 non-pathogenic fusions reported in literature studies. We tested FusionRank on 27 in-house fusion positive and two negative clinical cancer patient samples, ensuring a wide-ranging representation of GFs for improved oncogenicity predictions. In summary, while most GF prioritization tools primarily utilize sequence-based features and functional annotations, FusionRank integrates diverse sources of information on cancer and tissue-specific annotations, pathway information, retained or lost protein domains, available therapies, clinical information, and an interpretable visualization component. Combining these annotations with machine learning aids in a more holistic understanding of GFs, facilitating oncogenicity prediction and subsequent prioritization.

Session Title: Cancer Poster Friday Session

Board 8026F: Gene-Environment Interaction: Small Deletions and Transcriptomic Profile in Non-Melanoma Skin Cancer (NMSC) in a Population Exposed to Arsenic

Authors:

M. Kibriya[1], F. Jasmine[1], M. Argos[2], Y. Khamkevych[1], T. Islam[3], M. Rakibuz-Zaman[4], M. Shahriar[1], W. Verret[5], C. Shea[6], H. Ahsan[6]; [1]Univ Chicago, Chicago, IL, [2]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [3]UChicago Res. Bangladesh, Dhaka, Bangladesh, [4]Pulse Infoframe, London, ON, Canada, [5]Gilead Sci., Foster City, CA, [6]Univ. of Chicago, Chicago, IL

Abstract:

Background: Exposure to inorganic arsenic (As) may be a risk factor for non-melanoma skin cancer (NMSC) in non-white population. In our previous study including 7000 adults exposed to As, we found that over six-year follow-up, 2.2% of the male and 1.3% of the female participants developed basal cell carcinoma (BCC), while 0.4% of the male and 0.2% of the female participants developed squamous cell carcinoma (SCC). We have reported the association of frequently found single nucleotide variation (SNV) in tumor tissue and gene expression. Here, we identify the associations of somatic small deletions (DEL) and transcriptomic profile in NMSC with high and low As-exposure. **Material & Methods:** Comparing the tumor tissues from the first 32 NMSC cases and corresponding blood from same patients, using targeted sequencing, we first identified SNV, DEL and insertions (INS) in tumors. Then we removed the variants identified in independent set of non-tumor apparently healthy skin (n=16) and paired blood sample. After filtering out other germline variants in public database, we identified NMSC-associated SNVs, DELs and INSs. **Results:** We detected DELs in 617 unique loci (with minimum quality score of Q60) in 236 genes including *MTR, MLH1, APC, ATM, MSH6*. Many of the KEGG gene pathways (*TNF* signaling pathway, *TGF* beta signaling pathway, *p53* signaling pathway, *IL17* signaling pathway, *PD-L1* expression pathway etc.) showed significant difference in the magnitude of differential expression in BCC tissue compared to non-lesional healthy skin depending on absence or presence of DEL and also by As-exposure status. For example, in low As-exposure cases, the presence of DEL in *APC* was associated with a fold change (FC, 95%CI) of 5.12 (4.2-6.2) in *TNF* signaling pathway compared to FC 3.91 (3.3-4.7) without *APC* DEL; in high As-exposure cases, the presence of *APC* DEL was associated with FC of 1.97 (1.7-2.3) compared to non-significant FC 1.13 (-1.01-1.30) without DEL (interaction $p<0.0001$). These molecular data suggested that targeted therapy like *TNF* inhibitors may show better response in NMSC cases with low As-exposure having APC DEL compared to other group. **Conclusion:** We report, for the first time, the significant association of small DELs in NMSC tissue and As-exposure with many dysregulated gene pathways. These findings may help selecting group of patients for potential targeted therapy like *PD-L1* inhibitors, *IL17* inhibitors, *TNF* inhibitors and *TGF* beta inhibitors for precision medicine in future.

Session Title: Cancer Poster Friday Session

Board 8028F: Germline genetic analysis of 43,248 individuals across nine less-investigated cancer types in *BRCA1* and *BRCA2*

Authors:

H. Sasagawa[1,2], M. Endo[1], Y. Iwasaki[1], Y. Usui[1], Y. Sugiyama[2], K. Numakura[2], K. Matsuo[3], J. Halder[4], M. T. Parsons[4], A. B. Spurdle[4], Y. Kamatani[5], K. Matsuda[5], Y. Murakami[6], T. Habuchi[2], Y.

Momozawa[1]; [1]RIKEN, Yokohama, Japan, [2]Akita Univ., Akita, Japan, [3]Aichi Cancer Ctr., Nagoya, Japan, [4]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [5]The Univ. of Tokyo, Tokyo, Japan, [6]Nippon Med. Sch., Tokyo, Japan

Abstract:

Introduction: The association of pathogenic variants in *BRCA1/2* with four cancer types, including breast cancer, is widely recognized. Our previous analysis revealed that pathogenic variants in *BRCA1/2* were also associated with the risk of biliary tract, gastric, and esophageal cancers (*JAMA Oncol 2022*). These results showed the potential for expanding indications for PARP inhibitors and surveillance. However, the clinical significance of other cancer types, which occupy about 20% of all cancers, remains unclear. Therefore, we evaluated nine less-investigated cancer types.

Methods: We analyzed all coding regions and two base pairs flanking intronic sequences of *BRCA1/2* in 4,406 patients with diagnosed or past medical history of nine cancer types (bladder, bone, brain, head and neck, sarcoma, skin, testis, thyroid, and ureter) and 38,842 controls in BioBank Japan and Akita University Hospital. Variants predicted as loss-of-function or registered as pathogenic and likely pathogenic in ClinVar were collectively designated as pathogenic variants. We evaluated the association of pathogenic variants in *BRCA1/2* with the risk of nine cancer types.

Results: We identified 114 pathogenic variants among 1,024 germline variants. Focusing on cancer types with more than one case carrying a pathogenic variant, carrier frequency with pathogenic variants was highest in *BRCA1* for thyroid cancer (0.7%) and *BRCA2* for skin cancer (1.5%). We observed four significant associations after Bonferroni-correction: *BRCA1* and thyroid cancer (OR, 4.82; $P = 9.32 \times 10^{-4}$); *BRCA2* and bladder cancer (OR, 3.07; $P = 3.47 \times 10^{-5}$), head and neck cancer (OR, 3.27; $P = 2.04 \times 10^{-4}$), and skin cancer (OR, 4.66; $P = 8.49 \times 10^{-4}$). For the subset of head and neck cancer, pathogenic variants in *BRCA2* were significantly associated with pharyngeal and laryngeal cancer (OR, 3.60; $P = 5.63 \times 10^{-4}$). Further analyses were performed based on known risk factors, including sex differences and smoking history. Analyses stratified by sex showed that the impact of pathogenic variants in *BRCA1* was larger for males in thyroid cancer (females: OR, 2.98; males: OR, 8.13) and that in *BRCA2* was larger for females in bladder cancer (females: OR, 14.28; males: OR, 1.93). There were no significant interactions between smoking history and pathogenic variants in *BRCA2* in relation to the risk of bladder and head and neck cancers for which smoking is a known risk factor.

Conclusions: Our analysis has identified associations between pathogenic variants in *BRCA1/2* and four less-investigated cancer types. These results suggested the potential for further expanding indications of PARP inhibitors and surveillance.

Session Title: Cancer Poster Friday Session

Board 8029F: Germline structural variants as risk alleles for high grade serous ovarian cancer

Authors:

N. Mazumdar[1], R. Scott[2], M. Diaz[1], N. Gull[1], P-C. Peng[1], J. Tyrer[3], J. Baierl[1], P-H. Lai[1], B. Rimel[1], A. Li[1], S. Gayther[4], P. Pharoah[1], G. Contino[2], M. Jones[1]; [1]Cedars Sinai Med. Ctr., Los Angeles, CA, [2]Univ. of Birmingham, Edgbaston, United Kingdom, [3]Univ. of Cambridge, Cambridge, United Kingdom, [4]Universtiy of Texas Hlth., San Antonio, TX

Abstract:

Epithelial ovarian cancer (EOC) is a lethal gynecologic malignancy. 80% of women suffering from EOC are diagnosed at stages III or IV with an aggressive histotype, high grade serous ovarian cancer (HGSOC), and have a 5-year survival rate of <30%. Expanding the compendium of genetic variants conferring risk for EOC will help identify women who are at risk and can benefit from prophylactic surgery and lifestyle interventions but are presently unidentifiable. Single nucleotide polymorphisms (SNPs) account for ~40% of EOC risk variants; ~30% of excess familial risk is attributed to high-moderately penetrant rare variants in genes of the homologous recombination repair pathway (BRCA1 or BRCA2; RAD51C, RAD51D, PALB2 or BRIP1) and <10% of heritability is attributed to low penetrant common variants across 30 loci identified from genome wide association studies (GWAS). There is much still to learn about the heritable factors for EOC; >90% of the known risk variants are non-coding, without known gene targets and the majority of risk variants are yet to be identified. We previously reported risk-associated copy number variants (CNVs) enriched at known EOC risk loci, identified through GWAS, and noncoding common CNVs enriched in regulatory elements using high-throughput genotyping array data across 13,000 cases and 17,000 controls. We aim to identify a larger catalog of structural variants (SVs) conferring risk for EOC and HGSOC using whole genome sequencing (WGS) data available through publicly available sources. In preliminary studies in the Genomics England workspace, we selected a set of 650 ovarian cancer cases and 2000 control participants matched for age, sex and ethnicity to perform a case-control association analysis. SVs and CNVs were called by running the Manta-Canvas pipelines on germline whole genomes. We identified 82 regions, spanning 10 kb in length, that demonstrated altered copy number states in ovarian cancer cases (Bonferroni corrected significance threshold < 1.3E-06). Forty-seven regions had CNVs in more than 10% of cases. More than half these regions demonstrate a loss in copy number, which are more widely spread across the genome than regions that demonstrate a gain in copy number, which are clustered on chromosomes 1, 14, 16, 17 and 21. Validation in UK Biobank, All of Us and newly generated internal datasets is underway. The results of this

analysis indicate germline regions associated with EOC and HGSOC heritability due to structural variation

Session Title: Cancer Poster Friday Session

Board 8030F: Identification of Genetic Network unveils the pathologic mechanism of UTUC:A multi-omics and longitudinalcohortstudy

Authors:

T. Fu[1,2,3], B. Liao[1], Y. Zhang[1,2], N. Xiang[1,2], L. Wang[1,2,3], S. SU[1,2], Y. Yang[4], G. Liao[5], L. Du[6], Y. Lin[1,7], J. Huang[1]; [1]Dept. of Urology, Innovation Inst. for Integration of Med. and Engineering, West China Hosp., Sichuan Univ., Chengdu, Sichuan, PR China, Chengdu, China, [2]Frontiers Sci. Ctr. for Disease-related Molecular Network, West China Hosp., Chengdu, Sichuan, PR China, Chengdu, China, [3]Key Lab. of Bio-Resource and Eco-Environment of Ministry of Ed.,Coll. of Life Sci., Sichuan Univ., Chengdu, Sichuan, PR China, Chengdu, China, [4]Hlth.Management Ctr., Gen. Practice Med. Ctr., Innovation Inst. for Integration of Med. and Engineering, West China Hosp., Sichuan Univ., Chengdu, Sichuan, PR China, Chengdu, China, [5]State Key Lab. of Oral Diseases, Natl. Clinical Res. Ctr. for Oral Diseases, West China Hosp. of Stomatology, Sichuan Univ., Chengdu, PR China, Chengdu, China, [6]Chinese Evidence-Based Med. Ctr., West China Hosp., Sichuan Univ., Chengdu, PR China, Chengdu, China, [7]Program in Genetic Epidemiology and Statistical Genetics, Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Background: Upper Tract Urothelial Carcinoma (UTUC) accounts for approximately 9.3%-29.9% of urothelial malignancies in China, a prevalence notably higher than that observed in European and American populations (5%-10%). Limited understanding of UTUC's molecular mechanisms hampers progress in clinical diagnostics and treatment strategies.**Patients and Methods:** This study initially cleaned and analyzed the clinical data of UTUC patients from West China Hospital of Sichuan University to assess the epidemiological characteristics of UTUC. Subsequently, 48 cases of UTUC along with adjacent normal tissues were selected for comprehensive exome and RNA sequencing. Utilizing machine learning, we constructed a genetic network to explore the pathological mechanisms of UTUC. Survival analysis was performed using cohort follow-up data to validate critical genetic markers.**Results:** Clinical data from 480 UTUC patients collected between January 1, 2019, and December 31, 2023, indicated a male predominance (57.5%) with an average age of 67.8 yr. The cohort exhibited significant histories of smoking (9.6%), alcohol consumption (5.4%), hypertension (40.6%), and diabetes (17.7%). Notably, 9.4% had a history of bladder cancer, and 31% presented with multifocal tumors. High-grade tumors were prevalent

in 81.7% of patients, with 62.2% displaying invasive characteristics at diagnosis. Concurrent hydronephrosis was observed in 49.0% of the cases. Through weighted gene co-expression network analysis (WGCNA), we identified four gene modules associated with tumor infiltration and highlighted nine hub genes (MSRB3, HSPB2, SCRG1, SYNPO2, FXYD1, PLXNA4, CD163, IFFO1, and FPR3) through integrative network analysis on Cytoscape/cytoHubba. Moreover, multivariate analysis unveiled 64 significant genes linked to tumor invasion (all FDR < 0.05), while comprehensive sequencing identified 3968 differentially expressed genes and 9276 somatic exon mutations. Enrichment analysis indicated that these differentially expressed genes are involved in the regulation of urogenital system development and calcium signaling pathways. By combining analysis, we found two potential genes, MSRB3 and SYNPO2. Crucially, survival analysis using both TCGA database and our cohort confirmed significant prognostic implications for two genes, MSRB3 and SYNPO2.**Conclusion:** Our findings underscore the potential of MSRB3 and SYNPO2 as biomarkers for elucidating the unique biological and clinical characteristics of UTUC, meriting further investigation in future prospective clinical trials.

Session Title: Cancer Poster Friday Session

Board 8031F: Immunoglobulin heavy chain variable region genes usage and mutational status in chronic lymphocytic leukemic patients in Nigeria ⭐

Authors:

I. Uzoma[1], T. A. Taiwo[2], M. A. Durosinmi[3], O. Akinloye[4]; [1]Molecular Hematology and Immunogenetic Lab., Dept. of Med. Lab. Sci., Faculty of Hlth.Sci. and Technology, Coll. of Med., Univ. of Nigeria Nsukka, Enugu, Nigeria, [2]Genetics Lab., Dept. of Cell Biology and Genetics, Faculty of Sci., Univ. of Lagos, Nigeria., Enugu, Nigeria, [3]Dept. of Haematology, Obafemi Awolowo Univ. Teaching Hosp. Complex, Ile-Ife, Nigeria, Enugu, Nigeria, [4]Clinical Chemistry and Molecular Diagnostics Lab., Dept. of Med. Lab. Sci., Faculty of Basic Med. Ctr., Coll. of Med., Univ. of Lagos, Nigeria., Enugu, Nigeria

Abstract:

Chronic lymphocytic leukemia (CLL) is an indolent B-cell lymphoid neoplasm characterized by the clonal expansion of CD5/CD23/CD19 and CD20-positive positive B-cells in blood, bone marrow and lymphoid tissues. It is commonly seen in the Western world compared to sub-Saharan African regions. The mutational status of the immunoglobulin heavy chain variable region (IGH) genes is a prognostic feature in CLL. The aims are to identify the most used alleles of the variable (V) diversity (D) and joining (J) gene segments of the IGH gene, to analyze

complementarity determining region 3 (CDR3) properties and the mutational status of the IGH in CLL. We analyzed IGH genes' mutational status by performing the sequencing of 8 Nigerian CLL patients out of 23 CLL cases obtained from three centers namely, University College Hospital, Ibadan, Oyo State, Enugu State University of Science and Technology Teaching Hospital (ESUTH), and Meena Histopathology and Cytology Laboratory, Jos, Plateau State. Diagnosis of CLL was confirmed according to the latest World Health Organisation (WHO) classification in our previous study. The samples were subjected to polymerase chain reaction (PCR) amplification of the IGH locus using the BIOMED-2 protocol. The research approvals were obtained from the University of Nigeria Teaching Hospital, Enugu, Ministry of Health, Ibadan, Oyo State, and Meena Histopathology and Cytology Laboratory, Jos, Plateau State. Of the 23 cases, 8 (30.4%) cases were amplifiable and showed common clonal origin using the FR1 primer mixes (BIOMED-2 protocol). The ages of the patients with the sequenced cases ranged from 50-78 years, with three females and four males. Most of the specimens were from nodal sites although two cases were from extranodal sites, namely the colon and the omentum. Clinical information on patients regarding the disease's staging, treatment, and overall survival were not available. The *IGH* analysis revealed the frequent usage of the IGHV4-34 allele (38%) for the V gene segment, IGHD6-13 (43%) for the D gene segment and IGHJ4-02 allele (43 %) for the J gene segment. The mutational analysis revealed that 62.5% of the CLL cases had mutated *IGH* and this occurred more within the older age bracket. The complementarity determining region (CDR3) properties showed a median length of 13 amino acids and longer CDR3 lengths were recorded in unmutated cases as opposed to mutated cases (P=0.002). The study demonstrated a higher prevalence of IGHV4-34 gene usage of the *IGH* in Nigerian CLL patients and more of mutated *IGH* than unmutated *IGH* which is indicative of favourable prognostic features in the studied Nigerian cases.

Session Title: Cancer Poster Friday Session

Board 8032F: Integrated multi-omics drive precision cancer care for pediatric CNS tumors: An institutional study of 582 tumors ⭐

Authors:

D. Wong[1], F. Lin[2], K. A. Cole[1], J. Wu[1], J. Schubert[3], J. Chen[1], C. Kline[1], N. Golenberg[1], M. Patel[1], K. Cao[1], W. Fu[4], E. H. Denenberg[1], E. Fanning[1], D. Gallo[1], K. Zelley[1], G. Wertheim[1], M. Luo[1], L. F. Surrey[1], Y. Zhong[1], A. Viaene[1], G. Brodeur[1], S. MacFarland[1], M. Santi[1], A. Resnick[5], P. B. Storm[1], M. Fisher[1], M. Li[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Gabriella Miller Kids First Data Resource Ctr., Philadelphia, PA

Abstract:

Introduction: Multi-omic information has become integral in providing a comprehensive and holistic framework for the diagnosis, prognosis, and therapeutics of central nervous system (CNS) tumors. Here, we characterize a single institution cohort of 582 CNS tumors from 540 pediatric patients (≤ 21 years old) using a clinically validated integrated multi-omics approach (targeted DNA/RNA fusion panel, RNA-seq, DNA methylation).
Results: Our cohort included 261 low grade glioma (LGG), 90 high grade glioma (HGG), 84 embryonal tumor (EMB), 55 ependymoma (EPN), and 50 rare entities grouped into "Other". Clinically significant variants were identified in 93.0% (502/540) of patients and 82.8% (447/540) of patients received clinical impact from their multi-omics profiling results. Clinical impact was determined if the findings provided diagnostic (419/540; 77.6%), prognostic (63/540; 11.7%; critical prognostic biomarkers for the tumor type), or therapeutic (198/540; 36.7%) information relevant to clinical management. In patients with a diagnostic finding, the results were supportive of the histological diagnosis in 93.1% (390/419) patients, provided further information for molecular subgrouping in 25.1% (105/419), or suggested a change in diagnosis in 4.3% (18/419) patients. Nearly all LGG are MAPK driven often harboring *BRAF* fusions 46.7% (122/261) or mutations in *BRAF* (22.9%, 60/261) or *NF1* (9.9%, 26/261). HGG are more heterogeneous with frequent epigenetic (*H3-3A*) and DNA damage repair defects (*TP53*) or *MET* fusions. Copy number variants were the most common alteration type providing diagnostic information which is especially important for subtyping EMB and EPN. Though RNA-seq and methylation profiling were not performed in all cases, they play significant roles in diagnosing rare entities and tumors with non-canonical genomic alterations and equivocal histology ("Other"). In addition, 23.1% (125/540) of patients were suspected to have germline pathogenic variants with 58/540 (10.7%) being confirmed (45/125 germline test not done yet), facilitating cascade testing and enrollment into surveillance programs. The most frequent genes with germline alterations were *TP53* (HGG), *NF1* (LGG), and *SMARCB1* (atypical teratoid/rhabdoid tumor).
Conclusions: These data highlight the importance of multi-omics information as a component of precision medicine for the integrated diagnosis, prognosis, and therapeutic management of pediatric CNS tumor patients. The high rates of germline findings also emphasize the importance of identifying cancer predisposition within this population which may affect treatment and tumor surveillance.

Session Title: Cancer Poster Friday Session

Board 8033F: Integrative proteogenomic analyses reveal insights into subtype-specific glioma risk

Authors:

T. Nakase[1], K. Smith-Byrne[2], C. L. Cario[1], Q. T. Ostrom[3], G. A. Guerra[4], B. S. Melin[5], M. Rentoft[5], M. Wrensch[4], J. K. Wiencke[4], R. B. Jenkins[6], J. E. Eckel-Passow[6], Glioma International Case-Control Study (GICC), M. L. Bondy[1], S. S. Francis[4], L. Kachuri[1]; [1]Stanford Univ., Stanford, CA, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Duke Univ., Durham, NC, [4]Univ. of California San Francisco, San Francisco, CA, [5]Umea Univ., Umea, Sweden, [6]Mayo Clinic, Rochester, MN

Abstract:

Gliomas are a heterogeneous group of brain tumors with few known risk factors. Tumors with *IDH* mutations and 1p19q co-deletion have a relatively favorable prognosis, while IDH-wildtype glioblastomas (IDH_{wt} GBM) have a dismal median survival of 14 months. We undertook an integrative analysis of the plasma and brain proteome using protein quantitative trait loci (pQTL) to decipher disease mechanisms and explore potential therapeutic targets.
Our plasma proteome-wide association study (PWAS) examined 1,362 proteins in 11,304 glioma cases and 304,523 controls and identified 29 candidate proteins (FDR<0.1), including 9 with high posterior probability (PP) of colocalization (SuSiE-coloc PP>0.70). Most new risk signals were subtype-specific, such as CRELD1 ($p=2.5\times10^{-5}$) for triple-negative GBM, which has been shown to modulate chemosensitivity in GBM CRISPR screens. We also found three risk proteins for IDH-mutant 1p19q-intact ($IDH_{mut\text{-}intact}$) glioma: immunoglobulin glycoprotein BCAM ($p=9.6\times10^{-8}$), neurotrophin receptor SorCS2 ($p=3.6\times10^{-5}$) and immune checkpoint PD-1 ($p=1.1\times10^{-5}$). Interestingly, PD-1 inhibitors have not demonstrated clear benefits in GBM patients. Our results suggest this may be related to the larger effect of PD-1 on $IDH_{mut\text{-}intact}$ (OR=2.75, 95% CI: 1.61-4.69) than IDH_{wt} (OR=0.73, 95% CI: 0.51-1.05) tumors. For all colocalized candidates, plasma protein abundance was cis-regulated by transcriptional activity in brain tissue. For instance, genetically-predicted *SORCS2* expression in brain tissue was associated with SorCS2 levels in plasma ($z=16.1$, $p=1.1\times10^{-68}$), consistent with its role as a neurotrophin receptor in the brain. Analyses of brain pQTL data for 1,776 proteins identified 9 associations (PP>0.70), including both known and new risk signals for GBM, such as EGFR (OR=0.63 $p=1.9\times10^{-35}$) and ENPP6 (OR=0.62, $p=7.6\times10^{-5}$), respectively. Novel findings for glioma overall included galectin-3 (OR=1.37, $p=1.3\times10^{-5}$), a validated drug target and tumor fitness gene confirmed by CRISPR screens of glioma proliferation.
Taken together, our study identified multiple candidate risk proteins for glioma involved in immune response and neuronal proliferation. Triangulation with experimental data provided further support for these pQTL-based associations, which may help facilitate the prioritization of targets for future mechanistic and therapeutic investigations.

Session Title: Cancer Poster Friday Session

Board 8034F: Investigating the Association Between Clonal Hematopoiesis and Lung Cancer Risk

Authors:

J. Ulloa[1], M. Betti[1], T. Mack[1], Y. Pershad[1], A. Manful[1], A. Bick[2], M. Aldrich[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Introduction: Clonal hematopoiesis is an age-related condition characterized by the expansion of blood cells. This is driven by somatic mutations including single nucleotide mutations and mosaic chromosomal alterations (mCA), which range from a few kilobases to entire chromosomes in length. mCAs appear to be associated with increased lung cancer risk, however this has only been demonstrated in a limited number of studies. We seek to not only replicate prior associations between mCAs and lung cancer, but to also explore potential racial differences. Methods: Using a nested case-control design, we identified 1,016 controls (matched on age, sex, self-identified race, and recruitment site) and 606 incident lung cancer cases from the prospective observational Southern Community Cohort Study. DNA was obtained from blood and participants were genotyped using the Illumina MEGA genotyping array. mCA mutations were called using the established MOsaic CHromosomal Alterations (MoChA) analytic pipeline. Smoking behaviors were collected using surveys. Two-way global genetic similarity was estimated using ADMIXTURE software with CEU and YRI 1000 Genomes reference populations. Associations between mCAs and lung cancer were evaluated using logistic regression, controlling for age, sex, race, global genetic similarity, smoking status, and smoking pack-years. Results: Among 1,622 adults, 47% were male and 48% self-identified as Black/African American with a mean age of 55 years. Overall, 4% of individuals had at least one autosomal mCA. The frequency of mCA mutations was higher among lung cancer cases than controls (5.9% vs. 2.9% respectively). Analyses found that individuals with at least one mCA had a nearly 90% greater odds of developing lung cancer (odds ratio (OR) (95%CI): 1.89 (1.09-3.31), p = 0.023) compared to those without a mCA mutation. When stratified by self-identified race, there was no evidence of an association between mCA and lung cancer among White individuals (OR (95%CI): 1.25 (0.60-2.63), p = 0.54). Among Black/African American individuals, those with at least one mCA had a 3-fold greater odds of developing lung cancer compared to those without an mCA (OR (95%CI): 3.29 (1.43-7.97), p = 5.91x10-3). An assessment of interaction between self-identified race and mCA was not statistically significant (p = 0.10). Conclusion: Our study confirms prior findings of an increased risk of lung cancer associated with the presence of mCA mutations and

highlight opportunities to explore potential contributions of social determinants of health, environment and genetic similarity to this association.

Session Title: Cancer Poster Friday Session

Board 8035F: Investigating the cell-type specific regulation of IRF4 and its role in lung cancer via a lung cancer risk-associated pleiotropic variant

Authors:

J. Yin[1], A. Kane[1], E. Long[2], J. Feng[1], C. Gräwe[3], B. Li[1], H. Patel[1], J. Byun[4], C. Amos[4], K. Brown[1], J. Xia[5], M. Vermeulen[3], J. Choi[1]; [1]Natl. Cancer Inst./NIH, Rockville, MD, [2]Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, [3]Radboud Inst. for Molecular Life Sci., Oncode Inst., Radboud Univ., Nijmegen, Netherlands, [4]Baylor Coll. of Med., Houston, TX, [5]Dept. of BioMed. Sci., Creighton Univ., Omaha, NE

Abstract:

A recent cross-ancestry lung cancer GWAS have identified a locus associated with lung cancer risk at 6p25.3 tagged by a common germline variant, rs12203592 (C/T). Subsequent quantitative trait loci (eQTL) colocalization analysis using lung and blood tissues (GTEx v8) identified IRF4 as a likely target of rs12203592. rs12203592 has previously been linked with diverse traits, including pigmentation phenotypes and skin cancers, blood and immune traits, and blood cancers. rs12203592 also appears to have varied allelic effects in different cell types. Namely, the lung cancer risk-associated T allele is correlated with higher IRF4 expression in lung tissue and blood cells (GTEx v8) but lower expression in primary melanocytes (n = 106; Zhang et al). Previous studies had shown the melanocyte-specific allelic function of this SNP through lineage-specific transcription factors (TFs), suggesting a mechanism in this cell type relevant to pigmentation phenotype. We hypothesized that context specific regulation of IRF4 in cell types of lung tissue microenvironment contributes to the lung tumorigenesis through rs12203592. To identify potential cell-type-specific regulators, we performed mass spectrometry using the nuclear extracts from three cell lines representing lung, blood, and melanocyte lineage. The results nominated the NF-kB protein families as the most prominent T-specific binders in the A549 lung cancer cell line, which was validated by electromobility shift assays. Consistent with these findings, tissue-based SNP-gene interaction analysis demonstrated that the mRNA levels of NF-kB family members (3 of 5 genes) were correlated with IRF4 expression with a significant interaction with rs12203592 in GTEx lung tissue ($P < 0.007$), where the correlation is mainly observed with the T allele. In A549 cells, NF-kB pathways were induced by TNFα treatment,

accompanied by a >5-fold increase of IRF4 expression. To investigate the potential target genes of IRF4 in immune cells that could influence lung tumorigenesis, we analyzed 6 RNAseq datasets of IRF4 knockdown or over-expression. Among 314 genes significantly altered in >=2 datasets, 88 were predicted to be secreted, and 20 of them were further supported by IRF4 and RELA co-binding in the promoters based on a chromatin immunoprecipitation sequencing data. We will further investigate the effect of NF-kB pathway induction on the allelic IRF4 expression in lung epithelial and immune cell lines. These data will help elucidate the context-specific roles of IRF4 in lung cancer development and provide clues to gene regulation in this pleiotropic GWAS locus.

Session Title: Cancer Poster Friday Session

Board 8037F: Isoform-specific transcriptome-wide association studies of 6 cancers uncover extensive risk mechanisms undetectable by gene-level methods

Authors:

Y-H. Chang[1], B. Pasaniuc[2], S. Lindstroem[3], A. Bhattacharya[4]; [1]MD Anderson UTHlth.Houston Graduate Sch., Houston, TX, [2]UCLA, Los Angeles, CA, [3]Univ. of Washington, Seattle, WA, [4]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract:

Transcriptome-wide association studies (TWAS) have identified potential mediators for genetic risk of multiple cancers by integrating GWAS and tissue-specific reference panels of total gene expression. However, traditional TWAS relies on total gene expression, overlooking alternative splicing that can produce multiple isoforms from the same gene, each with potentially different effects on cancers.

Here, we identify transcriptomic mechanisms underlying genetic associations with risk of 6 cancers (breast, endometrial, colorectal, lung, ovarian, prostate) that gene-level methods do not detect. We applied isoTWAS, a multivariate, stepwise framework that integrates genetic and isoform-level transcriptomic variation with publicly available European-ancestry GWAS summary statistics (all N > 20,000) to identify both isoform- and gene-trait associations. Multi-tissue isoform- and gene-level expression was quantified from the GTEx Project (N = 838) using Salmon and GENCODE v38 transcriptomic annotations. We developed predictive models for (1) isoform expression using multivariate methods from isoTWAS that model shared genetic architectures for isoforms of the same genes and (2) gene expression using univariate methods from gene-level TWAS.

isoTWAS substantially increases discovery beyond TWAS. First, isoTWAS identified 2714 more

unique genes than TWAS, with most associations in breast (3575 isoforms, 1845 unique genes) and prostate (1755 isoforms, 1053 genes) cancers. Second, across the 6 cancers, we estimated a 34.8% (jackknife SE = 1.61%) increase in effective sample size for isoTWAS versus TWAS, calculated by converting gene-level association P values into χ^2 statistics. Third, isoTWAS prioritized 942 constrained genes (pLI ≥ 0.9) compared to 199 by TWAS, suggesting that isoTWAS captured true disease signals. Fourth, both TWAS and isoTWAS commonly tagged 112 GWAS-significant loci with transcriptomic associations, but isoTWAS identified 64 additional isoform-associated loci missed by TWAS, while TWAS found only 3 additional loci, marking a 56.6% increase in transcriptome-mediated GWAS loci discovery with isoTWAS. Fifth, Mediated Expression Score Regression estimated that isoform eQTLs (iso-eQTL, 20.6 ± 7.8%) explain nearly 63% more of SNP heritability than gene eQTLs (g-eQTL, 11.6 ± 4.7%). Lastly, we highlight several salient isoTWAS associations that exhibit iso-eQTL colocalization but no g-eQTL, including CLPTM1L (lung cancer), LAMC1 (colorectal), and LSP1 (breast).
In sum, our results strongly demonstrate that modeling isoform expression is crucial to maximize discovery of genetic risk mechanisms for cancers.

Session Title: Cancer Poster Friday Session

Board 8038F: Longitudinal characterization of 1,763 mosaic chromosomal alterations identifies factors influencing clonal dynamics of leukocytes ⭐

Authors:

R. Kelly[1,2], D. W. Brown[1,2], W. Zhou[3,1], A. K. Hubbard[1], C. L. Young[1], A. Klein[1], W-Y. Huang[1], N. D. Freedman[4], S. J. Chanock[1], P. S. Albert[1], M. J. Machiela[1]; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, [2]Cancer Prevention Fellowship Program, Natl. Cancer Inst., Rockville, MD, [3]Leidos BioMed. Res./NCI, Bethesda, MD, [4]Div. of Cancer Control and Population Sci., Natl. Cancer Inst., Rockville, MD

Abstract:

Circulating leukocytes with large mosaic chromosomal alterations (mCAs) are detectable in the general population, increase in frequency with age and are strongly associated with incident lymphoid and myeloid malignancies. Detectable mCAs are a type of clonal hematopoiesis in which clones with structural alterations change in relative abundance due to selective pressures. Understanding the epidemiologic factors and alterations influencing clonal trajectory could provide novel etiologic insights on the progression of normal leukocytes into malignant clones. We investigated mCAs in blood-derived DNA from 56,324 participants in the Prostate,

Lung, Colorectal, Ovarian Cancer Screening Trial, a prospective cohort of individuals between 55 and 74 years of age. mCAs were detected utilizing genotyping array intensity data to produce virtual karyotypes. In total 2,186 (4.6%) individuals had autosomal mCAs, 3,559 (12.7%) females had mosaic loss of the X chromosome (mLOX) and 5,355 (18.9%) males had mosaic loss of the Y chromosome (mLOY). Autosomal mCAs were associated with age (P value=2.0e-41), smoking (P=8.0e-4), and male sex (P=0.049). For a subset of 1,498 participants with detectable mCAs and samples collected over a 6-year time span, we performed array genotyping on an additional 3,880 blood samples to track the clonal trajectories of 1,763 mCAs. The effect of sex, smoking status, and mCA characteristics on clonal trajectory across age was examined using linear mixed models. Overall, mLOY had the highest rate of expansion (1.9% increase per year increase in age), followed by autosomal mCAs and mLOX (0.7% and 0.3%). For autosomal mCAs, chromosome 20 had the highest rate (1.5%) followed by chromosomes 9 and 3 (both at 1.3%). Rates also differed by mCA type, with autosomal gains and losses having higher rates (0.9 and 1.0%), relative to copy-neutral loss of heterozygosity (0.4%). We observed limited evidence for an effect of individual factors on rate of expansion of autosomal mCAs. Exploratory analyses suggest that individuals with an incident hematologic malignancy had elevated rates of clonal expansion relative to cancer-free individuals (1.1% vs. 0.7%). We observed wide variation in mCA clonal trajectories with differences primarily attributable to chromosomal location and copy number state. Additionally, though individual epidemiologic factors were associated with risk of autosomal mCAs, they had little impact on rates of mCA expansion. These results suggest that intrinsic factors related to gene abundance are important contributors of mCA clonal expansion, with some high-risk chromosomal alterations leading to large increases in clonal fraction.

Session Title: Cancer Poster Friday Session

Board 8039F: Long-range somatic structural variation calling from matched tumor-normal co-assembly graphs

Authors:

M. K. Le[1], Q. Qin[2], H. Li[3,4]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Dana-Farber Cancer Inst., Boston, MA, [4]Harvard Med. Sch., Boston, MA

Abstract:

The accurate identification of somatic structural variants (SVs) is a computational biology problem with significant applications to clinical cancer research. Though technologies such as

long read sequencing have facilitated the development of more accurate SV calling methods, existing somatic SV callers still struggle with achieving high precision. In this work, we present colorSV, a long read-based method for calling long-range SVs (those with a size of at least 1 Mb) by examining the local topology of joint assembly graphs from matched tumor-normal samples. colorSV is the first somatic SV calling method that uses a co-assembly approach, as well as the first SV caller that identifies variants by examining assembly graph topology. We demonstrate near-perfect precision and sensitivity for calling translocations on the COLO829 cell line, outperforming four existing somatic SV callers (Severus, Sniffles2, nanomonsv, and SAVANA) in both metrics and achieving an up to 0.67 increase in sensitivity over previous methods. We also evaluated colorSV for calling translocations on the HCC1395 cell line, finding that our method achieved a good balance between sensitivity and precision (where the sensitivity was only outperformed by Severus, and the precision was only outperformed by nanomonsv). Our work establishes a novel joint assembly-based strategy for characterizing long-range somatic variation, which could be further expanded or modified for the identification of SVs of different types and sizes.

Session Title: Cancer Poster Friday Session

Board 8040F: Loss of Y chromosome as the only cytogenetic abnormality in myelodysplastic syndromes

Authors:

N. Abdelmoula; LR23ES07 Med. Univ. of Sfax, Sfax, Tunisia

Abstract:

Background: Recent studies offer valuable insights into the fundamental biology of Y chromosome loss, contributing to advancements in cancer immunotherapy. Although the occurrence of Y chromosome loss is recognized as common in the somatic cells of elderly men, recent findings indicate that the Y chromosome is stepping beyond its role in sex determination. Notably, Y chromosome loss in peripheral blood has been linked to an elevated risk of all-cause mortality and various diseases, including different forms of cancers. Our aim was to report through a Tunisian cohort, the malignant myeloid disorders cases with loss of the Y chromosome as the only cytogenetic abnormality. We aim to discuss the genetic basis and the innovative importance of the screening of this chromosomal abnormality in hematological malignancies. Methods: Blood and marrow samples were cytogenetically studied from a cohort of 39 Tunisian patients presenting with CML (n=27) and MDS (n=12). Results: The occurrence of Y chromosome loss as the only cytogenetic abnormality in tested Tunisian patients was

detected in only one patient (8,3 percent of MDS patients). The patient was a 75-year-old man with refractory anemia (RA) and clinical MDS in progress for 20 months. The clonal population was homogenous: 45,X,-Y. Conclusion: Given the limited therapy options for sustainable responses in the frontline setting for patients with MDS, this work provides insights into the significance of Y chromosome loss, because this abnormality seems to be more substantial predictor of cancer presence than age. Extensive documentation reveals that Y chromosome loss is identified in 4-10 percent of male patients, often as the sole cytogenetic abnormality in MDS. Equally, Y chromosome loss is noted in various cancer types, encompassing 10-40 percent of bladder cancers. In tumor tissue, this cytogenetic abnormality experiences a significant increase, aligning with a generally unfavorable prognosis. Within hematologic malignancies, mosaic loss of chromosome Y in blood cells stands out as one of the most frequent chromosome alterations in adult males. This loss is strongly associated with clonal hematopoiesis and hematopoietic malignancies. While the molecular significance of Y chromosome loss remains unidentified, it is suggested that aberrant expression of Y chromosome genes might elucidate some mechanisms responsible for its contribution to male cancers. Our work highlights the potential of exploring the Y chromosome using FISH or molecular biology to improve the precision of detecting such abnormalities during hematopoietic malignancies, which can be managed through future immunotherapy approaches.

Session Title: Cancer Poster Friday Session

Board 8041F: Lynch Syndrome Pathogenic Variant Prevalence in a Population Biobank: an Analysis of Results from the Colorado Center for Personalized Medicine Biobank

Authors:

L. Axell; Univ. of Colorado, Aurora, CO

Abstract:

Introduction: Lynch syndrome (LS) due to pathogenic variants (PV) in MLH1, MSH2, MSH6, PMS2 or EPCAM is the most common cause of inherited predisposition to colorectal cancer, endometrial cancer, and other LS related cancers. Most descriptions of LS incidence and cancer risks are in registries where the proband was identified based on their cancer diagnosis. We describe the incidence of LS related variants in our population-based biobank using research data from Whole Exome Sequencing and SNP technology. We describe how our registry-based LS PV incidence compares to a calculated estimate from 2017 of 1 in 279 of the population

(MLH1 1/1,946, MSH2 1/2,841, MSH6 1/758, PMS2 1/714). Methods: To date 94,826 samples have been eligible for inclusion into the biobank database. 72,428 samples come from Regeneron GC Whole Exome Sequencing (WES), and 33,864 samples come from MEGA (Illumina Multi-Ethnic Genotyping SNP Array). Clinical variants within the 4 LS-related genes on the ACMG 3.2 secondary findings list (MLH1, MSH2, MSH6, PMS2) were considered if they were pathogenic or likely pathogenic (PV/ LPV) from an expert panel or multiple submitters from Clinvar and had a minor allele frequency <0.01. Rare or private variants would not be identified by our process. EPCAM is not one of the genes listed in the ACMG therefore not included. Results: In our biobank a total of 225 LS PV/LPV were identified. 33 MLH1(14.7%), 22 MSH2 (9.8%), 85 MSH6 (37.8%) and 85 PMS2 (37.8%). MLH1 1/2873, MSH2 1/4310, MSH6 1/1115 and PMS2 1/1115. Conclusions: In existing LS registries, MLH1 and MSH2 PV are most common, with fewer MSH6 PV and PMS2 PV (27-41% MLH1, 39-48% MSH2, 13-26% MSH6, 3.8-6.4% PMS2). In our biobank, we identified, most PV in MSH6 (37.8%) and PMS2 (37.8%) and fewer in MLH1(14.7%) and MSH2(9.8%). This finding supports the higher population prevalence of MSH6 and PMS2 calculated in 2017. Based on most cases of CMMRD being due to PV in MSH6 or PMS2, it was also expected that MSH6 and PMS2 be common in the general population as our registry data supports. PMS2 and MSH6 PV are more common in the general population than MLH1 and MSH2. Identifying all individuals with LS is important for their health management as well as being able to describe the varied cancer risks more accurately across the Lynch syndrome genes. The incidence of PV found in our population-based registry is lower than the published calculated incidence (1 in 421 vs. 1 in 279) Determining the incidence of PV in the general population across the LS genes will be more accurately determined with the continued testing through general population biobanks. Defining the incidence of LS across different populations is an important next step.

Session Title: Cancer Poster Friday Session

Board 8042F: Machine learning approaches to understand regulatory mechanisms of methylation dysregulation

Authors:

J. Tye, K. Clement; Univ. of Utah, Salt Lake City, UT

Abstract:

DNA methylation is an important epigenetic mark associated with gene regulation. Changes in methylation are common across many different cancer types, sometimes producing similar methylation signatures in different cancers. However, the drivers of these changes are not

completely understood. We studied methylation in a pan-cancer study to determine the relationship between methylation dysregulation, gene expression, and mutations. We used a collection of 9714 samples from 38 studies from The Cancer Genome Atlas (TGCA). These patient samples included methylome data from Infinium HM450K arrays as well as transcription and genetic mutation data. We quantified methylation dysregulation in samples at partially methylated regions (PMRs), which represent sites of methylation dysregulation. We selected regions with partial methylation in 30-60% of our cancer samples. We filtered the PMR set to those regions that were dynamic in a subset of samples and called these dynamic PMRs (dPMRs). UMAP clustering analysis on average methylation levels in dPMRs revealed a common methylation dysregulation signature shared across some types of cancers, such as thyroid and lung cancers, while other cancers, such as breast cancer, had a unique methylation dysregulation signature. Next, we trained machine-learning models on this dataset to determine the relationship between gene expression, mutations, and average methylation levels in each dPMR. We verified these models with additional studies to ensure model robustness. Finally, we used SHAP values to discover genes whose expression and mutation state contribute to methylation dysregulation. These genes included known methylation regulators and genes whose function as epigenetic modifiers has not been reported. Further investigation of these genes and the mutations from our models will yield an improved understanding of methylation dysregulation and its role in genome regulation and disease.

Session Title: Cancer Poster Friday Session

Board 8043F: MicroRNA signature predicts prognosis in patients with acute myeloid leukemia

Authors:

S. Yerukala Sathipati[1], T. Carter[1], R. Sharma[2]; [1]Marshfield Clinic Res. Inst., Marshfield, WI, [2]Marshfield Clinic Hlth.System, Marshfield, WI

Abstract:

Acute myeloid leukemia (AML) is a hematologic malignancy characterized by significant molecular heterogeneity and the abnormal proliferation of myeloid progenitor cells. While microRNAs (miRNAs) play a crucial role in the diagnosis, pathogenesis, and prognosis of AML, there remains a lack of accurate prognostic indicators, especially for pediatric and adolescent patients. Identifying miRNA signatures could enhance the prediction of overall survival and aid in developing targeted treatments, aligning with the goals of precision medicine research. In this study, we developed an evolutionary learning-based AML survival estimator (AMLS) to

predict the overall survival time of AML patients. AMLS integrates machine learning methods with an optimal feature selection algorithm, the inheritable bi-objective combinatorial genetic algorithm. We analyzed miRNA expression profiles from 156 AML patients, sourced from The Cancer Genome Atlas database. Through rigorous data engineering, AMLS identified 28 miRNAs as biomarkers for overall survival, achieving a correlation coefficient of 0.85 and a mean absolute error of 0.62 years between actual and estimated survival times. AMLS outperformed standard machine learning methods. The top 10 ranked miRNAs identified by AMLS include hsa-miR-99a, hsa-miR-942, hsa-miR-1249, hsa-miR-21, hsa-miR-30c, hsa-miR-183, hsa-miR-320c, hsa-miR-29c, hsa-miR-624, and hsa-miR-1978. These miRNAs provide key insights into AML disease progression by targeting several AML-associated genes and participating in critical biological pathways such as the p53 signaling pathway, chronic myeloid leukemia, and acute myeloid leukemia pathways, which impact tumor progression and survival. In conclusion, the identified miRNA signature can estimate prognosis in AML patients and has the potential to serve as a biomarker for AML prognosis, thereby improving therapeutic interventions.

Session Title: Cancer Poster Friday Session

Board 8044F: Multi-ancestry transcriptome-wide association analyses on single-cell deconvoluted expression data to identify cell-type specific risk genes for colorectal cancer ⭐

Authors:

Q. Li[1], J. Wang[1], V. Moreno[2], C. Li[1], J. Choi[3], Z. Chen[1], W. Wen[1], X. Shu[4], X-o. Shu[1], J. Long[1], Q. Cai[1], B. Singh[1], K. S. Lau[1], U. Peters[5], W. Zheng[1], X. Guo[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]L'Hosp.et de Llobregat, Barcelona, Spain, [3]Korea Univ. Coll. of Med., Ansan, Korea, Republic of, [4]Mem. Sloan Kettering Cancer Ctr., New York, NY, [5]Fred Hutchinson Cancer Ctr., Seattle, WA

Abstract:

Background: Transcriptome-wide association studies (TWAS) have identified approximately 300 potential risk genes for colorectal cancer (CRC), yet many target genes at genome-wide association study (GWAS) risk loci remain unknown. Traditional TWAS use gene expression predictions based on bulk normal colon tissues, potentially compromising performance due to tissue heterogeneity. Integrating single-cell RNA sequencing (scRNA-seq) data to deconvolute bulk RNA-seq datasets allows for quantifying gene expression profiles in distinct cell types, enhancing gene expression predictions. However, TWAS analysis using deconvoluted cell-type expression data in CRC remains largely unexplored. **Materials and methods:** We utilized CIBERSORTx tool (PMID: 31061481) to deconvolute bulk colon tissue RNA-seq datasets from 707

European-ancestry individuals (423 from BarcUVa-Seq project and 284 from GTEx), and 364 Asian-ancestry individuals (ACCC), using referenced scRNA-seq data from the COLON MAP (PMID: 34518541). Using the deconvoluted expression data, we used our sTF-TWAS (PMID: 36402776) to build genetic prediction models for potential CRC cell origins in major cell types, including absorptive cells (ABS) and stem cells (STM). TWAS analyses were conducted based on the population-specific prediction models and GWAS summary statistics from European (78,473 CRC cases and 107,143 controls) and Asian populations (23,942 cases and 217,267 controls). Putative risk genes were identified through a meta-analysis of TWAS results using the ACAT approach. Using scRNA-seq data from normal colon tissues, differential gene expression analyses between the target cell type (e.g., ABS) and other cell types were conducted. Pathway enrichment analysis was conducted via Ingenuity Pathway Analysis (IPA). **Results:** We identified 132 putative CRC susceptibility genes, including 102 from the prediction models built in ABS and 57 from the models in STM, at a Bonferroni-corrected $P < 0.05$. Of these, 35 genes were not previously implicated in CRC susceptibility, including 6 genes at 6 loci previously not reported by CRC GWAS (>1Mb away). Total 60 of the identified risk genes (45%) showed significant differential expression in the target cell type compared to other cell types. IPA analysis indicates significant enrichment of the identified risk genes in CRC-related pathways, notably including HEY1 and BMPs signaling pathways. **Conclusion:** This study discovered new putative susceptibility genes of CRC through multi-ancestry TWAS analyses on the deconvoluted expression data. Our findings provide novel insights into the roles of these genes in CRC development.

Session Title: Cancer Poster Friday Session

Board 8045F: Nanoparticle-mediated CRISPR-Cas9 Delivery Targeting BCR/ABL for Chronic Myeloid Leukemia

Authors:

K. Chen, Y. Cao, M. Saha; Georgia Inst. of Technology, Atlanta, GA

Abstract:

Chronic myeloid leukemia (CML) makes up 15% of all leukemia cases and affects over 34,000 people worldwide. Moreover, 30% of CML patients develop resistance to tyrosine kinase inhibitors, the accepted treatment, due to leukemic stem cells. Nanoparticle-encapsulated CRISPR-Cas9 mRNA presents a prospective alternative CML treatment by knocking out the BCR/ABL oncogene in CML stem cells. Remarkably, few studies have employed CRISPR-Cas9 mRNA for CML treatment. These studies use the industry standard, lipid nanoparticles, for Cas9

mRNA encapsulation; however, many alternative nanoparticles with potential tumor-inhibiting effects, such as Chitosan and hybrid exosome-liposome nanoparticles (HELNs), have additional benefits and remain untested for CML treatment. For instance, exosome nanoparticles resist immune system-mediated degradation and do not provoke adverse immunogenic responses. As such, this initial research tests the cytotoxicity of various novel nanoparticles in vitro by synthesizing liposomes, chitosan, and HELNs. All nanoparticle groups were delivered to a K562 human myeloid leukemia cell line; as a non-cancerous control, all nanoparticle groups were also administered to a BHK-21 hamster fibroblast cell line. Cell viability was measured among both cell lines and compared. Chitosan nanoparticles were synthesized by dissolving chitosan with acetic acid and loading with Cas9 mRNA. Lipofectamine LTX and Lipofectamine 3000 nanoparticles were loaded with Cas9 mRNA as positive controls. K562 and BHK-21 derived exosomes were prepared by centrifuging their respective conditioned media with PEG-8000. Exosomes and liposomes were fused using three freeze-thaw cycles to synthesize HELNs with 1:0, 1:1, 4:1, and 1:4 ratios of number of exosomes to number of liposomes with either Lipofectamine LTX or Lipofectamine 3000 as the liposome component. Saline was tested as a negative control. K562 and BHK-21 cells were treated with each group, and the cytotoxicity of each treatment was determined with CellTiter-Glo 2.0® luminescence viability assay. A positive relationship was found between exosome concentration and K562 and BHK-21 cell viability with R^2 values of 0.64 and 0.85, respectively. Furthermore, BHK-21 cells treated with Chitosan nanoparticles were significantly more viable than K562 cells treated with Chitosan nanoparticles with a p-value less than 0.001. These results suggest the promising potential of HELNs and Chitosan nanoparticles, as both have tumor-inhibiting effects that can be leveraged for CML treatment in combination with their encapsulated CRISPR-Cas9 mRNA targeting the BCR/ABL oncogene.

Session Title: Cancer Poster Friday Session

Board 8046F: Novel ETV6 fusion transcripts in acute leukemia were identified by Next Generation Sequencing (NGS) test

Authors:

N. Htun, S. Aslam, P. Kongtim, D. Jeyakumar, X. Zhao, S. Rezk, F. Quintero-Rivera, J. Chan, Y. Zhang; Univ. of California Irvine, Irvine, CA

Abstract:

The ETV6 gene, a transcription factor, can become a proto-oncogene through mutations and gene fusions. More than 40 ETV6-containing fusion transcripts have been reported, with the

majority identified in hematologic malignancies and soft tissue tumors. Here we present three acute leukemia cases with identified novel ETV6 fusion transcripts.Case 1: A 53-year-old male with a history of Burkitt's lymphoma and diffuse large B-cell lymphoma was transferred to our institution with fatigue. A bone marrow biopsy (BMBX) showed markedly hypercellularity with 31% myeloblasts, consistent with therapy-related acute myeloid leukemia (t-AML). A targeted NGS panel detected an ETV6-TMEM57 fusion. A highly complex karyotype was also detected. Vyxeos and Decitabine/Venetoclax were given but did not achieve remission. The patient's condition was further complicated due to neutropenic fever and infections. He was transferred to the intensive care unit and passed away later.
Case 2: A 61-year-old woman with a history of breast ductal carcinoma was found to have leukocytosis. BMBX showed AML with 40% blasts. Karyotype and FISH studies showed translocation (11;19) leading to KMT2A rearrangement. She went through chemotherapy and gained remission. However, she had a relapse and salvage with azacitidine plus Venetoclax were started and during cycle 1 of the treatment, BMBX showed 3-4% blasts and plasma cell dyscrasia. Karyotype revealed t(11;19) and a subclone with both t(11;19) and t(4;12), NGS identified a novel fusion gene ETV6-MAML3. Her treatment was changed to FLAG-IDA+Venetoclax before a stem cell transplant was performed. The patient remains in remission after the transplant.
Case 3: A 65-year-old male was admitted to our hospital due to newly diagnosed Acute Lymphoblastic Leukemia (B-ALL). BMBX showed B-ALL with Philadelphia-like features involving a 90% markedly hypercellular marrow with almost all cells being blasts. FISH results showed IKZF1 deletion, CRLF2 and IGH rearrangements, and deletion of the distal side of ETV6 gene. NGS results revealed an ETV6-LOH12CR1 fusion. The patient was enrolled in a clinical trial at UCI with Blina + low-intensity chemotherapy, however, the patient developed subdural hematoma and global cerebral edema, deteriorated and passed away.Conclusion: NGS technologies enable us to discover new fusion transcripts, which can provide new insights into the underlying molecular mechanism for hematological malignancies. In addition, all cases have complex karyotype or other chromosome abnormalities, thus the presence of the ETV6 fusions may represent a secondary event during clonal evolution and disease progression.

Session Title: Cancer Poster Friday Session

Board 8047F: Pangenome references improve biomarker estimation from tumor sequencing data

Authors:

E. Arslan, R. Sears, D. Turgut, A. Jain; Velsera, Boston, MA

Abstract:

It has recently been shown that patients from non-European ancestries are at a higher risk of inappropriate clinical intervention because of inaccurate biomarker estimation, arising from the reference bias inherent in standard methods for determining the tumor genome from sequencing data. Here we show that these inaccuracies can be reduced by using a pangenome reference appropriate for the patient's population. We constructed a novel secondary analysis workflow where the pangenome reference serves as a scaffold for mapping the sequencing reads, and is also included in the relevant 'panel of normals' needed to discriminate between germline and somatic mutations in tumor-only sequencing. We show that this approach detects known somatic mutations in tumor-only sequencing more accurately than the standard somatic calling workflow (Broad's Best Practice) on a standard benchmark tumor sample, HCC1395 (33% relative increase in F1 score). We also assessed the expected clinical impact of our approach by comparing the Tumor Mutational Burden (TMB) calculated from missense somatic mutations called in tumor/normal samples from 5 patients self-reported as belonging to African and 3 patients self-reported as belonging to European populations, where Broad's standard workflow has estimated the TMB value greater than 10 in each case. We find that the TMB values calculated from the tumor-only sequencing data analyzed by our workflow more closely approximate the TMB values calculated from the tumor-normal analysis of the same sample, being 17% higher on average (STD ±12%), whereas Broad's tumor-only analysis generates TMB values 33% higher on average (STD ±35%) than the tumor-normal analysis of the same sample. Tumor-normal TMB values calculated by the two methods do not vary as drastically, Broad's workflow resulted values being 15% higher on average (STD ±11%), indicating that Broad's standard tumor only analysis leads to significant overestimation of TMB values, which can be largely corrected by using our workflow when tumor-normal sequencing is not available. These results indicate that our approach of pangenome-based analysis has the potential to become the new standard for unbiased processing of somatic sequencing samples, following on from its increased adoption for germline sequencing analysis.

Session Title: Cancer Poster Friday Session

Board 8048F: Patterns of hepatitis b virus mutations associated with hepatocellular carcinoma is controlled by HLA genotype.

Authors:

M. Sugiyama[1], N. Nishida[2], K. Tokunaga[3]; [1]Natl. Ctr. for Global Hlth.and Med., Ichikawa, Japan, [2]Natl. Ctr. for Global Hlth.and Med., Ichikawa, Chiba, Japan, [3]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Abstract:

Background: Several viral mutations have been reported to predict the risk of developing hepatocellular carcinoma (HCC) in chronic hepatitis B. In particular, the C1653T and A1762T/G1764A mutations in the core promoter region have been widely reported, but these mutations may not be reproducible between studies. The reasons for the uncertain reproducibility remain unclear, partly because it is unclear whether this is due to differences in patient populations between studies. On the other hand, genetic polymorphisms in the HLA-DPB1 region have been reported in patients with HCC in GWAS-based analyses, highlighting the importance of HLA genotypes. **Aims:** In this study, patients were classified according to their HLA-DPB1 genotypes, and the patterns of HBV mutations associated with hepatocarcinogenesis were analysed. **Method:** A total of 2996 cases were enrolled in this multicentre study in Japan. These included 408 cases of CHB, 307 cases of HCC and 2281 healthy subjects. In this study, among the HLA-DPB1 genotypes, we focused on HLA-DPB1*05:01 (about 40%) and HLA-DPB1-02:01 (about 20%), which are commonly carried by Japanese, and established four groups of homo- and heterozygotes of each. Within these four groups, HBV sequence comparisons were made between CHB and HCC matched for age, sex and HBV genotype. **Results:** In the analysis of the patient carrying HLA-DPB1-05:01 as homozygote, the A1762T/G1764A mutation in the core promoter region was significantly detected in the HCC population compared to the CHB population ($p < 0.001$, OR=8.8: 95%CI 4.4-17.8). In addition, the C1653T mutation was significantly detected in the HCC population with heterozygous HLA-DPB1*05:01 compared to the CHB population ($p < 0.001$, OR=4.4: 95%CI 2.0-9.6). In contrast, analysis of the homozygous HLA-DPB1*02:01 population revealed a significant C129T mutation in the PreS2 region ($p < 0.02$, OR=23.8: 95%CI 4.5-119.8). Similarly, analysis of the HLA-DPB1-02:01 heterozygote population revealed a significant C129T mutation in the PreS2 region ($p < 0.01$, OR=33.1: 95%CI 11.3-96.5). The C129T mutation of the HBV genome in the PreS2 region was novel. **Conclusion:** The pattern of HBV mutations associated with HCC development varies according to HLA genotype. Stratification using human disease-related genes partly explains why viral mutations differ between individuals. This stratification approach to risk assessment may lead to useful prognostication.

Session Title: Cancer Poster Friday Session

Board 8049F: $PM_{2.5}$ exposure is associated with EGFR somatic mutations in non-small cell lung cancer

Authors:

Y. Ding[1], J. LoPiccolo[1], K. Taraszka[1], Y. Wei[2], J. Schwartz[3], A. Gusev[1]; [1]Dana-Farber Cancer Inst., Boston, MA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Background: Air pollutants, especially particulate matter with a diameter ≤ 2.5µm ($PM_{2.5}$) have been linked to lung cancer incidence, while the underlying mechanism remains largely unknown. Recent studies have shown $PM_{2.5}$ exposure can promote lung cancer development in healthy lung tissues carrying pre-existing epidermal growth factor receptor (EGFR) somatic mutations. However, the mutagenic effect of $PM_{2.5}$ on EGFR has not been explored.
Methods: In this study, we investigate the association between $PM_{2.5}$ exposure with EGFR mutation in a multi-ancestry retrospective observational study of 3192 non-small cell lung cancer (NSCLC) patients, who underwent OncoPanel sequencing of lung tissue between April 2013 and March 2024 at the Dana Farber Cancer Institute. The $PM_{2.5}$ exposure was retrieved from NASA's Socioeconomic Data and Applications Center and linked to each patient based on their residential ZIP code.
Results: We observe a significant association between $PM_{2.5}$ and EGFR mutation among patients of European ancestries (N = 2820 and the majority of the cohort). Within this group, 23.9% live in areas where $PM_{2.5}$ levels exceed United States air quality guidelines (> 9 µg/m³). These patients exhibit a notably higher EGFR mutation rate compared to those from lower exposure areas (22.8% vs 16.9%, p = 0.0006). After adjusting for sex, age at sequencing, smoking status, cancer stage and technical variables, 1 µg/m³ increase of $PM_{2.5}$ is associated with an odds ratio (OR) of 1.10 (95% CI 1.02-1.19, p = 0.018) for EGFR mutation incidence. The association is even stronger among non-smokers (OR = 1.15, 95% CI 1.04-1.28, p = 0.009, N = 985). Furthermore, we identified organic carbon, a key component of $PM_{2.5}$ resulting from biomass combustion, as the main driver of the observed association out of 15 main PM2.5 constituents (OR=1.99, 95% CI 1.34-2.95, p = 0.0006). In contrast, KRAS mutation is not associated with $PM_{2.5}$ or organic carbon.
We further investigate if differential $PM_{2.5}$ exposure explains the different prevalence of EGFR mutation across ancestral groups. While patients of East Asian ancestries have a much higher EGFR mutation rate (55.1%) than those of European ancestries (17.4%), the residential $PM_{2.5}$ level is only marginally higher for the former (9.4 vs 8.3 µg/m³), indicating that ambient $PM_{2.5}$ level is unlikely to explain the difference of EGFR mutation incidence across ancestries within US.
Conclusion: Our results highlight a potential role of $PM_{2.5}$ in inducing EGFR mutation among NSCLC patients, particularly in non-smokers, demonstrating the intricate influence of the environment in cancer development.

Session Title: Cancer Poster Friday Session

Board 8050F: Prediction of anthracycline-induced cardiomyopathy in long-term childhood cancer survivors: the power of polygenic risk scores

Authors:

N. Sharafeldin[1], X. Wang[2], L. Zhou[1], P. Trainor[1], P. Singh[1], W. Landier[1], L. Hageman[1], K. Oeffinger[3], E. J. Chow[4], S. Armenian[5], J. Neglia[6], A. Ritchey[7], D. Hawkins[4], J. P. Ginsberg[8], S. Bhatia[1]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Florida Intl. Univ., Miami, FL, [3]Duke Univ., Durham, NC, [4]Fred Hutchinson Cancer Ctr., Seattle, WA, [5]City of Hope, Durate, CA, [6]Univ. of Minnesota, Minneapolis, MN, [7]Children's Hosp. of Pittsburgh of UPMC, Pittsburgh, PA, [8]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Anthracyclines are associated with dose-dependent risk of cardiomyopathy in childhood cancer survivors. The inter-individual variability in risk of anthracycline-induced cardiomyopathy is explained in part by genetic variation at the host level. In this study, we aim to test the prediction performance of survivor-specific and general population polygenic risk scores of cardiomyopathy in childhood cancer survivors. Study cohort consisted of 421 non-Hispanic White, childhood cancer survivors diagnosed with cancer at age ≤21y (140 with cardiomyopathy; 281 without matched on primary cancer diagnosis, year of diagnosis, and duration of follow-up). Median age was 7.8y for cases and 6.9y for controls at time of primary cancer diagnosis and 22.2y for cases and 20.7 for controls at time of enrollment, mean anthracycline dose was higher for cases (307 v 162 mg/m2 for controls) and cases were more likely to have received chest radiation (27.1% v 16.7% for controls). Genotyping was performed using Illumina arrays. The Michigan Imputation Server was used for genotype imputation using 1000Genomes reference panel, and for polygenic score calculation of general population PGS for cardiomyopathy (PGS002051) derived from UKBB participants of European ancestry. We curated a list of 57 candidate single nucleotide polymorphisms (SNPs) previously implicated in anthracycline-induced cardiomyopathy to form a standardized survivor-specific PRS (PRS57). Conditional logistic regression models adjusted for age at diagnosis, sex, anthracycline dose, and chest radiation were used to estimate PRS associations. Area under the receiver operating characteristic curve (AUC) was used to evaluate the performance of models including the PRS compared to clinical only models. Both PRS57 (adjusted odds ratio (aOR)=2.5 per standard deviation (SD), 95%CI=1.8-3.5, p<0.001) and PGS002051 (aOR=1.5 per SD, 95%CI=1.1-2.0,

p=0.005) were significantly associated with cardiomyopathy. Compared to a model with clinical variables only AUCclin=0.66, 95% CI: 0.61-0.71; a model including a PRS had significant predictive performance for both PRSs (PRS57: AUCclin+PRS=0.77, 95% CI: 0.73-0.81, p<0.001; PGS002051: AUCclin+PRS=0.72, 95% CI: 0.67-0.77, p<0.019. A combined clinical and survivor-specific PRS57 model significantly improved risk prediction performance compared to the combined clinical and general population PGS002051 model (p=0.015). Superior performance of PRS57 suggests a need to validate the survivor-specific PRS in an independent population prior to utilization in targeted intervention strategies aimed to predict cardiomyopathy risk in childhood cancer survivors.

Session Title: Cancer Poster Friday Session

Board 8052F: *TERT*p.K1050N: A founder variant in Ashkenazi Jewish populations associated with telomere biology disorders

Authors:

K. De Andrade[1], E. Pinto[2], L. Zeigler P.[3], J. Kim[1], T. Zhao[3], N. Giri[1], J. Haley[4], O. Florez-Vargas[1], D. Smelser[4], L. McReynolds[1], G. Urban[4], G. Zambetti[2], T. Bryan[5], D. Carey[4], C. Garcia[6], D. Stewart[1], S. Savage[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Cancer Genomics Res. Lab., Leidos BioMed. Res., Frederick Natl. Lab. for Cancer Res., Frederick, MD, [4]Dept. of Genomic Hlth., Geisinger Clinic, Geisinger, Danville, PA, [5]Children's Med. Res. Inst., Faculty of Med. and Hlth., Univ. of Sydney, Westmead, Australia, [6]Div. of Pulmonary and Critical Care Med., Columbia Univ., New York, NY

Abstract:

Telomeres are nucleoprotein complexes essential for maintaining the stability of chromosomal ends. Germline variants in telomerase reverse transcriptase (*TERT*) can disrupt telomerase function, leading to a spectrum of illnesses known as telomere biology disorders (TBDs). Typical TBD-phenotypes include bone marrow failure, acute myeloid leukemia, myelodysplastic syndrome, pulmonary fibrosis, and liver diseases. Two clinically diagnosed TBD families, including six heterozygous and one homozygous family member, were detected with the *TERT* p.K1050N (c.3150G>C, rs373400596) variant in the participating institutions. The homozygous individual developed bone marrow failure as a teenager and avascular necrosis of the hips in the 20s. Heterozygous individuals developed diseases as adults, including liver disfunction, oral leukoplakia, peripheral neuropathy, pulmonary fibrosis, among others. The p.K1050N variant is rare in gnomAD v4.1.0 (non-UK Biobank, UKB) at a global minor allele frequency (MAF) of 0.0087%. However, it is highly enriched in the Ashkenazi Jewish population

(MAF = 0.2%). Two additional population databases contain a total of 37 seemingly unrelated (up to the third degree) individuals with the p.K1050N variant, 22 (60%) of whom were genetically determined to be of ASH ancestry. Of note, one of the TBD families self-reported as of ASH ancestry. Based on these observations, our hypothesis was that p.K1050N may be a founder ASH variant. We analyzed whole-genome sequencing from 32 heterozygous individuals to evaluate a potential shared haplotype. We identified a common haplotype spanning a region of 734,219bp (hg 38 chr5:540,514-1,277,733) which includes six rare SNVs (rs988282022, rs1004401785, rs1315386864, rs548800022, rs993878614, and rs925606770). The STR D5S1981-allele 260 was identified as the marker within the haplotype that segregates along with the six rare SNVs in all available DNA samples (n=12 individuals). Y-chr STR markers and mitDNA haplogroups were assessed to determine lineage and ancestry. Y-STR analysis identified the R1a, R1b, and G1b haplogroups while mitDNA sequencing detected the H, K, L2, and T haplogroups, all commonly found in the ASH population. Our findings support the designation of *TERT* p.K1050N as a founder variant of ASH ancestry. Assays to define the consequences of p.K1050N on telomerase function are underway. Additional studies to further investigate the clinical phenotypes are required to determine if genetic screening for the p.K1050N variant is warranted in the ASH population.

Session Title: Cancer Poster Friday Session

Board 8053F: The Multiethnic Cohort: A Resource for the study of Genetic and non-Genetic Cancer Risk Across Populations

Authors:

D. Bogumil[1], X. Sheng[1], P. Wan[1], L. Xia[1], S. Streicher[2], B. Z. Huang[1], C. W. K. Chiang[1], F. Chen[1], L. R. Wilkens[2], L. Le Marchand[2], C. A. Haiman[1], D. V. Conti[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of Hawai'i Cancer Ctr., Honolulu, HI

Abstract:

Introduction: The Multiethnic Cohort Study (MEC) was established to identify cancer risk factors across five self-reported racial and ethnic groups (African American [AA], Japanese American [JA], Latino [LA], Native Hawaiian [NH] and White [WH]) in the U.S. Here we use the new large-scale MEC genetics data set for genetic risk prediction across this diverse population.
Methods: The MEC includes over 215,000 individuals aged 45-75 at baseline (1993-1996) from California and Hawaii, with genome-wide germline genotyping and imputed data for 73,139 participants (10,962 AA, 24,234 JA, 17,242 LA, 5,488 NH, and 14,649 WH). In prospective analyses, we estimated the association between genetic risk and incident cancer using weighted

genetic risk scores (GRS) based on the most recently published cancer-specific GWAS. We fit Poisson time-to-event models for each cancer and estimated the cumulative risk of each cancer across age and GRS quantiles.
Results: There were 1,880 breast (BC), 1,340 colorectal (CRC), and 2,899 prostate (PCa) incident cancer cases diagnosed following biospecimen collection. The largest increase in risk was seen among WH for PCa, where a 1-SD increase in GRS was associated with a 2.13-fold (95%CI 1.93, 2.34) increase in incidence. The cumulative risk of PCa and BC is considerably higher than that of CRC due to differences in baseline risk and GRS effect across populations. For example, at age 80, a self-reported African American male in the 95% GRS quantile had a 43.1% (95% CI: 36.9%, 50.9%) lifetime risk of PCa, whereas a white male with the same age and GRS percentile had a risk of 35.9% (29.9%, 43.8%). For CRC, the cumulative risk for a GRS in the 95% percentile ranges from 3.8% (2.9%, 5.2%) for WH to 9.0% (7.2%, 12.2%) for JA, among males and females. BC had similar estimates of cumulative disease risk across groups ranging from 19.9% (15.6%, 26.7%) for NH to 14.4% (11.9%, 18.3%) for JA.
Discussion: Inadequate representation of individuals from populations hinders the global translational impact of genetic findings. These results highlight the unique and valuable characteristics of the MEC genetic database and cohort. These data offer the ability to conduct genetic discovery and fine-mapping and to explore the interface of genetic factors, non-genetic factors, and health disparities within a single prospective cohort developed with methods validated across participating populations.

Session Title: Cancer Poster Friday Session

Board 8054F: The Palestinian founder allele TP53 p.R181C predisposes to high risk of breast cancer but not to the pediatric cancers of Li-Fraumeni or Li-Fraumeni-like syndrome

Authors:

S. Lolas-Hamameh[1], S. Lieberman[2], T. Walsh[3], M. Lee[3], R. Beeri[2], A. Said Sarahneh[1], H. Fridman[2], G. Lazer Derbeko[2], T. Klopstock[2], O. Freireich[2], A. Aburayyan[3], S. Gulsuner[3], A. Lahad[4], M-C. King[3], E. Levy-Lahad[2,4], M. Kanaan[1]; [1]Bethlehem Univ., Bethlehem, Palestinian Territory, [2]Shaare Zedek Med. Ctr., Jerusalem, Israel, [3]Univ. of Washington, Seattle, WA, [4]The Hebrew Univ. of Jerusalem, Jerusalem, Israel

Abstract:

Loss-of-function mutations of *TP53* are responsible for young-onset breast cancer and other cancers of Li-Fraumeni syndrome (LFS) and Li-Fraumeni-like syndrome (LFLS), but the clinical

consequences of partial-loss-of function *TP53* variants are not fully understood. Among women of Palestinian ancestry, one such variant, TP53 p.R181C, is heterozygous in 1.3% of all breast cancer patients and in 10% of breast cancer patients diagnosed by age 40y, while absent from female Palestinian population controls and virtually absent worldwide. Using information from 25 Palestinian families carrying R181C, we calculated the risk of breast cancer among TP53 p.R181C heterozygotes as 50% (SE 7%) by age 50y and 81% (SE9%) by age 80y. For comparison, we analyzed 103 Palestinian families ascertained through a breast cancer patient matched in age at diagnosis to an index case heterozygous for TP53 p.R181C, but with no P/LP variant in any known breast cancer gene. We compared cancer history of first-degree relatives (FDRs) of these two sets of index cases. (We note that FDRs of TP53 carriers who were not genotyped have a 50% chance of being a carrier, and this would be reflected in the risks observed). Among female FDRs of TP53 mutation carriers, breast cancer risk by age 70 was 29% (SE 6%) versus 11% (SE 6%) among female FDRs of index cases of comparison families (RR=3.7, P<0.00001). To understand pediatric cancer risks we compared frequencies of pediatric cancers (age dx<20y) among FDRs of confirmed TP53 p.R181C heterozygotes from the 25 TP53 families versus FDRs of the index case of the 103 comparison families. Pediatric cancers were reported in 3 of 519 relatives of TP53 p.R181C carriers and in 7 of 1082 relatives of the comparison patients: OR = 0.90, 95%CI [0.23, 3.49], Fisher P=0.9 (2-tailed). Pediatric cancers in TP53 families were brain cancer, renal cancer, and an ophthalmic tumor. No cases of adrenocortical or choroid plexus tumors were observed. Pediatric cancers in the comparison families were leukemia (3 cases), bone cancer (2 cases), brain cancer and testicular cancer. For families harboring TP53 p.R181C, this result is important because it suggests that it is not necessary to carry out genetic testing of children, nor to attempt onerous follow-up LFS surveillance in heterozygotes. For cancer genetics generally, the result suggests that a TP53 missense allele can predispose to very high risk of breast cancer with no pleiotropic effects.

Session Title: Cancer Poster Friday Session

Board 8055F: The role of pre-diagnostic circulating metabolites in prostate cancer risk: A cross-population meta-analysis of untargeted metabolomic studies

Authors:

H. Fuller[1], O. P. Agasaro[1,2], P. Wan[3], L. Pooler[3], L. R. Wilkens[4], L. Le Marchand[4], D. Albanes[5], D. V. Conti[3], C. A. Haiman[3], B. F. Darst[1]; [1]Fred Hutch Cancer Ctr., Seattle, WA, [2]Univ. of Washington, Seattle, WA, [3]Univ. of Southern California/Norris Comprehensive Cancer Ctr., Los Angeles, CA, [4]Univ. of Hawaii Cancer Ctr., Honolulu, HI, [5]Natl. Cancer Inst., Bethesda, MD

Abstract:

Introduction: Prostate cancer (PCa) is the 2nd most common cancer in men and presents a major health disparity. While several epidemiologic studies have shown that metabolic dysregulation may contribute to PCa pathogenesis, pre-diagnostic metabolite evidence has not been quantitatively aggregated. **Methods**: We conducted a metabolome-wide association study (MWAS) to examine associations between pre-diagnostic circulating metabolites (n=831) and PCa risk in African ancestry individuals (355/360 cases/controls) from the Multiethnic Cohort (MEC). Logistic regression models adjusted for batch, birth year and sample collection year were performed. We then conducted a systematic review of pre-diagnostic untargeted circulating metabolomic studies of overall, aggressive, and lethal PCa risk in any population, and meta-analyzed identified studies with MEC findings. A Bonferroni adjusted P-value threshold was used to define significance. **Results**: In African ancestry individuals from MEC, 16, 7, and 7 metabolites were nominally (P<0.01) associated with overall, aggressive, and lethal PCa risk, respectively. Asparagine and taurousodeoxycholate were nominally associated with increased risk of all outcomes. No metabolites were significant following multiple testing correction. Our systematic review identified 14 studies (11,768/12,184 cases/controls) of predominately European ancestry. In total, 498, 297, and 591 metabolites were meta-analyzed for overall, aggressive, and lethal PCa, respectively. In European ancestry studies, four metabolites were significantly associated with lethal PCa risk: three lipids (3-hydroxybutyrate, glycerol, and stearoyl ethanolamide) were associated with increased risk and one nucleotide (dihydroorate) was associated with decreased risk. These four metabolites were nominally associated with lethal PCa risk in cross-population analyses, with no effect heterogeneity between populations. In cross-population meta-analyses, the fatty acid ethylmalonate was significantly associated with lethal PCa (OR=1.45, CI=1.24-1.69, P_{adj}=0.002). Ethylmalonate was nominally associated with increased lethal PCa risk in European but not African ancestry participants, although the effect direction was consistent. **Conclusions**: Pre-diagnostic levels of four metabolites were associated with lethal PCa risk in European ancestry participants. Ethylmalonate was associated with increased risk of lethal PCa across populations, a fatty acid previously associated with risk of breast and ovarian cancer. Future work is needed to evaluate the clinical utility of these metabolites as lethal PCa biomarkers.

Session Title: Cancer Poster Friday Session

Board 8056F: Towards personalised medicine for Caribbean breast cancer: the Bermuda cohort

Authors:

C. Weldon[1,2], K. Burchall-Busby[1], K. Fox[1], S. Pacheco[1], D. Dahary[1], P. Burgess[1], T. Tucker[1], J. Hall[1], S. Beck[1], P. Costello[3], K. Hughes[4], M. Jain[5,1]; [1]CariGenetics, Hamilton, Bermuda, [2]Bermuda Coll., Paget, Bermuda, [3]Wessex Clinical Genetics Service, Southampton, United Kingdom, [4]Med. Univ. of South Carolina, Charleston, SC, [5]Northeastern Univ., Boston, MA

Abstract:

There is evidence to suggest that black women are diagnosed with more advanced breast cancers and those that have fewer treatment options as compared to other ethnic groups. Four of the top 10 countries dying from breast cancer are Caribbean, including the number one spot. Currently all genes linked to inherited breast cancer risk are from European genomic studies. Although previous studies have shown BRCA1/2 testing of Caribbean women missed 13% of inherited breast cancers, to date there has been no genomic study to find unique Caribbean genetic markers for breast cancer.

We conducted a first-of-its-kind study on Caribbean genetic markers of breast cancer in a Bermuda cohort by generating 102 whole genomes (51 patient group and 51 control group) using long-read nanopore sequencing at approximately 30X coverage. Each genome was analysed using Claire, Sniffles and Spectre then annotated using Geneyx Analysis. Using computational filtering of the whole genome data we reported on 30 high risk cancer genes, observing an inherited breast cancer rate 2X higher than expected at 20%. Non-BRCA genes accounted for 60% of the mutations, including DNA mismatch repair genes, an observation not reported elsewhere in the Caribbean or world. Structural variant calling identified additional pathogenic deletions missed by normal SNV calling, highlighting the value of long-read sequencing for clinical applications.

Furthermore, we searched for frameshift, nonsense, and missense variants as predicted to be damaging by at least 3 prediction tools using Geneyx Analysis, revealing RAD50, RAD51 and RAD54 novel candidate variants found more in patients than controls. The RAD family of proteins are involved in DNA repair, increasing the accumulation of deleterious mutations that can lead to higher cancer risk. Some candidate variants are also unique to the cohort, seen at higher rates than global genomic databases, likely being founder mutations unique to the Bermuda population.

Large structural variants and methylation were analysed for the first time in a Caribbean population, uncovering new findings linked to small island population structure and inheritance. This work is the purest Caribbean genomics study to date (inclusion criteria was four Caribbean grandparents) and the first time an entire clinical genomics research study was conducted in the Caribbean by the Caribbean, from recruitment to analysis. We anticipate this to serve as a model for more genomic studies in nations where local genomics capacity is needed, fuelling more diversity in genomics and better accessibility for all.

Session Title: Cancer Poster Friday Session

Board 8057F: Trio exome sequencing identifies increased incidence of clonal hematopoiesis of indeterminate potential in Bloom Syndrome probands and carriers.

Authors:

A. Wei[1], I. Lin[1], T. Gebo[1], P. Boutros[1], M. Flanagan[2], N. Kucine[2], C. Cunniff[2], V. Arboleda[1], V. Chang[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Weill Cornell Med. Coll., New York, NY

Abstract:

Background: Bloom Syndrome (BSyn) is an autosomal recessive disorder caused by recessive or compound heterozygous variants in BLM. BLM is a RecQ helicase, a family of helicase enzymes that maintain genomic stability. The genomic instability caused by biallelic germline variants in BLM drives the incidence of early-onset cancer, particularly hematologic malignancies, in patients with BSyn. The malignancy risk in BSyn carriers (heterozygous for deleterious BLM variants), remains controversial and understudied. The underlying mechanisms leading to increased cancer risk may involve increased somatic mutation rate, de novo mutations that are not inherited from parents, aberrant epigenetic alterations which can be caused by mutations in DNA methylation (DNAm) genes, or through tissue-specific somatic mutations that drive clonal expansion. For example, in blood, clonal hematopoiesis of indeterminate potential (CHIP) is caused by somatic mutations in leukemia-related “CHIP” genes in individuals without overt evidence of leukemia but associated with increased risk of leukemia.
Methods: In this study, we performed exome sequencing on BSyn probands (n=10), BSyn carriers (n=19), control children (n=19) and control adults (n=38). We obtained these 19 age- and sex-matched control trios from the publicly available dbGAP study phs001272.v1.p1 deposited by the Broad Institute Center for Mendelian Genomics. We analyzed the prevalence of CHIP, the presence of somatic variants in DNAm genes, and the occurrence of de novo germline mutations using a multinomial model.
Results: We discovered significantly increased occurrence of low-frequency, putative somatic CHIP variants in both BSyn probands (p-value = 0.0017) and BSyn carriers (p-value = 1.4093E-06) compared to age-matched controls. We also identified an increased occurrence of somatic variants in DNAm genes in BSyn carriers. These findings suggest that the presence of even just one germline pathogenic variant in BLM could be sufficient to increase the risk of clonal hematopoiesis, which may be used as a biomarker of aging, cancer, cardiovascular disease,

morbidity and mortality.
Conclusion: We find statistically-significant, increased CHIP and somatic DNAm gene variants in BSyn probands and Bsyn carriers compared to age-matched controls. Our findings contribute to growing literature that carriers for genes important in maintaining genomic integrity also have a higher somatic mutation rate and increased cancer risk. Further studies are warranted to confirm these findings in population-scale cohorts and to follow up on the impact of BLM loss-of-function variants on cancer risk.

Session Title: Cancer Poster Friday Session

Board 8058F: Unveiling the Gut Microbiome Network Fingerprint: A Deep Learning Approach to Predict Immunotherapy Response in Melanoma Patients.

Authors:

J. Sun[1], Z. Fan[2], H. Park[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Stanford Univ., Stanford, CA

Abstract:

Immune checkpoint inhibitors (ICI) are designed to help the immune system fight cancer. Among multiple important biological layers, the human gut microbiome, a complex ecosystem of trillions of microorganisms, has emerged as a key player influencing the development of the immune system. Despite extensive research across diverse cohorts, there remains a lack of consensus on the specific microbiome characteristics associated with ICI responses, impeding the translation of these findings into clinical practice. We hypothesize that the extensive and nonlinear interactions among gut bacteria are key drivers of the response to ICIs. To test this, we developed a novel deep neural network (DNN) model-driven network analysis method, termed Deep Learning-based Gut Microbiome Network (DL-gutMicroNET). The DNN component of DL-gutMicroNET identifies nonlinear interactions between gut bacteria, while the network analysis component systematically decodes the gut microbiome interaction patterns. Utilizing data from five cohorts of advanced melanoma patients undergoing ICI therapies, we demonstrate how DL-gutMicroNET characterizes bacterial interactions, predicts immunotherapy responses, and elucidates potential functions in host biology. For instance, we found that highly connected gut bacteria, such as Faecalibacterium.prausnitzii, Enterobacteriaceae, Ruminococcus_gnavus, Blautia_massiliensis, are more strongly associated with positive responses to ICIs. Also, such bacteria are associated with key immune pathways, such as antigen processing and presentation. Altogether, DL-gutMicroNET offers novel insights into the gut microbiome

mechanisms underlying immunotherapy responses, paving the way for improved predictive models and personalized treatment strategies.

Session Title: Cancer Poster Friday Session

Board 8059F: Use of national genome strategy to determine rate of germline pathogenic mutations among Emirati women with early onset breast Cancer

Authors:

B. Alkaf[1], A. Al Awadhi[1], A. Saad[1], A. El-Khani[1], c. cremin[1], D. Sanchez[1], A. Attia[1], A. Al Mannaei[2], F. Al Marzooqi[1], F. Khan[3], G. Katagi[1], H. Wu[1], H. Al Mabrazi[1], I. Eltantawy[4], J. Quilez[1], J. Mafofo[1], L. Petalidis[1], M. Ibrahim[4], M. Al Ameri[2], R. Rawashdeh[3], S. De Sylva[5], S. Khan[1], T. Magalhaes[1], V. Zvereff[1], V. Kusuma[1], W. Abdulrahman[2], M. Al Manei[2], M. Pimental[1], A. Koshy[1], H. Al Nowais[1], S. Grobmyer[3]; [1]M42, Abu Dhabi, United Arab Emirates, [2]Dept. of Hlth.of Abu Dhabi, Abu Dhabi, United Arab Emirates, [3]M42; Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates, [4]M42; IROS, Abu Dhabi, United Arab Emirates, [5]M42; Malaffi, Abu Dhabi, United Arab Emirates

Abstract:

Background: Early onset breast cancer (EOBC) (<40 years) is more common among Emirati and other Arabic women compared to Western populations. The etiology of EOBC in this population remains unclear, and the rate of pathogenic germline mutations among Arab patients with breast cancer is not known. In the United Arab Emirates, the Abu Dhabi health information exchange platform, Malaffi, holds an estimated 1.8 billion records for 7.4 million women, while the Emirati Genome Program (EGP) has sequenced ~70,000 women with ~30X whole-genome sequencing. These government-sponsored programs together form one of the world's largest population genomic resources and serve as an integrated approach that is pivotal for identifying novel pathogenic genetic variants that can have a significant impact on the Emirati healthcare system, and enhances understanding of disease etiology in this unique and understudied population. Methods and Results: Following IRB approval, we analyzed the number of patients in the Malaffi health information exchange who also had whole genome sequencing as part of the EGP (n=1,100). The mean age of this population was 51 years, with 188 patients (17.1%) having EOBC. The analysis of germline pathogenic and likely pathogenic mutations among these 1,100 patients is currently ongoing to determine the overall rate and types of germline mutations and to assess whether there is an increased rate of mutations among patients with EOBC compared to other populations. Conclusions: Malaffi and Emirati Genome Program are

powerful tools for precision medicine, showcasing the potential of genome-first approaches. In this case, it allowed us to determine the rate of germline genetic mutations among a large population of Emirati breast cancer patients. Information from this analysis can be helpful in implementing public health strategies to reduce the burden of breast cancer and ovarian cancer and improve outcomes.

Session Title: Cancer Poster Friday Session

Board 8060F: Variation in bulk RNA-seq and estimated cell type proportion using deconvolution when comparing pancreatic cancer samples within the same individual ⭐

Authors:

R. Jansen[1], S. A. Munro[1], S. O. Antwi[2], K. G. Rabe[3], H. Sicotte[3]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Mayo Clinic, Saint Johns, FL, [3]Mayo Clinic, Rochester, MN

Abstract:

The promise/utility of using genomic data to inform individualized cancer therapy has been a longstanding goal for clinicians and patients. However, little attention has been given to using genomic data to inform sampling strategies to enhance the diagnostic and prognostic yield of tissue sampling. By assessing intratumor genomic heterogeneity, some studies have shown it may be possible to target biopsies to tumor subclones driving disease progression or treatment resistance. In this study, we investigated whether targeted sampling may be important for genomic analysis by analyzing multiple surgical samples from the same individual. We used bulk RNA-seq data on 17 patients that were part of the TCGA and performed a separate bulk RNA-seq using FFPE samples from the same 17 patients (repeat sequencing study). Three deconvolution methods (dtangle, nnls, and gprogwc) were implemented in the granulator R package to compare cell type proportions for these paired samples. We normalized study-specific gene expression values within each gene by calculating transcripts per million to compare average expression values between the two bulk RNA-seq samples from the same patient. We selected *KRAS*, *TP53*, *SMAD4*, and *CDKN2A* for evaluation as these are most frequently mutated in pancreatic cancer. We selected *ADAM9*, *BMI1*, *CBX4*, and *CDH1* as these have been validated to have altered expression in pancreatic tumors compared with adjacent normal tissue. We observed that average intra-patient cell type proportion varied most between the TCGA and repeat sequencing study for NK cells and macrophages. For the differential expression analysis, we observed that the average intra-patient gene expression of *KRAS*, *SMAD4*, *CBX4*, and *CDH1*, significantly varied between the TCGA data and the repeat

sequencing data, after adjusting for multiple testing. Further, we used the median as a cut point for each gene expression value to calculate expression scores for each patient per gene. Then we determined concordance in expression profiles between the TCGA data and the repeat sequencing data. We observed very low concordance (kappa = 0.005) across the two RNA-seq datasets, suggesting repeated sampling of tissues, as opposed to one-time sampling, may be a better approach to developing biomarker panels to inform treatment plans for patients. Although these two RNA-seq studies used resected tissues from the same patients, our sample size was small and different sequencing platforms were used; thus, the findings need verification in other larger studies. The long-term goal of this area of research is to harness the power of genomic analysis to optimize biopsy/sampling decisions and treatment plans.

Session Title: Cancer Poster Friday Session

Board 8061F: Whole-genome RNAi screen to identify modulators of prometastatic gene ASAP1

Authors:

S. Claffey[1], S. Lin[1], Y-C. Chen[1], P. A. Randazzo[2], K-C. Cheng[1]; [1]Natl. Ctr. for Advancing Translational Sci., NIH, Rockville, MD, [2]Natl. Cancer Inst., NIH, Bethesda, MD

Abstract:

High expression of the Arf GTPase activating protein ASAP1 correlates with metastasis and poor prognosis in several cancers, including uveal melanoma and soft tissue sarcomas. ASAP1 is a known regulator of actin structures and mesenchymal tissue differentiation. The pathways that regulate ASAP1 expression are largely unexplored despite its correlation with poor prognosis in cancers with high expression levels. To find genes that modulate the expression of ASAP1, we conducted a series of RNAi screens using a HiBiT reporter assay. The team used CRISPR-Cas9 to modify a U2OS cell line with a HiBiT luminescence reporter on the C-terminus of ASAP1. HiBiT is a small 11-amino acid reporter peptide that will bind with high affinity to its larger complement protein, LgBiT, forming the luminescent enzyme, NanoBiT. This system can be used to quantify the amount of our HiBiT-tagged protein of interest because the luminescence intensity is proportional to the amount of HiBiT-tagged protein. Using this U2OS-HiBiT cell line, we validated the assay using negative and positive siRNA controls and optimized three assay parameters in the 384-well format: cell number, lipid transfection reagent concentration, and incubation time point. We then conducted a pilot kinome RNAi screen format to confirm that the assay parameters chosen were optimal. Next, we performed a whole-genome RNAi screen targeting 21,584 human genes. Through analysis of the resulting luminescence readout, we

found genes that, when silenced, could downregulate ASAP1 expression. The resulting gene hits from this screen will provide insight into ASAP1 regulation and could be potential therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5001F: A Comparative Study of Molecular Landscape of Prion Diseases in Humans Suggests Novel Drug Repurposing Opportunities

Authors:

S. Mehruf[1], A. Titus[2], S. Khader[3], K. Yadav[4]; [1]Allen D. Nease High Sch., Saint Johns, FL, [2]Dept. of Preventative Cardiology, Houston Methodist, Houston, TX, [3]Sch. of Publ. Hlth., Faculty of Med., Imperial Coll. London, London, United Kingdom, [4]Sch. of Engineering Med., Texas A&M Univ., Houston, TX

Abstract:

Prion disease is a broad group of progressively fatal neurodegenerative diseases associated with the misfolding of a host-encoded cellular prion protein, PrP. The most common form of prion disease, Creutzfeldt-Jakob Disease (CJD), affects approximately 1 million Americans yet there is no treatment or cure. In this study, we thoroughly delve into the intricate molecular correlations between different subtypes of prion disease and seek to identify the common genes and biochemical pathways associated with prion disease to pinpoint potential opportunities for computational drug repositioning. Among humans, the subtypes are CJD, sporadic CJD (sCJD), variant CJD (vCJD), iatrogenic CJD (iCJD), Gerstmann-Sträussler-Scheinker disease (GSS), and Fatal Familial Insomnia disease (FFI). Disease-gene associations were meticulously compiled from surveys of published genome-wide association studies (GWAS) data from the DisGenet database. Enrichr was utilized to identify significantly enriched biological pathways and functional similarities. Among the different subtypes, notable examples of the common genes are the PRNP gene (which encodes a prion protein) and the MDH1 gene (which encodes an enzyme that catalyzes NAD/NADH-dependent oxidation in metabolic pathways). An instance of a common pathway across the diseases includes ferroptosis, an intra-cellular iron-dependent form of cell death. We have also identified potential drugs that could be thoroughly explored to treat prion disease: gossypol, a phenolic aldehyde that permeates cells and acts as an inhibitor for several dehydrogenase enzymes including MDH1. Further meticulous study and exploration of these genes, pathways, and proteins provide substantial evidence of significant genetic

correlations between different subtypes of prion disease, suggesting common genetic foundations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5002F: A compendium of splicing variation in the Alzheimer's disease risk gene *TREM2*

Authors:

M. Lukes[1], K. Kiianitsa[2], O. Korvatska[2]; [1]Dept. of Med., Div. of Med. Genetics, Univ. of Washington, Seattle, WA, [2]Dept. of Psychiatry and Behavioral Sci., Univ. of Washington, Seattle, WA

Abstract:

A substantial fraction of genetic variation in Alzheimer's disease (AD) are noncoding variants that regulate gene expression, but unlike changes in protein code, impact of regulatory variants is not easily predictable. Efforts of the Alzheimer's Disease Sequencing Project (ADSP) amassed over 60,000 exomes/genomes from multi-ethnic cohorts, identifying many novel variants that await functional assessment. Up to 50% of variation associated with human diseases are variants that affect gene splicing. Currently known codon-altering risk alleles in AD genes explain at best 30% of AD susceptibility, emphasizing the need to investigate variants that regulate gene expression and splicing.
TREM2, an immune receptor on microglial cells, is associated with a spectrum of neurodegenerative conditions, including AD. *TREM2* is subject to species-specific alternative splicing. We recently discovered that multiple missense *TREM2* variants linked to early-onset dementia or late-onset AD risk and thought to inactivate the protein also change RNA splicing, decreasing functional transcript. Herein, we performed a comprehensive characterization of *TREM2* variants identified in ADSP cohorts for their effect on splicing. Variants with strongest predicted splice-altering scores were validated experimentally by a full-gene *TREM2* splicing reporter assay in microglia-derived cells and their effects on splicing mechanism were further characterized.
701 SNPs in the *TREM2* locus were retrieved from VariXam, an aggregated ADSP database of variants identified in genome/exome analyses of 56,864 individuals. For splicing effect prediction, we used pre-computed SpliceAI scores made for three changes at each nucleotide position of 4.5 kb *TREM2* genic sequence identifying 181 variants with positive scores. We further prioritized 70 variants and screened them with a full-gene *TREM2* splicing reporter assay in HEK293T and microglia-derived HMC3 cells. We confirmed loss-of-function phenotypes for

SNPs residing in donor/acceptor sites and in polypyrimidine tracks. We identified multiple missense and synonymous variants that increased exon skipping and/or changed exon starts. We also discovered a novel cryptic *TREM2* exon whose usage was enhanced by two deep intronic variants.
Our efforts provide a comprehensive, nucleotide resolution map of splice-altering variation in *TREM2* that will guide assessment of the clinical significance of rare SNPs found in AD cohorts, for which association-based testing is either impossible or impractical. Moreover, the strategy is expandable (or can be adapted) for studies of splicing variation in other AD risk genes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5003F: A comprehensive MHC-class-II region association-scan for the study of DQ allotypes and race influence on the risk of FVIII inhibitor development in hemophilia-A patients

Authors:

M. Almeida[1], V. Diego[1], K. Viel[2], B. Luu[3], A. Ameri[4], K. Haack[5], R. Rajalingam[6], C. Ramsey[7], E. MANUSOV[8], J. Galan[1], J. Powell[3], J. Peralta[9], H. Meade[10], M. Escobar[11], S. Kumar[12], S. Williams-Blangero[1], C. Kasper[13], L. Almasy[14], S. Cole[15], J. Blangero[16], T. Howard[17]; [1]Univ. of Texas Rio Grande Valley, Brownsville, TX, [2]Histonis Inc., Portsmouth, NH, [3]Haplogenics Corp., Brownsville, TX, [4]Georgia Hlth.Sci. Univ., Augusta, GA, [5]Texas BioMed. Res. Inst., San Antonio, TX, [6]Univ. of California San Francisco, San Francisco, CA, [7]Clever Culture Systems, San Diego, CA, [8]UTRGV, brownsville, TX, [9]STDOI / UTRGV, Brownsville, TX, [10]BioMarin Pharmaceutical Inc., San Raphael, CA, [11]Univ. of Texas Hlth.Sci. Ctr., Houston, TX, [12]UTRGV, Sch. of Med., McAllen, TX, [13]Univ. of Southern California, Los Angeles, CA, [14]Children`s Hosp. of Philadelphia, Philadelphia, PA, [15]Texas BioMed. Res. Inst, San Antonio, TX, [16]Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX, [17]UTRGV Sch. of Med., Brownsville, TX

Abstract:

Hemophilia A (HA) is a clotting disorder caused by the Factor VIII (*F8*) gene mutations. The standard HA patient care is the administration of exogenous FVIII (tFVIII) infusions to manage bleeding events. Long-term tFVIII infusions elicit an undesired immune response 25% of the time where F8 inhibitory (F8I) antibodies hinder the treatment. Genetic factors influencing affecting F8I's presence include but are not limited to the following three major classes: (1) Causal mutations in the *F8* gene; (2) genetic variation in cellular and molecular features of the immune system; and (3) relevant racial variation, manifested as race-specific genetic variations.

The HLA region on chromosome 6 harbor multiple DQ genes where complex splicing events greatly increases the antibodies variability. To characterize the exact role of the HLAII region for the development of FVIII inhibitor's status, we performed a F8I-status association scan focused at the ~750 kb extended MHCII region. This short chromosomal region was genotyped by 926 high-quality SNPs/SNVs, present at a Immunochip GWAS platform, distributed across 16 densely packed protein genes critical to immune system function including the 11 HLAII loci encoding the α- and β-chains of *DPA1*, *DPB1*, *DQA1*, *DQB1*, *DRA*, *DRB1*, *DRB3/4/5*, *DMA*, *DMB*, *DOA*, and *DOB*. The F8I status genetic association of each SNPs/SNVs was individually analyzed using a linear mixed model considering the pairwise genetic relationship of all samples as implemented on SOLAR. A subset of 534 SNPs is located near the seven classical HLAII loci. Under the best overall model, the variant rs1049225 (C/T) located at position 32659970 on chromosome 6 (*DQB1* gene), had the highest association signal (P.value = $5.75*10^{-7}$), giving an odds ratio (OR) and 95% confidence interval (CI) of 1.19 [1.10, 1.28]. The best overall model proposed in this report accounted for the following covariates (each followed by OR [95% CI]), age: 1.00 [0.99,1.00]; baseline severity: 0.88 [0.75, 1.05], exposure to tFVIII: 1.04 [0.93, 1.17], race (Caucasian coded 0 and African-ancestry (AA) coded 1): 1.10 [1.00, 1.20]; and a six-level *F8* mutation rank (MR) variable (where the intron 22 inversion was the reference group coded MR1): MR2: 1.29 [1.14, 1.46]; MR3: 1.18 [0.99, 1.42]; MR4: 1.25 [1.08, 1.45]; MR5 1.46 [1.16, 1.85]; and MR6: 2.26 [1.60, 3.18]. Our results confirm a previous scientific report where a greater F8 inhibitor risk was observed of African American patients. The novel association, located at *DQB1* gene, identified needs to be further replicated but, if confirmed, will greatly increase our understanding about the genetic factor influencing F8I development.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5004F: A genome-wide analysis of pleiotropy between morning circadian preference and BMI reveals the tissue-specific rhythmicity of ADCY3 in adipose tissue.

Authors:

C. Tchio[1,2,3,4], M. Maher[1], C. Moth[5], J. Meiler[5], T. Herman[6], J. Lane[7], R. Saxena[8]; [1]Massachussetts Gen. Hosp., Ctr. for Genomic Med., Boston, MA, [2]Broad Inst., Program in Med. and Population Genetics, Cambridge, MA, [3]Morehouse Sch. of Med. Cardiovascular Res. Inst., Atlanta, GA, [4]Harvard Med. Sch., Boston, MA, [5]Ctr. for Structural Biology, Vanderbilt Univ., Nashville, TN, [6]Morehouse Sch. of Med., Cardiovascular Res. Inst., Atlanta, GA, [7]BWH, Boston, MA, [8]Mass Gen. Hosp., Harvard Med. Sch., Boston, MA

Abstract:

Background: Chronotype is the behavioral manifestation of our internal circadian clock, influencing whether individuals are early birds (morning people) or night owls (evening people). Chronotype significantly impacts complex diseases such as obesity. This study dissects the relationship between circadian preference and body mass index (BMI) using a genome-wide pleiotropy approach.
Methods: We conducted a genome-wide pleiotropy analysis to explore the associations between morning circadian preference and BMI, focusing on the tissue-specific expression and rhythmicity of key genes.
Results: We identified a lead locus, rs11676272 in *ADCY3*, associated with increased morning preference and decreased BMI. The rs11676272 (S107P) is a missense variant that destabilizes the ADCY3 protein structure, highlighting the gene's role in both circadian behavior and metabolic processes. In mouse models, *Adcy3* expression was rhythmic in white and brown adipose tissues during constant darkness, and it oscillated anti-phase to the canonical circadian genes Clock and Bmal1. The presence of the Bmal1-specific E-box motif CACGTG on the Adcy3 gene suggests Bmal1 regulation of this rhythmicity, indicating a novel mechanism of circadian control within these tissues.
Conclusion: Findings from our genome-wide pleiotropy approach underscore a novel role of ADCY3 at the intersection of circadian regulation and metabolic health. This study offers new insights into the genetic architecture of these complex traits, highlighting the significance of tissue-specific rhythmicity in adipose tissue.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5005F: A genome-wide association study of cancer-related cognitive impairment in patients who received chemotherapy for non-metastatic breast cancer

Authors:

A. Olowofela[1], K. Mustian[2], M. Janelsins[2], E. Culakova[2], P. Auer[3], S. Kerns[4]; [1]Med. Coll. of Wisconsin, Milwaukee, WI, [2]Univ. of Rochester Med. Ctr., Rochester, NY, [3]Med. Coll. OF WISCONSIN, Milwaukee, WI, [4]Med. Coll. of Wisconsin, Wauwatosa, WI

Abstract:

Cancer-related cognitive impairment (CRCI) is reported in up to 75% of patients during chemotherapy. It may persist for many years post-chemotherapy and it impacts the quality of

life of survivors. The knowledge on genetic predisposition to CRCI is limited. Thus, we aimed to identify genetic variants associated with CRCI in women who receive chemotherapy for non-metastatic breast cancer (NMBC). We performed a genome-wide association study (GWAS) using data from two NCI cooperative group clinical studies (URCC07012 and URCC100553) of women with NMBC who received chemotherapy and were assessed prospectively using the Functional Assessment of Cancer Therapy - Cognitive Function (FACT-Cog) questionnaire version 2. Participants completed the FACT-Cog pre-chemotherapy, and at 6 months post-chemotherapy. A decreased FACT-Cog score implied cognitive decline. We performed linear regression to identify non-genetic factors associated with change in FACT-Cog score ($p < 0.10$). Then we performed a GWAS adjusting for these factors and principal components to identify variants associated with change in FACT-Cog score. Among 802 women, median age was 55 years (range (R): 22 years - 81 years), median BMI was 29.5 kg/m2 (R: 15.3 kg/m2 - 65.4 kg/m2), most were White (86.5%), and postmenopausal (53.4%). Treatment included previous surgery (85%) and/or anthracycline-based chemotherapy (46%). There was a median decline in FACT-Cog score of -9 (R: -141 - 74). Age (Coefficient (C): 0.48; 95% CI: 0.28 - 0.67; $p <0.001$), race (Black vs White; C: 6.81; 95% CI: 0.69 - 14.31; $p = 0.075$), educational status (high school vs college or higher; C: 5.35; 95% CI: 0.34 - 10.37; $p = 0.036$), menopausal state (post vs pre; C: 7.56; 95% CI: 2.84 - 12.29; $p = 0.002$), Karnofsky Performance Score (80 vs 100; C: -8.30; 95% CI: -18.09 - 1.5; $p = 0.097$), anthracycline-based chemotherapy (vs non-anthracycline; C: -3.93; 95% CI: -8.16 - 0.30; $p = 0.069$), and pre-chemotherapy FACT-Cog (C: -0.20; 95% CI: -0.27 - -0.12; $p <0.001$) were associated with change in FACT-Cog score. An adjusted GWAS among 664 individuals with available genetic and clinical data identified a region on chr6q27 tagged by rs9365961 (chr6:166149994:C>A; C: -10.05; $p = 1.46 \times 10^{-8}$) as associated with cognitive decline. This region is downstream from LOC124901524 and upstream from TBXT, which encodes an embryonic nuclear transcription factor previously associated with susceptibility to neural tube defects. The region has no previously known clinical role in CRCI and warrants validation testing.
Funded by NCI Contract No. HHSN261201500003I, Task Order No. HHSN261000039 and by UG1CA189961, URCC NCORP Research Base.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5006F: A meta-predictor for causal gene identification in GWAS overcomes limitations of existing computational approaches ⭐

Authors:

D. Usoltsev, I. Molotkov, M. Artomov; Nationwide Children's Hosp., Columbus, OH

Abstract:

Introduction: Genome-Wide Association Studies (GWAS) represent a powerful tool for identifying genetic loci associated with diseases. However, linking association signals to causative genes is challenging since they often fall into non-coding regions of DNA. Fine-mapping can narrow down causal variants, but connecting a variant to a gene typically requires additional data like transcriptomics, proteomics, and epigenomics to prioritize genes within a locus. To reduce the high cost of generating functional data, computational techniques like transcriptome-wide association studies (TWAS) have been developed. Coupling of such data with machine learning enables integration of multiple data types with GWAS to pinpoint disease genes. However, results can be inconsistent across existing solutions, especially for complex loci or phenotypes. *We hypothesized that creating a meta-predictor incorporating previous solutions, TWAS, and a graph representation of protein interactions would provide an efficient and stable solution for selecting the most appropriate candidate gene from every GWAS locus independently of the phenotype.* **Methods:** The meta-predictor takes two essential features as input: TWAS results and posterior probabilities for each gene calculated using the Markov chain Monte Carlo (MCMC) algorithm on a protein-protein interaction network. Additionally, it incorporates previously developed Locus to Gene (L2G), Polygenic Priority (PoPS), and GPrior scores where available. These features are analyzed using XGBoost, resulting in scores that provide the probabilities of genes being causal within a locus. **Results:** We trained the model on three GWAS summary statistics: schizophrenia, inflammatory bowel disease (IBD), and coronary artery disease (CAD) and respective known disease genes. In each training instance, a meta-predictor was trained to select the correct gene from gene candidates in the ±200 Kb region around the true disease gene. As an independent validation we evaluated the ability of meta-predictor to select the genes that were not used in training and were identified by schizophrenia exome analysis and more recent GWAS (PGC3). The meta-predictor consistently showed ROC AUC of 0.73 and outperformed each individual score: PoPS - 0.59, L2G - 0.64, GPrior - 0.64. **Conclusions:** Meta-predictor provides increased performance in identifying disease genes and greater stability of the solutions across the phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5007F: A multi-omics study of cerebral vasospasm in peripheral blood from patients with aneurysmal subarachnoid hemorrhage

Authors:

J. Zhang, R. Du; Brigham and Women's Hosp., Boston, MA

Abstract:

Cerebral vasospasm (CVS) is the predominant complication of aneurysmal subarachnoid hemorrhage (aSAH), yet its molecular characteristics within blood remain inadequately defined. This study aims to unravel the hematologic biomarkers associated with CVS from a multi-omics perspective. Here, we profiled gene expressions, microRNA (miRNA) expressions, and DNA methylations, in human blood samples collected at intervals during and after the period of risk for CVS following aSAH. In-silico analysis agreed a difference between the clinically asymptomatic and symptomatic CVS groups based on sample molecular similarities at multiple omics levels. A further cross-omics analysis indicates that dysregulation in the assembly of iron-sulfur clusters and in the functional activity of T cells is intricately associated with the onset of CVS. Additionally, we established an integrated multi-omics model using the XGBoost approach, which is capable of predicting CVS severity and identifying key contributory features from heterogenous data resources. Among these identified features, we found a set of transcriptomic features, including ISCA1, miR-15b-5p, GLRX5, CA1, miR-106b-3p, showing an early brain injury-irrelevant expression pattern robustly linked with CVS. The subsequent ROC analysis demonstrated ISCA1, a core protein involved in the biogenesis and assembly of iron-sulfur clusters, as the most powerful candidate in independently discerning CVS events. Collectively, our study offers a multimodal omics resource and introduces novel biomarkers for future vasospasm investigations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5008F: A phenome-wide association study reveals the impacts of germline copy number variations on multiple traits in Taiwan

Authors:

S-H. Liu[1], T-H. Hsiao[1], Y-H. J. Shao[2]; [1]Dept. of Med. Res., Taichung Veterans Gen. Hosp., Taichung, Taiwan, [2]Graduate Inst. of BioMed. Informatics, Taipei Med. Univ., Taipei, Taiwan

Abstract:

Germline copy number variations (CNVs), a type of structural variation originating in germ cells, are rare but can have significant impacts on disease susceptibility. This study investigates germline CNVs and their associations with diseases in the Taiwanese population, utilizing large-scale genetic data from the Taiwan Precision Medicine Initiative (TPMI). By applying array-based CNV detection method, the Taiwanese germline CNV profile was characterized, revealing more duplications than deletions. We identified 111,825 across 48,489 individuals, most containing less than three CNVs, while several high-frequency CNV hotspots were also detected, including

well-known regions on chromosome 6 or 22. Top disease associations were found for nephritis and nephropathy, allergic rhinitis, esophagitis, gastroesophageal reflux disease, essential hypertension, and diabetes mellitus (DM). Furthermore, whole genome sequencing (WGS) data with certain traits or phenotypes were selected for CNV detection validation. This comprehensive analysis of Taiwanese germline CNVs provides insights into their landscape and potential roles in disease predisposition within this population.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5009F: Almost all rare genetic associations identified through whole-genome sequencing occur at loci identified from GWAS ⭐

Authors:

A. Wood[1], G. Hawkes[1], H. Wright[1], R. Beaumont[1], K. Chundru[1], K. Patel[1], L. Jackson[1], T. Frayling[2], A. Murray[1], C. Wright[1], M. Weedon[1]; [1]Univ. of Exeter, Exeter, United Kingdom, [2]Université de Genève, Geneva, Switzerland

Abstract:

Background
In 2022, the GIANT consortium published a saturated map of common genetic variation associated with adult height. In total, 12,111 single nucleotide polymorphisms (SNPs) were reported, spanning 21% of the genome based on windows +/-35kb around index SNPs. However, less is known about the proximity of potentially rare (minor allele frequency (MAF) <1%) genetic associations for height relative to previously reported common SNP associations.
Aim
Using 575,055 individuals with whole-genome sequence data in the UK Biobank (UKB) and All of Us (AoU), we aimed to assess whether rare genetic associations for height occur near common SNPs previously reported.
Method
Using 449,489 individuals from UKB, we performed single-variant association analysis of ~260 million variants with a minor allele count >=5. We obtained conditionally independent variants using GCTA-COJO. We subsequently performed aggregate-based association testing for rare variants, adjusting for the conditionally independent single-variant associations. Genetic associations reaching statistical significance (single-variants: P<3x10-10; aggregate-based: P<9x10-9) were put forwards for replication using 128,566 individuals in AoU. Proximity of associations with nominal levels of significance (P<0.05) to previously reported common SNPs was subsequently calculated.

Results
In total, 105 rare variants (19 with MAF<0.1%) demonstrated nominal evidence of replication, with 100% directional effect consistency, explaining 0.6% of the variance in height in both UKB and AoU. In total, 90/105 (86%) rare variants (15/19 with MAF <0.1%) are within 35kb of a previously reported common SNP, and 101/105 (96%) rare variants (18/19 MAF<0.1%) are within 100kb of a previously reported common SNP. In addition, we identified 81 aggregate-based associations (73 coding, 8 non-coding) with nominal evidence of replication. Overall, 60/73 (82%) coding and 8/8 (100%) non-coding aggregate associations are within 35kb of a previously reported common SNP, and 70/73 (96%) of coding aggregate associations are within 100kb. All 105 variants and 81 aggregate-based associations are within 500kb of a previously reported common SNP.
Discussion
Our results suggest that the 21% of the genome previously associated with height through array-based genome-wide association studies harbour most rare genetic associations identified through whole-genome-sequencing.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5010F: Altered TNF-α response in Down syndrome

Authors:

A. Cozart; Univ. of Colorado Anschutz, Aurora, CO

Abstract:

Down syndrome (DS) is conferred by triplication of human chromosome 21. Individuals with DS exhibit widespread immune dysregulation, accompanied by alterations in their cytokine profiles and peripheral blood transcriptome. Work over the past decade has cemented a role for interferon (IFN) signaling in this immune dysregulation, however, questions remain about the contributions of other inflammatory cytokines such as Tumor Necrosis Factor Alpha (TNF-α). Here we report plasma cytokine profiling and transcriptome analysis of whole blood for ~400 individuals from the Human Trisome Project Biobank. For genetic and pharmacological mechanistic investigations of the cellular response to TNF-α, we employed pairs of euploid/trisomy 21 fibroblast and retinal pigment epithelial cell lines. We also report the effects of TNF-α treatment on the Dp16 mouse model of DS and littermate controls. Proteomic and transcriptomic analysis of individuals with T21 revealed elevated levels of TNF-α and accompanying activation of the TNF-α signaling via NFκB gene expression program. Western blot analysis revealed a hyperactive IFN response to TNF-α treatment in T21 cell lines that is

dependent on the cytosolic DNA sensor, cGAS. Finally, treatment of Dp16 animals with TNF-α resulted in greater weight loss than in WT littermates. IFN activation may be influenced by TNF-α, indicating an interaction that could modify responses to immune challenges. Further studies will be necessary to define the significance of the TNF-IFN interaction in DS.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5011F: Alzheimer's Disease Problem History Heterogeneity and the Relationship with Mitochondrial Genetic Risk

Authors:

C. Mansel, R. Townley, D. Mazzotti, M. E. Sardiu, R. H. Swerdlow, R. A. Honea, O. J. Veatch; Univ. of Kansas Med. Ctr., Kansas City, KS

Abstract:

The inconsistent efficacy of drugs targeting Alzheimer's Disease (AD) in clinical trials may relate to the heterogeneous clinical history of patients. The medical and substance use history (i.e., problem history) likely interacts with genetic risk for AD to alter disease risk, progression, and treatment response. Mitochondrial dysfunction is a key pathophysiological mechanism underlying many human diseases including neurodegenerative and cardiovascular disease. Previous studies have found that a polygenic risk score of nuclear-encoded mitochondrial proteins (mitonuclear PRS) significantly predicts AD risk. We hypothesize that mitochondrial genetic variation explains not only risk for AD but also the underlying heterogeneous problem history and clinical presentation.
The address this hypothesis, we used an unsupervised machine learning approach to cluster 2,794 AD patients (mean (SD) age of 76 yrs, 1,237 (44.9%) male, 2,337 (84.9%) white, 2,601 (94.4%) non-Hispanic, and 15.7 (6.0) years of education) in the National Alzheimer's Coordinating Centers Database (NACC) into three distinct problem history clusters: minimal problem history (n = 1,380), substance use history (n = 1,038), and cardiovascular problem history (n = 336). The minimal problem history cluster had more individuals with earlier-onset AD, female sex, and Hispanic ethnicity. The cardiovascular problem history cluster were older at age of AD onset and more individuals were of male sex and white race; this cluster also had a lower proportion of individuals with E4/E4 *APOE* genotypes. The substance use history cluster had fewer individuals with E3/E3 *APOE* genotypes. Moreover, we used linear mixed effects modeling to show that cluster membership significantly contributes to cognitive decline variability while controlling for covariates ($F = 4.19$, $p = 0.015$). To examine mitochondrial genetic variation, we called mitochondrial haplogroups and calculated a mitonuclear PRS using

19,630 SNPs tagged to nuclear encoded mitochondrial proteins (+/- 50kb) using MAGMA . We are currently investigating not only how problem history clusters differ with regard to haplogroup and mitonuclear PRS, but also how this genetic variation interacts with cluster membership to drive AD cognitive decline.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5013F: Analyses of GWAS and subthreshold loci lead to the discovery of dendrite development and morphology dysfunction underlying schizophrenia genetic risk

Authors:

R. Chen[1], B. Siciliano[2], Q. Wang[1], C. Xu[3], Q. Wei[4], H. Yang[1], J. Sutcliffe[5], Y. Jiang[1], Y. Ji[6], C. Liu[7], F. Cheng[8], E. Cook[9], N. Cox[10], X. Zhong[11], Z. Wen[12], B. Li[13]; [1]Vanderbilt Univ., Nashville, TN, [2]Emory Univ., Atlanta, GA, [3]Vanderbilt Univ., Atlanta, GA, [4]Vanderbilt Univ, Nashville, TN, [5]Vanderbilt Univ Sch Med, Nashville, TN, [6]Vanderbilt Genetics Inst., Nashville, TN, [7]SUNY Upstate Med. Univ., Syracuse, NY, [8]Cleveland Clinic, Cleveland, OH, [9]Univ Illinois Chicago, Chicago, IL, [10]Vanderbilt Univ Med Ctr., Nashville, TN, [11]Vanderbilt Univ. Med. Ctr., Nashville, TN, [12]Emory Univ. Sch. of Med., Atlanta, GA, [13]Vanderbilt, Nashville, TN

Abstract:

Schizophrenia (SCZ) is highly polygenic, thousands of genes contribute to risk, and its biological underpinnings remain largely unclear. In this study, we explore a cost-effective strategy to increase the inference power of novel pathways, by expanding the analysis to include sub-threshold GWAS (subGWAS) loci. We observed that more than a half of the subGWAS loci from a smaller GWAS turned into significant loci (sigGWAS) after increasing sample size and the subGWAS loci are enriched in brain tissue enhancers, indicating subGWAS loci contain true signals. We identify 180 subGWAS loci (i.e., $5 \times 10^{-8} < P \leq 10^{-6}$) based on SCZ summary statistics of 40,675 cases and 64,643 controls from the CLOZUK and PGC datasets. By merging GWAS (sigGWAS) and subGWAS loci and jointly modeling the expanded set of loci, we identify in total 304 high-confidence risk genes (HRGs) using the Integrative Risk Gene Selector (iRIGS), an integrative Bayesian framework to nominated risk genes from a set of GWAS loci. Risk genes inferred from SCZ subGWAS harbor genuine SCZ risk genes and show temporal and spatial expressions related to brain development. Interestingly, we discovered a novel category of biological processes implicated in SCZ*, i.e. dendrite development and morphogenesis* (DDM)*,* that is only detected in subGWAS loci. The DDM enriches the heritability of SCZ and other related diseases, i.e. bipolar disorder and major depressive

disorder. We select three candidate DDM genes (*CUL7, DCC* and *LRP8*) for validation in human induced pluripotent stem cell (hiPSC)-derived neurons, and observe that manipulating these genes leads to abnormal dendrite development, in particular reduced neurite length compared to controls. We further reveal that the perturbation of the DDM genes leads to disrupted regulatory programs of the transcription factors *CUX1/2, NEUROD1* and *SOX5*. Collectively, our study identifies and validates DDM as a novel biological process implicated in SCZ susceptibility, and reveals regulatory programs involved in perturbation of three candidate DDM genes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5014F: Are GWAS maximizing potential discoveries with non-European ancestry data?

Authors:

E. Ha[1], C. N. Rotimi[2], A. A. Adeyemo[2], A. R. Bentley[2]; [1]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [2]NIH, Bethesda, MD

Abstract:

Genome-wide Association Studies (GWAS) are a central genomic research method, and, as is well appreciated, studies of European ancestry populations (EUR) dominate the literature. Efforts to diversify the populations in which genomic research is conducted have been made, motivated by several factors including novel discovery and the need to avoid exacerbating already unable health disparities. As more non-EUR populations are included in GWAS, evaluating how these data are being incorporated and if existing study designs maximize scientific findings is crucial. We investigated all GWAS studies published from 2019 of type 2 diabetes (T2D) and related cardiometabolic traits (retrieved from the GWAS Catalog 3/28/2024) and sought to characterize these studies in terms of if and how the study design allows for novel associations in non-EUR data.

Of the 65 publications that met our study criteria, 62% (n=40) studied a single ancestry group, of which 50% focused exclusively on EUR. 85% of single-ancestry, EUR-exclusive studies were drawn from data that included non-EUR, yet only 3 studies provided a rationale for excluding other ancestry groups. Of the 20 single ancestry, non-EUR focused publications, 35% (n=7) focused on an underrepresented population. A total of 25 studies included multiple ancestry groups. Of the 45 studies including at least one non-EUR population, ancestry groups represented included East Asian (n=18), African/African American (n=16), Hispanic/Latino (n=10), South Asian (n=9), Arabic (n=3), Native Hawaiian and Pacific Islander (n=2), Native American (n=1), and isolated populations (n=2). Three publications only included non-EUR

groups to evaluate generalizability of EUR GWAS findings, limiting non-EUR associations to those already discoverable in EUR. 50% of studies including multiple ancestry groups only presented results for combined-ancestry analyses, increasing the potential for missing ancestry-specific findings. In total, 32% of the studies including non-EUR populations were not optimized for facilitating genomic discoveries in these populations.
Despite the improvements in the inclusion of underrepresented populations in genomic research and the increasing availability of non-EUR samples, GWAS continue to be dominated by EUR-exclusive studies. Additionally, there is considerable variability in the way ancestry groups are evaluated in analyses and in how results are presented, which can limit the value of the inclusion of diverse samples. To maximize the potential for novel findings, multi-ancestry studies should prioritize evaluating data separately by ancestry groups in addition to combined analyses.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5015F: Assessing Non-Communicable Disease Risk Factors in an Admixed Population: Insights from the UAE Healthy Future Study

Authors:

G. Begum, Y. Idaghdour, U. Healthy Future Study Investigators; New York Univ. Abu Dhabi, Abu Dhabi, United Arab Emirates

Abstract:

The UAE faces major public health challenges in tobacco use, obesity, diabetes, and cardiovascular disease (CVD). These issues have complex co-determinants, including individual health habits (e.g., diet, smoking, physical activity), genetic susceptibility, the microbiome, and social and environmental conditions. To better understand these risk factors and examine their relative contributions to non-communicable diseases (NCDs), we established the UAE Healthy Future Study, the country's first prospective cohort study, focusing on approximately 15,000 UAE national men and women aged 18 and above. To date, the study generated baseline data from participants including, whole genome genotyping and sequencing data to assess population structure, relatedness and the landscape of common and rare genetic variation that has been increasingly implicated in NCDs. Our analysis, using various methods, revealed the highly admixed nature of the UAE population, alongside relatively high levels of relatedness. We conducted association analysis of blood lipid traits to investigate the impact of lipid levels on NCD risk and severity. This integrated approach provides valuable insights into how common

and rare genetic variation, and lifestyle factors interact to influence individual risk to NCDs. These results will help pave the way for targeted public health interventions in the UAE.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5016F: Assessing the role of rare pathogenic variants in heart failure progression by exome sequencing in 8,190 patients

Authors:

K. Carss[1], C. Salvoro[1], O. Chazara[2], M-P. Dube[3], A. Nag[1], Q. Wang[4], L. Middleton[5], D. Vitsios[6], A. Walentinsson[7], Q-D. Wang[7], K. Hansson[7], C. Granger[8], J. Kjekshus[9], S. Petrovski[10], C. Haefliger[2], J-C. Tardif[3], D. Paul[1]; [1]AstraZeneca, Cambridge, United Kingdom, [2]AstraZeneca, Melbourn, United Kingdom, [3]Montreal Heart Inst., Montreal, QC, Canada, [4]AstraZeneca, Chapel Hill, NC, [5]AstraZeneca, Melbourn, Royston, United Kingdom, [6]AstraZeneca, CAMBRIDGE, United Kingdom, [7]AstraZeneca, Gothenburg, Sweden, [8]Duke Univ., Durham, NC, [9]Univ. of Oslo, Oslo, Norway, [10]AstraZeneca, Fulbourn, United Kingdom

Abstract:

Most therapeutic development is targeted at slowing disease progression, often long after the initiating events of disease onset. Heart failure is a chronic, life-threatening disease and the most common reason for hospital admission in people over 65 years of age. Some common genetic factors that influence heart failure progression have been identified, but the contribution of rare variants remains mostly uncharacterised. Here, we expand the first rare-variant study on heart failure progression with larger sample sizes and updated analyses. We investigated exome sequences from 8,190 patients with heart failure across two clinical trials, CHARM and CORONA, and one population-based cohort, UK Biobank. For each cohort, we performed gene-level collapsing analyses of rare variants, using ten genetic models. Then, we assessed the contribution of gene-level genotypes on the outcomes 'time to cardiovascular death' and 'time to cardiovascular death and/or hospitalisation' using Cox proportional hazards regression with Firth's correction. From a meta-analysis of the three cohorts, we identified multiple independent exome-wide-significant associated genes, including *STIMATE* (HR [95% CI] = 4.72 [2.80, 7.93], $p = 5.02 \times 10^{-9}$), *CUTC* (Hazard Ratio (HR) [95% confidence interval (CI)] = 4.71 [2.69, 8.24], $p = 5.75 \times 10^{-8}$), *IFIT5* (HR [95% CI] = 4.59 [2.62, 8.04], $p = 9.65 \times 10^{-8}$), and *BLK* (HR [95% CI] = 6.57 [3.27, 13.19], $p = 1.19 \times 10^{-7}$). We observed a general lack of overlap with genes known to be involved in heart failure onset, suggesting that the genetic architecture underlying progression of heart failure is largely distinct from its onset. Leveraging public genomic data resources, transcriptomic and pathway analyses, as well as a machine-learning

approach, we annotated and prioritised the identified genes for further target validation experiments. Together, these findings advance our understanding of the molecular underpinnings of heart failure progression and reveal putative new therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5017F: Association of Genetic Susceptibility Factors with Coronary Heart Disease in a Diverse Cohort: an eMERGE Network Study

Authors:

M. Naderian[1], M. E. Hamed[1], J. J. Cimino[2], J. B. Cortopassi[3], A. Espinoza[4], Q. Feng[5], R. Irvin[3], G. P. Jarvik[6], L. C. Kottyan[7], T. A. Manolio[8], E. M. McNally[4], B. Namjou-Khales[9], M. J. Roy-Puckelwartz[4], H. K. Tiwari[10], W-Q. Wei[11], N. A. Limdi[3], I. J. Kullo[1]; [1]Mayo Clinic, Rochester, MN, [2]Columbia Univ., New York, NY, [3]Univ. of Alabama at Birmingham, Birmingham, AL, [4]Northwestern Univ., Chicago, IL, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Univ. of Washington, Seattle, WA, [7]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [8]Natl. Human Genome Res. Inst., Bethesda, MD, [9]Children s Hosp. Med. Ctr., Cincinnati, OH, [10]Univ. of Alabama at Birmingham, Birmingham, IN, [11]Vanderbilt Univ., Nashville, TN

Abstract:

Background: We studied associations of genetic factors with the presence of coronary heart disease (CHD) and assessed their added predictive value when incorporated into the Pooled Cohort Equations (PCE) in a diverse cohort of ~10,000 adults in the eMERGE study. **Methods:** Participants provided family history (FamHx) of CHD, were genotyped for a multi-ancestry polygenic risk score (PRS) (PGS004696) for CHD, and sequenced for familial hypercholesterolemia genes (*LDLR*, *APOB*, *PCSK9,* and *LDLRAP1)*. We used PCE to estimate the 10-year risk of CHD, and excluded those with missing age, sex, self-reported race, or genetic factors. CHD was ascertained by self-report via participant-filled surveys. Multivariable logistic regression and the area under the curve (AUC) analyses were performed. **Results:** Of 9,788 participants (mean age: 51.2±15.0, 67% female), 418 (4.3%) had CHD. Based on self-reported race/ethnicity, 56% were White, 20% were Hispanic/Latino, 16% were Black, and 7% were Asian. The prevalence of high PRS for CHD (top 5th percentile), monogenic etiology, and FamHx was 5.1%, 0.7%, and 5.1%, respectively, in adults aged <40 years and 4.0%, 0.8%, and 11.9%, respectively, in adults aged ≥40 years. In a model including genetic factors and adjusted for age, sex, and self-reported race/ethnicity, odds ratio (OR) (95% confidence interval) for having CHD for a 1 standard deviation (SD) increase in PRS was 1.31 (1.18-1.46, P=5.9×10^{-7}); for monogenic etiology 2.57 (1.14-5.24, P=0.014); and for FamHx 2.64 (2.05-3.36, P=1.2×10^{-14}). The OR for 1 SD

increase in PRS for CHD in participants with or without FamHx of CHD was 1.45 (1.15-1.82, *P*=0.001) vs. 1.28 (1.13-1.44, $P=7.6\times10^{-5}$), respectively (*P*-diff=0.346). Those with high PRS and FamHx had the highest OR for CHD: 4.28 (2.04-8.29, $P=4.3\times10^{-5}$). PRS and PCE were minimally correlated (R= 0.04, *P*=0.001), and the AUC (95% CI) of the PCE for predicting presence of CHD improved from 0.729 (0.705-0.754) to 0.751 (0.728-0.775) after incorporating PRS and FamHx. **Conclusion:** PRS and FamHx were independently and additively associated with the presence of CHD and their addition to PCE improved CHD status discrimination. These findings motivate integration of PRS and FamHx into CHD risk algorithms.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5018F: Benchmarking ancestry-aware PRS strategies for optimal prediction of PTSD in admixed samples

Authors:

H. Lin[1], A. Maihofer[2], N. Shah[1], J. Honorato Mauer[1], C. Nievergelt[3], E. Atkinson[1]; [1]Baylor Coll. of Med., Houston, TX, [2]UCSD, La Jolla, CA, [3]Univ California San Diego, La Jolla, CA

Abstract:

Post-traumatic stress disorder (PTSD) is a debilitating psychiatric disorder linked to various health problems and is an increasing public health burden. Given its polygenic nature, polygenic risk scores (PRS) hold promise in estimating an individual's genetic liability to PTSD. However, admixed cohorts, despite suffering disproportionately from PTSD among US populations, have been largely excluded from prior genomic studies. Given the complex genetic architectures in admixed groups, PRS derived from Eurocentric genome-wide association studies (GWAS) have shown poor applicability to admixed cohorts, potentially leading to unreliable predictions and exacerbating health disparities.

In this study, we applied various PRS methods to assess binary PTSD diagnosis within two-way African/European admixed African American samples from the *All of Us* Research Program, leveraging GWAS summary statistics from corresponding cohorts in the Psychiatric Genomics Consortium for PTSD (PGC-PTSD). We compared the predictive performance of PRS methods with or without consideration of local ancestry. Our results suggest that incorporating local ancestry enhances genomic resolution and prediction accuracy for PRS in admixed populations. This underscores the effectiveness of ancestry-aware approaches for complex traits in admixed individuals. This project provides the potential for broader applications across phenotypes and among other ethnic minorities, promoting health equity in underserved groups.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5020F: Blood Cell-Type Specific eQTL Analysis in a Multi-Ethnic Cohort Provides Novel Insights into Cardiovascular Disease Risk

Authors:

M. Johnson[1], Y. Cai[1], A. Molstad[2], A. Reiner[3], W. Sun[1], C. Kooperberg[1], L. Hsu[1]; [1]Fred Hutch Cancer Res. Ctr., Seattle, WA, [2]Univ. of Minnesota, Minneapolis, MN, [3]Univ. of Washington, Seattle, WA

Abstract:

Background
Cardiovascular disease (CVD) including stroke, coronary heart disease and heart failure are the leading cause of mortality in the United States, disproportionately impacting racial and ethnic minorities. In conjunction with socioeconomic factors, cardiovascular diseases share a large hereditary component. However, identifying causal single nucleotide polymorphisms (SNPs) remains a challenge due to their complex polygenic architecture. Expression quantitative trait loci (eQTL) analyses can help identify causal SNPs across populations. However, gene expression measured in bulk tissues may vary between cell types, potentially obscuring cell type-specific effects. This study aims to investigate cell type-specific eQTLs (ct-eQTLs) in a diverse cohort to enhance our understanding of genetic risk factors for CVD.

Methods
Whole genome sequencing and RNA-seq data was measured in whole blood from 1400 participants, including 350 African Americans and 1050 white non-Hispanics, enrolled in the Women's Health Initiative Long Life Study. We performed cell type-specific eQTL mapping using the CSeQTL method, leveraging allele-specific expression to enhance the detection power of eQTLs. Cell type proportions including neutrophils, monocytes, T and B-cells were estimated using CibersortX. Functional genome annotation, along with pathway enrichment was used to characterize ct-eQTLs. Finally, GWAS catalog data was integrated to identify eQTLs relevant to CVD and its risk factors including lipid levels.

Results
We identified 145,904 ct-eQTLs across 3,714 genes, with the highest number detected in neutrophils. Functional enrichment analysis revealed cell type-specific pathways relevant to disease pathogenesis. For instance, eQTL genes identified in B-cells were enriched for antigen processing pathways, while T-cell genes were associated with immune response activation. Moreover, we found significant overlap between ct-eQTL variants and cell type specific histone markers, suggesting unique genetic regulatory mechanisms. Among the identified ct-eQTLs, 871 variants across 248 genes overlapped with CVD trait loci, these included the ORMDL3 gene

associated with lipid metabolism and T-cell inflammation, and the GSDMB gene involved in the pyroptosis pathway.

Conclusions

Our study illustrates potential cell-type specific regulatory mechanisms which may play a role in disease pathogenesis. Consequently, these insights may be used to identify novel therapeutic targets related to cardiovascular disease.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5021F: Building the path to clinical polygenic risk scoring for adult-onset autoimmune diabetes across diverse ancestries in the biobank at the Colorado Center for Personalized Medicine

Authors:

M. Brasher[1,2], M. Fisher[1], J. A. Shortt[1,2], C. Sanchez Wild[1], S. Raghavan[3,4], R. K. Johnson[2,5], N. Rasouli[6], N. Rafaels[1], K. R. Crooks[1], L. K. Wiley[1,2], C. R. Gignoux[1,2], M. Lin[1,2], Colorado Center for Personalized Medicine; [1]Colorado Ctr. for Personalized Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Dept. of BioMed. Informatics, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]VA Eastern Colorado Hlth.Care System, Aurora, CO, [4]Dept. of Med., Div. of Gen. Internal Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Dept. of Epidemiology, Colorado Sch. of Publ. Hlth., Aurora, CO, [6]Dept. of Med., Div. of Endocrinology, Metabolism and Diabetes, Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Diabetes has well-defined subtypes, but differentiating among them can be challenging, particularly amidst atypical clinical presentation. A classic example is adult-onset autoimmune diabetes, where a misdiagnosis of type 2 diabetes may result in inappropriate management, increased time to diagnosis, and a greater risk of complications. Predictive PRS may assist in risk stratification, early detection, tailored treatments, and proactive monitoring, yet barriers remained, including clinical validation. Here we develop the path to clinical translation of type 1 diabetes (T1D) PRS as a potential pilot for implementation of PRSs with available intervention and clinical decision support in a large health system. We evaluated two established T1D scores: the 67-SNP GRS2 developed in a European population which includes tagging variants to call HLA DR-DQ haplotypes, HLA variants, and non-HLA variants; and a seven-SNP score developed in an African American population (AA7). We identified tag proxies for the T1D PRS SNPs on the Illumina Global Diversity Array (GDA) by maximizing linkage disequilibrium (LD) in the 1000 Genomes to find best proxies for SNPs (GRS2 mean R^2 = 0.76, AA7 mean R^2 = 0.92). We applied

both proxy scores and an additive combination of the two to 73,346 participants in the biobank at the Colorado Center for Personalize Medicine (CCPM) to evaluate predictive performance for clinician-curated diabetes subtypes. All three scores strongly differentiated between T1D cases (N=16-784) and T2D cases (N=162-5723) across genetically defined ancestry groups when correcting for demographic covariates of age, sex, batch, and first 10 genetic PCs (AUC > 0.80). All scores performed similarly well in the UK Biobank (AUC > 0.75) in differentiating T1D (N_{AFR} = 15, N_{EUR} = 732) from T2D cases (N_{AFR} = 1154, N_{EUR} = 27575) based on ICD-10 codes in genetically defined European and African ancestry groups from the Pan-UK Biobank Project. Importantly, this "case-case" setting is closer to the clinical application of disease subtyping than the commonly reported comparison to healthy controls. CCPM individuals in the top decile of scores for each ancestry had OR > 2.5 compared to those outside of the top decile. Of individuals in the top decile of the AA7 score, <10% across ancestry groups have some DM-related diagnosis but no previous T1D code in electronic health records. The results of our evaluations demonstrate the potential utility of PRS for diabetes subtyping in a clinical setting and represent critical first steps toward a standardized system for future translation of PRS to equitable clinical use.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5022F: Characterising genetics of gynecological pain conditions and interstitial cystitis

Authors:

K. Johnston[1], R. Signer[2], E. M. Hicks[3], L. M. Huckins[4]; [1]Yale Univ. Sch. of Med., New Haven, CT, [2]Icahn Sch. of Med. at Mount Sinai, Fort Collins, CO, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Yale Univ., New Haven, CT

Abstract:

Chronic overlapping pain conditions (COPCs) are a subset of chronic pain conditions commonly inter-comorbid and where pain may be predominantly nociplastic. Several conditions in this subset involve chronic pain in female genitalia, and genitourinary chronic pain. These gynecological/ genitourinary chronic pain conditions are highly prevalent (e.g., up to one third of people with vulvas in the US have vulvodynia), yet relatively understudied compared to other chronic pain conditions, with a lack of large genome wide association studies (GWAS). We carried out some of the first GWAS, SNP-heritability, and genetic correlation analyses of vulvodynia (case sample size range,13-142), dyspareunia (14-1086), any N94 (71-3464), N94 (non-menstrual) (31-1818) and IC (6-289) using International Statistical Classification of Diseases

and Related Health Problems 10th revision (ICD10) hospital inpatient data in UK Biobank (UKB) as a feasibility/ exploratory study. 'N94' refers to ICD10 category 'Pain and other conditions associated with female genital organs and menstrual cycle'. We assigned non white British UKB participants to genetic ancestry groups using 1000 Genomes superpopulations data, ADMIXTURE, and genetic principal component plots, followed by SAIGE case/control GWAS analyses per trait within each genetic ancestry group with sufficient case sample size (white British, European, African, South Asian, American). We then performed linkage-disequilibrium (LD) score regression analyses (ldsc) to calculate SNP-heritability and estimate genetic correlations for significantly heritable traits. We found 'any N94' and dyspareunia to be significantly heritable in white British UKB participants (observed scale SNP-h2 0.13, SE = 0.038, and 0.23, SE = 0.11 respectively). Any N94 was significantly ($p < 0.05$) genetically correlated with ADHD (rg = 0.45), PTSD (0.42), type 2 diabetes (0.15), menopause (0.20), and chronic widespread pain (0.46). Dyspareunia was significantly genetically correlated with multisite chronic pain (0.39), ADHD (0.45), multiple sclerosis (0.28), PTSD (0.42), broad headache (0.38), and major depression (0.35). We also find one suggestive IC SNP on chromosome 6 (rs4409171, OR = 0.05, $p = 4.76 \times 10^{-8}$) in the African ancestry subset. Caveats include small case sample sizes and significant case-control imbalance. In addition, the available UKB sample did not include non-hospital-inpatient ICD10 information. Despite these caveats, we believe this work demonstrates common genetic variation contributes to variation in gynecological pain conditions, and is an important first step in characterizing this variation.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5023F: Characterizing the Genetic Landscape of Indigenous Australians: Association of Regional-wide Autozygosity with ACR Reveals a novel region in *FTO* Gene

Authors:

V. Arunachalam[1], N. K. Tran[1], W. Hoy[2], R. Lea[1], S. H. Nagaraj[1]; [1]Ctr. of Genomics and Personalized Hlth., Queensland Univ. of Technology, Brisbane, Australia, [2]Ctr. of Chronic disease, Faculty of Med., The Univ. of Queensland, Brisbane, Australia

Abstract:

Background: The genetic distinctiveness of Indigenous Australian populations continues to garner increased attention in genomic research. Despite this, Indigenous people remain underrepresented in all genetic research. Moreover, these populations are prone to elevated Runs of homozygosity (a continuous stretch of homozygous genotypes) due to their prolonged

geographic isolation. The identification of runs of homozygosity (ROH) provides valuable insight into genetic diversity and susceptibility to diseases. In this study, we investigate the genetic diversity of the Tiwi population (a Founder population from Northern Australia), who has been separated from the mainland of Australia over decades, using ROHs, and their associations with health-related traits. **Method:** We employed whole-genome sequencing data from 455 Tiwi individuals to conduct principal component analysis, which allowed us to identify population structures in comparison to cohorts from the UK Biobank (UKBB), the Norfolk Isolate, Melanesian, and Polynesian ethnicities. Further, we assessed runs of homozygosity and computed both genome-wide and regional measures of homozygosity. Furthermore, we investigated the association between the clinical traits and autozygosity, utilizing regional-wide autozygosity as the primary predictor. **Results:** Our analysis revealed distinct genetic characteristics of the Tiwi population, notably in their kidney genetic profile compared to the UKBB British population. Tiwi individuals exhibited an increased burden of ROHs, particularly on chromosomes 2, 17, and 18, associated with gene *LINC0109*, *FMLN1*, and *RPL17P45* respectively, indicating prolonged isolation and genetic drift. Moreover, a positive correlation between genomic FROH and albumin-to-creatinine ratio (ACR) levels, suggests a potential link between the autozygosity and renal health markers. Further, the regional-wide autozygosity association identified an association between elevated ACR and region in the gene FTO, implicating its role in obesity, kidney disease, and cardiovascular conditions. **Conclusions:** This is one of the largest studies on ROHs, highlighting the genetic distinctiveness and elevated burden of ROHs in the Tiwi community. The identified associations between ROH and clinical traits emphasize the crucial importance and utility of ROH mapping. This research lays a robust foundation for further exploration of ROH mapping and its implications for disease susceptibility within Indigenous communities, underscoring the need for continued investigation in this area.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5024F: Charting the molecular and spatial transcriptomic landscape of crypt variability in Inflammatory Bowel Disease

Authors:

E. Soda[1], F. Cisternino[1], D. Ricca[1], F. Paolo[1], M. Talisa[2], D. Oukrif[2], H. Lowes[3], M. Rodriguez-Justo[2], A. P. Levine[2], C. A. Glastonbury[1]; [1]Human Technopole, Milan, Italy, [2]Res. Dept. of Pathology, UCL Cancer Inst., Univ. Coll. London, London, United Kingdom, [3]Dept. of Cellular Pathology, Univ. Coll. London Hosp. NHS Fndn. Trust, London, United Kingdom

Abstract:

Inflammatory Bowel Diseases (IBD) are complex, multifactorial disorders marked by inflammation of the gastrointestinal tract. Despite extensive research, the pathogenesis of IBD remains poorly characterised. Classified into Ulcerative Colitis (UC) and Crohn's Disease (CD), these disorders exhibit unique clinical features with unclear molecular distinctions. Histological assessment of endoscopic gastrointestinal biopsies is essential in disease diagnosis. Amongst altered histological structures in IBD, crypts show significant variability. While crypt architecture severity has known diagnostic importance and is associated with disease chronicity, the influence of crypt architectural changes and their variability on patient prognosis and outcomes is not well understood. Here, we utilise a dataset composed of ~2k colonic biopsy Whole Slide Images and the largest Spatial Transcriptomics (ST) dataset of IBD generated to date, to characterise the landscape of crypt architectural variability. We implemented a supervised segmentation model able to segment colon substructures (lamina propria, crypt, muscle and surface). The model achieves a Jaccard index of 0.847 and a F1 score of 0.902 on a holdout test set. Leveraging this model, we extracted ~600k colonic crypts, analysed using self-supervised representation learning (DINOv2). Clustering and trajectory analyses revealed a continuum from normal to abnormal crypts, including identifying clusters with distinct cutting planes, enlarged lumens, and the presence of Paneth cells. To understand the molecular variation in crypt structure, we generated ST data (Visium 10X) on a cohort of 49 donors (16 CD, 17 UC, 2 Indeterminate Colitis and 14 healthy controls). Crypt spots were assigned to substructures using the segmentation masks. Differential expression across conditions was performed using per-donor spots assigned to crypts, which resulted in 147 differentially expressed genes in UC crypts and 71 in CD crypts ($|\log FC| \geq 1$ and FDR ≤ 5%). Notably, we replicated previously described changes including LCN2 and PLA2G2A upregulation ($\log FC \geq 3.59$ and $P=1.23 \times 10^{-13}$, $\log FC \geq 2.48$ and $P=6.73 \times 10^{-8}$) and AQP8 downregulation ($\log FC \geq 3.63$ and $P=1.63 \times 10^{-8}$). In an additional cohort of 72 samples, targeted spatial transcriptomic profiling of crypts (GeoMX DSP) was conducted to validate our findings. Ongoing work aims to model the transcriptomic and morphological variability of crypts jointly and determine the role crypt architectural distortion and its regulation plays in IBD.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5025F: Common and rare variant associations with diabetic retinopathy in 18,175 cases ⭐

Authors:

J. Otto, X. Zhang, E. Jorgenson, Colorado Center for Personalized Medicine - RGC Collaboration, GHS-RGC DiscovEHR Collaboration, MAYO-RGC Project Generation, Penn Medicine BioBank

(PMBB), UCLA-RGC ATLAS Collaboration, Regeneron Genetics Center, G. Coppola; Regeneron Genetics Ctr., LLC, Tarrytown, NY

Abstract:

Diabetic retinopathy is a common microvascular complication of diabetes and a leading cause of vision loss worldwide. Genetic factors are thought to underlie risk for diabetic retinopathy; however, no replicable genetic loci have been reported to date. The aim of the current study was to test for associations of common and rare variants with diabetic retinopathy and related outcomes. Three sets of phenotypes were evaluated. These included any diabetic retinopathy as well as its clinical stages and complications (vs. type 2 diabetes controls). The latter included mild, moderate, and severe non-proliferative diabetic retinopathy, as well as proliferative diabetic retinopathy and diabetic macular edema. Diabetic retinopathy (vs. all non-diabetic retinopathy controls) was also tested. Phenotypes were defined using ICD-10 diagnosis codes and case sample sizes ranged from 395 to 18,175 across seven cohorts. Cross-ancestry genome-wide and exome-wide association analyses were conducted for single common and rare variants, as well as their gene burdens using REGENIE, and results across cohorts were pooled using inverse-variance weighted meta-analysis. Within a type 2 diabetes population, diabetic retinopathy was associated with older age, longer duration of diabetes, higher HbA1c and blood pressure levels, and other risk factors and microvascular complications. After correction for multiple testing, two genome-wide significant ($P < 5e-08$) common variant loci near *TCF7L2* and *FTO* were identified in analyses of diabetic retinopathy vs. type 2 diabetes controls, implicating known gene regions for lipid and glucose metabolism and body mass index. No gene loci were associated with more than one clinical stage or endpoint after correction for multiple testing. In the analysis of diabetic retinopathy vs. all non-diabetic retinopathy controls, extensive genome-wide common variant signal was shared across type 2 diabetes, diabetic retinopathy, and related risk factors. Gene burden analyses of the same phenotype identified associations of putative loss of function and missense variants in *GCK*, *HFN1A*, and *HNF4A*, known risk genes for monogenic diabetes of the young. The current study represents the largest investigation of genetic factors underlying diabetic retinopathy and its complications to date. The analysis of diabetic retinopathy vs. type 2 diabetes identified genes and variants that likely represent common genetic signal shared across type 2 diabetes, HbA1c, and body mass index. Future studies might seek to elucidate whether these genetic signals reflect more severe diabetes in cases or an independent role in diabetic retinopathy pathophysiology.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5026F: Common genetic risk variants in clinical single nucleotide polymorphism data in autism spectrum disorder

Authors:

M. Torres Diz[1], S. Ganesan[1,2], M. Xie[1], A. Malakar[1], J. M. Miller[1], S. M. Haag[1,2], R. Rajagopalan[1,3], L. K. Conlin[1,3], M. Gandal[3,4], I. Helbig[1,3]; [1]The Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Drexel Univ., Philadelphia, PA, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Univ. of California, Los Angeles, CA

Abstract:

Single Nucleotide Polymorphism (SNP) arrays have been used in the clinical diagnostic context to identify disease-causing copy number variations for decades. SNP data collected in large scale by various research studies and biobanks have empowered Genome Wide Association Studies (GWAS) to identify common variants associated with a particular phenotype of interest, and phenome-wide association studies (PheWAS) have further emerged to relate variants across the phenotypic landscape. However, diagnostic genotyping represents a data source with relatively untapped potential to identify such associations across the phenome, especially given the wealth of clinical information readily accessible in the electronic medical record (EMR). Here, we leverage the unique opportunity afforded by secondary research use of clinical SNP arrays of 11,860 individuals with EMR data from the Children's Hospital of Philadelphia to detect epilepsy-associated variants and perform PheWAS. We identified 3,196 cases with F84 an autism-related ICD code. GWAS analysis revealed 17 genomic loci that achieved genome-wide significance, including 4 previously reported loci associated with autism spectrum disorder (ASD), including 18q loci in *FBXO15* gene. Lastly, PheWAS across 11,312 unique ICD10 diagnoses mapped to 1,547 phecodes identified 437 phenome-wide significant SNP-phenotype associations including autistic disorder with variants in *KIF1A* and *TRERF1*. Overall, our analyses indicate the relationship between common variants and phenotypes in individuals with autism spectrum disorder and entail for further evaluation. These associations reflect the potential for discovery from merging already available clinical and diagnostic genotyping datasets with corresponding deep EMR phenotyping.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5027F: Common polygenic risk for asthma is associated with multiple atopic phenotypes in individuals with inborn errors of immunity and immune dysregulation

Authors:

J. Yan[1], S. Srinivasan[1], **L. Amendola**[2], M. Similuk[1], M. Kamen[3], C. Jodarski[1], R. Moses[4], N. Reynolds-Lallement[1], K. Lewis[1], S. A. Bannon[1], E. Damskey[3], H. Orellana[1], S. Byers[1], R. Ghosh[1], B. A. Seifert[1], M. Tokita[1], J. Lack[1], W. Cao[1], G. Constantine[1], P. Khoury[1], A. Klion[1], J. R. Bergerson[1], A. Freeman[1], S. M. Holland[1], M. Walkiewicz[3]; [1]NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]NIH/NIAID, Bethesda, MD, [4]NIAID, NIH, Baltimore, MD

Abstract:

Background: Atopic diseases are characterized by complex multifactorial etiology, with common occurrence in the general population and increased risk for severe disease in certain inborn errors of immunity (IEI). Population-based studies of asthma suggest a shared etiology across multiple atopic phenotypes. Given significant common variant associations with asthma from genome-wide association studies and rare pathogenic/likely pathogenic variants conferring increased risk for severe atopic disease, we aimed to investigate effects of common polygenic risk for asthma on atopic disease in the setting of IEI and immune dysregulation.
Methods: Genome sequencing was performed via a centralized sequencing program in a cohort of 2,067 unrelated probands with known or suspected IEI and/or immune dysregulation. Variants with allele frequency > 0.01 were used to calculate polygenic risk scores (PRS) for asthma using effects estimated in the Global Biobank Meta-Analysis Initiative. PRS based on common 834,809 variants were assessed for association with asthma, atopic dermatitis (AD), eosinophilia >540 cells/ul),, and serum IgE levels. Association was determined using logistic models controlling for sex, age, and the first 10 ancestry principal components.
Results: The cohort included 363 (17.6%) participants with asthma, 277 (13.4%) with AD, 336 (16.3%) with eosinophilia and 50 (2.4%) with elevated IgE. Participants were 53.1% female, 25.0% of non-European ancestry, and had a median age of 38 years (range 0-94). The PRS was associated with asthma (OR=1.64 [1.44,1.87], *p*=1.56-12). After controlling for asthma presence, the PRS was associated with AD (OR=1.16 [1.006,1.34], *p*=0.04) and elevated IgE (OR=1.37 [1.02,1.84], *p*=0.038). After controlling for asthma and elevated IgE, the PRS was associated with eosinophilia (OR=1.27 [1.02,1.58], *p*=0.0323). After controlling for both asthma and *STAT3*-hyper-IgE syndrome (*STAT3*-HIES), the PRS had a nominal effect on elevated IgE (OR=1.15 [0.98,1.34], *p*=0.057 and a significant effect on eosinophilia (OR=1.27 [1.04,1.44], *p*=0.012). Within the *STAT3*-HIES subset (n=56), the PRS was not associated with asthma.
Discussion: Within a sample of probands with rare immune conditions, common polygenic risk

for asthma was associated with atopic disease. This study supports an independent effect of common genetic variants on risk for atopic disease even among individuals with known or suspected IEI. Additional studies are warranted to parse causal relationships underlying these results to better understand the genetic architecture of asthma and atopic disease within the context of rare immune disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5028F: Common variants at 22q12.2 are associated with susceptibility to Tuberculosis

Authors:

X. Chang[1], L. Zheng[2], V. Phan[3], T. Dang[3], T. Nguyen[4], M. Silcocks[1], C. Chee[5], T. Nguyen[4], C-K. Heng[6], Y. Teo[6], J-M. Yuan[7], P. Koh, Woon[8], M. Caws[9], C. Khor[2], R. Dorajoo[2], S. J. Dunstan[1]; [1]The Univ. of Melbourne, Melbourne, Australia, [2]Genome Institue of Singapore, Singapore, Singapore, [3]Pham Ngoc Thach Hosp., Ho Chi Minh City, Viet Nam, [4]Oxford Univ. Clinical Res. Unit, Ho Chi Minh City, Viet Nam, [5]Tuberculosis Control Unit, Tan Tock Seng Hosp., Singapore, Singapore, [6]Natl. Univ Singapore, Singapore, Singapore, [7]Univ. of Pittsburgh, Pittsburgh, PA, [8]Natl. Univ. of Singapore, Singapore, Singapore, [9]Liverpool Sch. of Tropical Med., Liverpool, United Kingdom

Abstract:

Background: Tuberculosis (TB) continues to be a leading cause of morbidity and mortality worldwide. Despite improvements in treatment and diagnostic methods, TB remains a formidable health challenge. Past Genome-wide association studies (GWAS) have explored TB susceptibility across various ethnic groups. Yet, a significant portion of TB heritability remains unexplained, suggesting the presence of additional, undiscovered susceptibility loci.**Methods:** In this study, we conducted individual GWAS in both the Singapore Chinese and Vietnamese populations, followed by a comprehensive meta-analysis incorporating independent East Asian data, encompassing a total of 11,841 prevalent pulmonary TB (PTB) cases and 197,373 controls.**Results:** We identified a novel PTB susceptibility locus at 22q12.2 [rs6006426, OR (95%Cl) =1.097 (1.066, 1.130), $P_{meta} = 3.31\times10^{-10}$]. Our lead single nucleotide polymorphism (SNP) was found to affect the expression of splicing factor 3a subunit 1 (*SF3A1*) in various immune-related cells, with the allele associated with increased risk for PTB linked to decreased expression of *SF3A1* (*P* ranging from 1.48×10^{-9} to 6.17×10^{-18}). Furthermore, a significant association was observed between rs6006426 and cigarette smoking, a well-established risk factor for TB ($P < 0.044$). When exploring the interplay between genetic marker, smoking and

TB, our findings indicated that smoking status significantly mediated the effect of rs6006426 on PTB; that is rs6006426 allele A decreased the likelihood of smoking, decreasing the risk of PTB ($\beta_{indirect\text{-}effect}$ = -0.004, $P_{indirect\text{-}effect}$ = 0.020).**Conclusions:** Our findings offer novel insights into the genetic factors underlying TB and reveals new avenues for understanding its etiology.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5029F: Context specific cell-cell communications reveal Aβ-related signals in human microglia for Alzheimer's disease

Authors:

Y. Tan[1,2], R. Chen[1,2], A. Tiwari[1,2], Y. Yan[3], X. Zhong[2,4], B. Li[1,2]; [1]Dept. of Molecular Physiology and Biophysics, Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Genetics Inst., Nashville, TN, [3]Dept. of Biostatistics, Vanderbilt Univ., Nashville, TN, [4]Dept. of Med., Div. of Genetic Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Microglia play a dominant role in the genetics of Alzheimer’s disease (AD). Microglia communicate with nearly all brain cell types to maintain homeostasis and keep the brain healthy, while abnormal cell-cell communications (CCCs) involving microglia can contribute to the pathogenesis of AD. Besides, microglia are central players in the brain's response to amyloid-beta (Aβ), the hall marker of AD. Nevertheless, specific interactions between microglia and other brain cells, such as astrocytes and neurons, in the pathogenesis or clearance of Aβ, remain largely unknown. In this study, we aim to identify microglia-centered CCCs associated with Aβ in AD to reveal how microglia influence disease progression through their interactions with other cell types. As the dissection of CCC requires modeling the signaling pathways, we generate a cell type-specific prior knowledge of ligands on their effects on the receiver cell for AD to better model cell type-specific functions. We integrate single-cell RNA sequencing (scRNA-seq) data with the cell-specific prior knowledge and calculate context-specific communication scores to identify CCC and the involved ligands and receptors associated with Aβ burden. Specifically, we define a communication score as the product of the mean expression of a ligand in the sender cell type and the mean expression of a receptor in the receiver cell type. We observe that cell type-specific CCCs from excitatory neurons to microglia, involving the ligands ADAM23 and the receptors ITGA5, are positively associated with the increased levels of Aβ. For astrocytes and Oligodendrocyte progenitor cells (OPC), we observe ligand-receptor pairs of MEGF10-ABCA1 and GPC6-INFRSF11B, respectively. We further investigated the ligand activity using NicheNet and replicated the CCC involving ligands ADAM23 and GPC6 whose activities are

significantly associated with Aβ. We are exploring the context-specific communication scores at the single-cell level to dissect the fine-scale CCC implicated in AD and map their downstream signaling and regulatory impact in microglia.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5030F: Contribution of common and rare variants to schizophrenia risk in East and South Asian ancestries ⭐

Authors:

S. Kim[1], A. Hassan[2], J. Vincent[3], B. Neale[4], M. Ayub[5], J. Knowles[6], S. Qin[7], H. Huang[4]; [1]Broad Inst., Cambridge, MA, [2]Univ. of Peshawar, Peshawar, Pakistan, [3]Ctr Add/Mental Hlth, Clarke Div, Toronto, ON, Canada, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Dept. of Psychiatry, Univ. Coll. London, London, United Kingdom, [6]Rutgers Univ. HGINJ, Piscataway, NJ, [7]Shanghai Jiao Tong Univ., Shanghai, China

Abstract:

Schizophrenia genetic studies have largely focused on European ancestries, leaving genetic variants in other populations underexplored and potentially increasing health disparities. Recent studies have somewhat rectified this for common variants in East Asian ancestries, but rare variants remain largely uncharacterized in non-European populations. Here, we report the initial results from the first large-scale schizophrenia sequencing study that examines both common and rare variants in East and South Asian populations, using an innovative, cost-effective sequencing technology. We developed Blended Genome Exome (BGE), a novel sequencing method merging low-pass whole genome (2-4x) and deep whole exome sequencing (~40x). This cost-effective technology accomplishes both sequencing tasks in a single experiment, outperforming arrays in non-European ancestries by removing the need to pre-select ancestry-biased variants of interest. We have conducted BGE sequencing on large cohorts of 44,450 individuals with East and South Asian ancestries, equally split between cases and controls. Moreover, we have access to a total of approximately 119,180 samples (55,148 schizophrenia cases and 64,032 controls), which includes not only BGE data but also existing data comprising SNP array and WES data. This diverse dataset will significantly enhance our research capabilities and enable us to make robust and comprehensive analyses. Our preliminary findings are derived from a comprehensive analysis of currently available summary statistics, SNP array, and BGE dataset. Across these three data platforms, we rigorously processed and conducted quality control, resulting in 85,415 samples and 17,019,507 variants. In our initial genome-wide meta-analysis of 40,304 individuals with schizophrenia and 45,111

controls, we identified 40 independent genomic loci associated with schizophrenia in East and South Asian ancestries. The current study discovered over five times the number of variants associated with schizophrenia in a previous EAS schizophrenia discovery GWAS study. Additionally, we identified significant variants within the MHC region, which were not detected in the previous study. Through the integration of both common and rare variants within these datasets, we can prioritize genes primarily associated with schizophrenia (gene prioritization). This study will provide, for the first time, deeper insight into the contribution of common and rare variants to schizophrenia risk and may have implications for selecting candidate genes in pharmacogenetics to intervene in or treat schizophrenia, especially in Asian ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5031F: Deciphering the complex cases of Early-Onset Parkinson's Disease: Advanced Insights from Next Generation Sequencing Technologies

Authors:

N. Tran[1], R. Rautsaw[2], M. Gavrielatos[1], D. Kammeraad[3], M. G. Heckman[1], A. Beasley[1], Z. K. Wszolek[1], O. Ross[1]; [1]Mayo Clinic, Jacksonville, FL, [2]PacBio, Menlo Park, CA, [3]Golden Helix Inc, Bozeman, MT

Abstract:

Background: Parkinson's disease (PD) is the second most common progressive neurodegenerative disease and currently there is no effective treatments that halts/stop disease progression. A subset of PD patients with early-onset symptoms (EOPD) presents unique challenges and opportunities for understanding the disease's underlying mechanisms. To date, less than 10% of patients with EOPD are genetically explained. Short-read whole-genome sequencing (SR-WGS) has been effective in identifying single nucleotide variants yet struggle with complex structural variations like repeats, inversions, and large deletions. Long-read sequencing (LRS) overcome short-read limitations to capture complex genomic regions, uncovering novel genetic loci.**Methods:** In our current study, we have generated both SR-WGS and PacBio LR-WGS for a subset of 24 Early-onset PD (EOPD) patients (AAO<39). For both Illumina short reads and PacBio HiFi long reads, we utilized a deep-learning based variant caller using a convolutional neural network approach DeepVariant (Google) to call genetic variations in aligned reads. We have also comprehensively analyzed copy number variants, tandem repeats, structural variants and CpG methylation on PacBio HiFi reads. We also conducted a sub-analysis where we overlapped SVs and SNVs for compound heterozygotes discoveries using

VarSeq/GeneyX as an annotation tool. **Results:** We find that at the individual sample level, there are significant higher number of coding variants and SNPs captured by PacBio LRS compared to Illumina SRS. We also found highly biological-relevant candidate variants located in *Slc6a3,* a gene encoding dopamine transporter, that partially explains the patient's clinical phenotypes. **Conclusions:** Our study provides a robust side-by-side comparisons for the conventional SR-WGS and state-of-the-art LR-WGS to answer to the question of the utility and need for long-read sequencing to identify variants that likely have a phenotypic impact, paving the way for novel discoveries of previously unknown relevant variants. We believe that understanding the genetic variation in both nuclear and mitochondrial genomes along with epigenetic modifications using long read sequencing could provide novel variants for unsolved complex cases.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5032F: Determination of the function of ZBTB40 in osteoblast maturation

Authors:

T. Orndorff[1], R. Maynard[2], C. Farber[3], C. Ackert-Bicknell[4]; [1]Univ. of Colorado, Aurora, CO, [2]Univ. of Kentucky, Lexington, KY, [3]Univ. of Virginia, Charlottesville, VA, [4]Univ. of Colorado Anschutz, Aurora, CO

Abstract:

Osteoporosis is the progressive loss of bone mass and strength with age, leading to increased risk of fragility fracture. This disease is exceedingly common, and approximately half of all women can be expected to suffer at least one fragility fracture after menopause. Bone mineral density (BMD), a clinically measurable trait that is inversely correlated with fracture, is highly genetic and up to 85% of the variance in this phenotype can be attributed to heritable factors. Bone is formed by osteoblasts that first generate the collagenous matrix of bone and then mineralize it. We conducted a GWAS for the phenotype of "in vitro bone formation" and identified a locus on mouse Chromosome 4 that was syntenic with a human Chromosome 1 GWAS locus for BMD. The most likely causative gene for these loci is *ZBTB40,* an uncharacterized gene that is presumed to be a transcription factor based on protein domain structure. To better understand the role of this gene in osteoblast function, we generated loss of function mice using CRISPR and mice homozygous for this mutation make a truncated version of *Zbtb40* that lacks the BTB domain. We collected calvarial osteoblasts from both *Zbtb40 mut/mut* and *wt/wt* mice, subjected these cells to differentiation media for 48

hour and collected cells for scRNAseq. Trajectory analysis conducted using Monocle 2.0 showed that the wildtype sample contained all cell types expected from immature to mature osteoblasts, whereas the mutant cells had an overabundance of immature osteoblasts relative to mature cells and a disrupted maturation trajectory with loops in the early stages. Using siRNA, we knocked down expression of *Zbtb40* in MC3T3 cells, a mouse pre-osteoblast like cell line and bulk RNA sequencing was conducted 2 days after differentiation was induced. DESeq2 analysis was used to identify differentially expressed gene with an adjusted p-value less than 0.05 and |log2FoldChange| > 1. To identify overrepresented categories of genes in the over and under expressed gene sets, gene ontology assessments using Visual Annotation Display (VLAD) was conducted. As expected, genes associated with gene translation (G0:0006412), gene expression (GO:00010467) and bone mineralization (GO:0030282) were decreased upon knockdown of *Zbtb40*. Interestingly, among the up regulated genes we saw an over representation of genes annotated to collagen-containing extracellular matrix (GO: 0062023) and integrin binding (GO:0005178). These results show the critical role of ZBTB40 in osteoblast function and physiology. A better understanding of the function of this gene may also provide new avenues for treatment of this common disease.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5033F: Development of a Polygenic Risk Score for Accurate Differentiation of Type 1 and Type 2 Diabetes in Adults Using At-Home DNA Testing

Authors:

Y. Matsuda[1], Z. Yu[1], M. Vernekar[1], V. N. Shah[2]; [1]Genomelink, Inc., Berkeley, CA, [2]Indiana Univ. Sch. of Med., Indianapolis, IN

Abstract:

Background: Type 1 Diabetes (T1D) affects approximately 2 million people in the United States and it's commonly underdiagnosed. Contrary to the earlier belief that T1D predominantly begins in childhood, recent studies have shown that nearly 50% of T1D is actually adult-onset, which leads to misdiagnosis as Type 2 Diabetes (T2D). Some studies showed that over 30% of adult T1D cases were initially being incorrectly diagnosed as T2D. This leads to inadequate treatment since T1D and T2D require different therapeutic approaches. The major challenge is that the current diagnostic methods rely on islet antibody tests (GAD, IA-2, IAA, ZnT8), which vary in sensitivity and often necessitate multiple tests; the process is not only costly but also time-consuming. There is a pressing need for more accessible, cost-effective, and accurate

diagnostic tests.
Objectives: The primary objective of this study is to develop an accurate genetic prediction model (Polygenic Risk Score - PRS) to distinguish T1D from T2D, which can be a promising tool for T1D diagnosis. Also, this study aims to establish cost-efficient DNA genotyping operations, including the logistics of DNA kit shipping, to facilitate the widespread adoption of this diagnostic method.
Methods: This study recruited 196 T1D cases and 187 T2D controls. DNA samples were collected using FDA-cleared at-home DNA kits. We analyzed genetic variants associated with T1D and developed a PRS model to differentiate T1D from T2D. The model's performance was evaluated using the area under the curve (AUC), and its diagnostic accuracy was measured in terms of sensitivity and specificity.
Results: The PRS model demonstrated high accuracy, with an AUC exceeding 0.85. The model also achieved better sensitivity and specificity, both over 80%, compared to current gold standards such as the GAD antibody test. Additionally, the feasibility of at-home DNA test operations was confirmed, showing that this method is practical and effective for large-scale diagnostic use.
Conclusion: The results of this study indicate that our at-home DNA test and PRS model provide a reliable and cost-efficient alternative to current diagnostic methods for distinguishing T1D from T2D. The at-home DNA test is much more accessible as it’s non-invasive and low-cost compared to the current islet antibody tests. This new approach has the potential to improve diagnostic accuracy, ensure timely and appropriate treatment, and reduce long-term healthcare costs associated with T1D misdiagnosis.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5034F: Discovering stimulatory state specific type 2 diabetes GWAS mechanisms with single-cell multi-omics on iPSC-derived fibro-adipogenic progenitors

Authors:

C. Ventresca[1], A. Varshney[1], P. Orchard[1], Y-C. Tsan[1], A. Monteiro da Rocha[1], M. Laakso[2], J. Tuomilehto[3], T. Lakka[2], K. Mohlke[4], M. Boehnke[1], L. Scott[5], H. Koistinen[3], F. Collins[6], T. Herron[1], S. Bielas[1], S. C. J. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Eastern Finland, Kuopio, Finland, [3]Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, [4]Univ North Carolina, Chapel Hill, NC, [5]Univ Michigan, Ann Arbor, MI, [6]NIH, Bethesda, MD

Abstract:

Variants and genes at type 2 diabetes (T2D) and related trait genome wide association study (GWAS) signals may impact T2D risk via skeletal muscle, a primary insulin-responsive tissue. We previously generated *in vivo* single-nucleus (sn-)multi-omics (RNA+ATAC) profiling of 286 skeletal muscle biopsies from the FUSION Tissue Biopsy Study and found that a subset of GWAS signals colocalized with cell-type specific e/caQTL in fibro-adipogenic progenitors (FAPs). In 39 of these individuals, we derived induced pluripotent stem cell (iPSC) lines from fibroblasts. To investigate mechanisms for the GWAS signals, we differentiated iPSC lines to FAPs, allowing us greater flexibility to dissect the underlying pathways using sn-multi-omics and CUT&Tag, a sensitive method to analyze genome-wide histone marks. We previously optimized both techniques on multiple cell types. We hypothesized that FAPs can be differentiated from iPSC lines to investigate the impact of stimulatory state specific FAP molecular mechanisms on T2D risk and T2D-related traits. Here, we demonstrate that FAPs can be derived from iPSC lines for deep molecular phenotyping within different environments. We have successfully differentiated iPSC lines from 39 donors into FAPs. These iPSC-derived FAPs displayed FAP morphology, expression of FAP marker genes, and loss of expression of pluripotency markers. After differentiation flow cytometry showed that the FAP marker *NT5E* was expressed in an average of 71% of cells across all lines with a standard deviation of 20 percentage points. In principal component analysis with bulk skeletal muscle cells types the iPSC-FAPs clustered with *in vivo* FAPs, and logistic regression analyses between cell type-specific chromatin accessibility peaks of the iPSC-FAPs and other skeletal muscle cell types revealed that the chromatin accessibility profile of iPSC-FAPs was very similar to *in vivo* FAPs (p-value = 6.221527e-252), indicating that the differentiated cells can be used as a model system to characterize FAPs. We are currently inducing high-glucose, high-insulin, and high-lipid states for the iPSC-FAPs and will be performing sn-multiomics on the iPSC-FAPs in the different stimulatory states. We will use this to investigate stimulatory state specific genetic regulatory effects (e/caQTL) within T2D pathways. We will validate our results by performing CRISPRi knockdown and CRISPRa activation on a subset of nominated loci.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5035F: Discovery of novel loci for blood pressure from genome-wide meta-analysis accounting for gene-alcohol interactions in African population cohorts

Authors:

M. Feitosa, K. Schwander, M. Province, on behalf of the CHARGE Gene-Lifestyle Interactions Working Group; 1Div. of Statistical Genomics, Washington Univ. Sch. of Med., St Louis, MO

Abstract:

Elevated blood pressure (BP) is a risk factor for cardiometabolic diseases influenced by genetic and environmental factors, including alcohol consumption. We investigated whether genome-wide gene-alcohol (GxA) interaction influences systolic BP (SBP), diastolic BP (DBP), and pulse pressure. We analyzed 4 dichotomous alcohol consumption variables: (i) current drinkers (yes/no), (ii) light versus never (LIGHTvsNEVER) drinkers, (iii) heavy versus never drinkers, and (iv) heavy versus light (HEAVYvsLIGHT) drinkers in up to 78,252 individuals in 19 African population cohorts. Firstly, we conducted genome-wide GxA interaction using a 1df freedom test and identified 5 loci ($P<5x10^{-8}$), including a novel *SH3BGRL2*. We detected the *SH3BGRL2*-alcohol interaction in HEAVYvsLIGHT drinkers for pulse pressure in ten cohorts (same β_{GxA} effect direction, N=30,862 individuals). *SH3BGRL2* locus harbors associations with smoking and obsessive-compulsive disorder. Secondly, we performed the 2df joint test of GxA interaction and main genetic effect and identified 16 loci for BP ($P<5x10^{-8}$). Fourteen loci of the 2df associations are suggested mainly due to a main genetic effect, while the novel *LOC124900602-SNCCA* and *CDIN1* loci were identified when the GxA interaction was accounted for with the main genetic effect. We found an association between *LOC124900602-SNCA* and SBP in LIGHTvsNEVER drinkers, which was previously associated with alcohol consumption. *CIDIN1* locus is associated with DBP in LIGHTvsNEVER drinkers and previously reported with alcohol use disorder, smoking, and major depressive disorder. Our findings demonstrate that gene-alcohol interaction and Human genetic diversity can contribute to gene discovery for BP and may provide insights into mechanisms by which alcohol consumption influences BP.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5036F: Dose Optimization of Probucol in Rescuing Clec16a Hypofunction Phenotype in KO Mice with Disrupted Mitophagy

Authors:

R. Pandey[1], B. Strenkowski[2], M. Matos[3], M. Bakay[4], H. Hakonarson[3]; [1]Childrens Hosp. of Philadelphia, PHILADELPHIA, PA, [2]Childrens Hosp. Of Philadelphia, Philadelphia, PA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Children Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Background: CLEC16A is linked to various autoimmune diseases. We generated an inducible whole-body knockout (KO), Clec16a$^{\Delta}$UBC mice to address the role of CLEC16A loss of function. CLEC16A KO mice exhibit disrupted mitophagy, a critical cellular process for maintaining

mitochondrial health and function. Published data from our inducible Clec16a KO mice suggests a derailed functional link between mitophagy, autophagy, lipophagy, ER stress, SOCS-induced JAK-STAT signaling and CLEC16A, predisposing to the immune dysregulation and the development of an inflammatory phenotype together with neurodegeneration. Here, we investigated Probucol, a lipid-lowering drug with antioxidant properties, for its potential to restore mitophagy and rescue the CLEC16A KO phenotype. **Methods:** We evaluated several doses of Probucol, ranging from 3 to 50 mg/kg, in a series of in vivo experiments using animal models with Clec16a loss of function. The efficacy of each dose was assessed by measuring key biomarkers related to immune response and metabolic function and rescue of KO phenotype. **Results:** The dose of 15 mg/kg Probucol was found to be optimal in rescuing the disrupted mitophagy in CLEC16A KO mice. This dose significantly restored the levels of mitophagy markers and improved mitochondrial membrane potential compared to both lower and higher doses. Additionally, the 15 mg/kg dose enhanced overall survival, indicating a successful restoration of mitochondrial function. **Conclusion:** The optimal dose of Probucol for rescuing the disrupted mitophagy phenotype in CLEC16A KO mice is 15 mg/kg. This finding highlights Probucol's potential as a therapeutic agent for treating autoimmune disorders associated with CLEC16A hypofunction and impaired mitophagy. Further research is necessary to explore the clinical relevance of these findings.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5037F: Elevated Prevalence of MODY Genetic Variants in Gestational Diabetes from the UK Biobank

Authors:

J. Lee[1], H. Lee[2], J. Choi[2], J. PARK[2], H. Yang[1], K. Park[1], S. Kwak[1]; [1]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [2]Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

Maturity-Onset Diabetes of the Young (MODY) exhibits a diverse phenotypic spectrum, leading to frequent misdiagnosis as other forms of diabetes, including Gestational Diabetes Mellitus (GDM). This study aimed to investigate the prevalence of MODY-related genetic variants, including *GCK*, *HNF1A*, and *HNF4A* in women diagnosed with GDM within the UK Biobank and assess the potential misdiagnosis of MODY during pregnancy. We analyzed whole exome sequence (WES) data from the UK Biobank, including the 200K interim release (n=200,643) and the latest release (n=469,835). The study population included women with GDM, women with type 2 diabetes (T2D) who had a history of childbirth, and a control group of women with a

history of childbirth but no diagnosis of GDM, type 1 diabetes, or type 2 diabetes. We compared the frequency of pathogenic or likely pathogenic monogenic diabetes variants using two variant sets: 1) variants identified by the ClinGen group and 2) variants specified by the Exeter group (Mirshahi et al. 2022). In the 200K interim WES dataset, 2.3% of the 522 women diagnosed with GDM had MODY variants according to the Exeter group criteria. Among women with T2D and those in the control group, the detection rates were 0.62% and 0.06%, respectively. In the latest WES dataset, which included 1,129 women with GDM, the detection rates were 1.86% for GDM, 0.67% for T2D, and 0.05% for the control group. Detection rates were consistently higher in women with GDM than those with T2D, with *GCK* being the most frequently identified gene in women with GDM. These findings suggest that approximately 2% of women diagnosed with GDM may actually have MODY. This underscores the need for improved diagnostic approaches to accurately identify and manage monogenic diabetes during pregnancy.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5038F: Environmentally-Driven gene expression variation influences NAFLD risk in Mexican Americans

Authors:

R. Perez[1], M. Almeida[2], V. Diego[2], S. Laston[3], T. Howard[4], J. Galan[1], E. MANUSOV[5], J. Curran[2], J. Blangero[6], S. Williams-Blangero[2]; [1]Univ. of Texas - Rio Grande Valley, Brownsville, TX, [2]Univ. of Texas Rio Grande Valley, Brownsville, TX, [3]Univ. of Texas Rio Grande Valley, South Padre Island, TX, [4]UTRGV Sch. of Med., Brownsville, TX, [5]UTRGV, brownsville, TX, [6]Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX

Abstract:

Non-alcoholic fatty liver disease (NAFLD) is prevalent (58%) among Mexican Americans in the Rio Grande Valley of South Texas. The Rio Grande Valley Family Study is a longitudinal epidemiological study of Mexican Americans focusing on the determinants of common complex diseases risk. For this study, we examine genetic and environmental factors influencing NAFLD progression. We aim to identify the environmental portions of the gene expression measures that are associated with liver steatosis. 525 participants were recruited from Mexican American families. Clinical data were collected, including biometric measurements and hepatic health assessments using Vibration-Controlled Elastography (FibroScan, Paris) to identify steatosis. Genome-wide RNA sequencing was employed to identify transcripts associated with NAFLD. To focus our examination on environmental sources of gene expression variation, we corrected transcript measures for inter-individual genetic variation using best linear unbiased prediction

of genetic values.
After genetic correction, *CD209* gene expression levels reached genome-wide significance (p =1.23x10-6), indicating a strong association with NAFLD. *CD209* is known for its role in immune response and inflammation. Additionally, *PNPLA3* (p = 2.73x10-05) was associated with NAFLD risk. This well-known NAFLD-influencing gene is linked to triglyceride metabolism and fibrosis in hepatocytes. *MAPRE3* (p = 4.57x10-05) levels were also associated with NAFLD. This gene is involved in microtubule stabilization and cellular processes, with emerging evidence suggesting a role in liver disease.
This study shows that environmentally driven variation in *CD209*, *PNPLA3*, and *MAPRE3* gene expression levels is associated with risk for NAFLD in Mexican Americans. *CD209*'s involvement in immune response highlights its potential as a biomarker for liver inflammation. *PNPLA3*'s significant role in lipid metabolism and fibrosis highlights its importance in NAFLD progression. *MAPRE3*'s emerging association with liver health also warrants further investigation. Given that we limit examination to only the malleable environmental component of gene expression variation, our findings have significant implications for targeted therapies and personalized medicine in this high-risk population.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5039F: Evaluating genetic and proteomic support for cytokine-targeting therapies of prevalent immunological diseases in the UK Biobank

Authors:

M. Luth, E. Bartell, D. Jakubosky, G. McInnes, B. Cajes, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Abstract:

Proinflammatory cytokines such as tumor necrosis factor (TNF) play a key role in the pathogenesis of many prevalent immunological diseases, such as rheumatoid arthritis (RA) and inflammatory bowel disease (IBD). Drugs targeting cytokines, either directly (e.g. TNF inhibitors such as adalimumab) or indirectly (e.g. JAK family inhibitors such as upadacitinib) are a major segment of approved and investigational therapies for these disorders. Large-scale biobanks comprising genetic data linked to rich phenotypic information, such as the UK Biobank, are a valuable resource for leveraging human genetics to predict the phenotypic effects of pharmacological modulation of therapeutic targets. The high-throughput proteomic data available in approximately 50,000 UK Biobank participants may offer complementary or

differentiated therapeutic insights, but the specific use-cases and degree to which proteomic data can aid in drug discovery and development have not been fully elucidated. To investigate the utility of both data types, we comprehensively evaluated the genetic and proteomic evidence supporting inhibition of cytokine and related targets of approved and investigational therapies for prevalent immunological diseases in the UK Biobank. Our findings reveal both shared and distinct associations between genetic variants in proinflammatory cytokine and related genes, circulating levels of the proteins they encode, and the immunological diseases for which cytokine-targeting and related therapies have been approved. For example, while TNF protein levels were robustly associated with numerous immunological indications for which TNF inhibitors have been approved, such as RA and Crohn's disease, nonsynonymous genetic variants in TNF did not share these associations. Conversely, nonsynonymous genetic variants in TYK2 were robustly associated with protection from a range of immunological diseases for which TYK2 inhibitors are either approved (e.g. psoriasis) or under investigation (e.g. IBD), but as an intracellular target, proteomic analyses of TYK2 were not possible. Notably, the absence of either genetic or proteomic robust associations with an immunological disease was predictive of clinical development failure; for instance, clinical development of the IL1B inhibitor canakinumab for RA was halted because its efficacy was inferior to TNF inhibitors. Overall, our analyses highlight the utility of leveraging both genetic and proteomic data when conducting analyses in support of drug discovery and development. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5040F: Evaluating kidney size as a marker for renal and hepatic health

Authors:

R. Roshani, L. Petty, A. Petty, C. Robinson-Cohen, J. Below; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Kidney disease affects 37 million Americans. Traditional markers for assessing kidney function, such as estimates of glomerular filtration rate (eGFR) and glomerular integrity (albuminuria), are often inadequate due to the complexity and variability of renal pathology. Evaluating kidney size offers a complementary perspective, as changes in kidney size can occur before significant alterations in kidney function. This study aims to understand how kidney size contributes to health and disease by evaluating its relationships with genetic factors, risk factors, and clinical

outcomes. We utilized phenotypic and genetic data from individuals with radiology-report or magnetic resonance imaging (MRI)-derived kidney size measurements within some of the largest biobanks in the world. Previous Genome-Wide Association Studies (GWAS) of kidney volume have identified 10 risk loci affecting eGFR, chronic kidney disease (CKD), body mass index (BMI), and type 2 diabetes, including loci near *UMOD* and *FTO*. To further investigate these associations, we performed Linkage disequilibrium score regression analysis (LDSC) on traits from European ancestry and kidney volume GWAS summary statistics of approximately 38,000 individuals from the UK Biobank (UKBB). LDSC quantifies the genetic correlation between traits by leveraging linkage disequilibrium patterns. Our analysis revealed substantial genetic correlations between body size, kidney function, and total kidney volume (P value < 10^{-8}). Additionally, we conducted stratified Phenome-Wide Association Studies (PheWAS) on approximately 43,000 male and female diabetic and non-diabetic individuals from Vanderbilt University Medical Center's biobank (BioVU). PheWAS involves adjusting for age and BMI to explore associations between kidney size and a wide range of clinical phenotypes and allows for the identification of phenotypic associations across various health outcomes with the kidney size. The PheWAS results demonstrated significant associations between kidney length and both subclinical and clinical cardiovascular outcomes, acute kidney injury, and liver diseases. These findings underscore the importance of kidney size as a relevant trait for kidney, liver and cardiovascular health. By integrating LDSC and PheWAS methodologies, this study highlights the genetic and phenotypic relationships involving kidney size, offering insights into its role in predicting health and disease outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5041F: Evaluatingthe impact of biological sex on ADHD presentation, prevalence, and genetic risk

Authors:

S. Guagliardo[1], M. Lundberg[2], A. Schork[2], M. Shuey[3], N. Cox[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Res. Inst. for Biological Psychiatry, Roskilde, Denmark, [3]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

Attention-deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder that impairs vigilance-attention, executive functioning, and motivation. Due to this, individuals with ADHD are at higher risk of addiction, poor academic and professional outcomes, and social deficits. The reported prevalence of ADHD is two to seven times higher in males than females. One

explanation is that “female sex” is protective against ADHD. Others argue, however, that the high heritability of ADHD (0.6-0.85) and similar reported genetic risk between sexes suggests that these trends may instead be due to heterogenous expression of symptoms in response to the environment (e.g. the female experience altering symptom manifestation). This aligns with the fact that men tend to be diagnosed with the more disruptive ADHD-hyperactive subtype, increasing likelihood of detection and decreasing age at diagnosis. Conversely, women are more often diagnosed with the more internalized ADHD-inattentive subtype, resulting in lower detection, later age of diagnosis, and increased chance of misdiagnosis. Using Vanderbilt University Medical Center’s (VUMC) biobank (n=3,285,882 electronic health records (EHR); n=119,750 genotyped samples (BioVU)) we identified trends in ADHD prevalence and genetic architecture between sexes. ADHD-associated ICD codes (n=38,419) were less frequent in EHR-recorded females relative to males (n=14,395 vs. 24,024) and the median age at first diagnosis was substantially older (21.72 years, IQR=20.96 vs. 15.05 years, IQR=9.1). In a genetically determined European-ancestry subset, ADHD polygenic risk scores (PRS) were a significant independent predictor of ADHD diagnosis, with stronger effects in females (males, beta= 13.47, p=0.03; females, beta=16.90, p=7.4E-7) with female cases having higher average PRS than male cases (p=0.04). In a sex-specific phenome-wide association study (PheWAS), ADHD PRS was associated with similar phenotypes regardless of sex, including substance/tobacco use, other psychiatric disorders, obesity, diabetes mellitus, and respiratory problems. Our findings that female patients with ADHD appear to have higher genetic liability despite lower rates of diagnosis are consistent with previous studies. ADHD PRS did not demonstrate differential comorbidity structures based on sex in this subset. Ongoing data acquisition will enhance our dataset, enabling more sex-specific analyses with stratification (race, decade of diagnosis, ADHD subtype) while preserving statistical power.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5042F: Examination of unreported *NOD2* variants in unsolved autoinflammatory cases

Authors:

C. Gill, J. Kemppainen, S. Sabrowsky, K. Sikkink Brown, K. Kolbert, L. Duncan, S. Kroc, E. Toering, A. Knudson, N. Modjeski, C. Vitek, D. Whalen, M. Basiaga, A. Orandi, M. Koster, J. Davis, K. Warrington, F. Vairo; Mayo Clinic, Rochester, MN

Abstract:

Background: Genetics plays a crucial role in autoinflammatory diseases, often driven by variants in genes regulating the innate immune system. Previous studies showed <30% diagnostic yield for suspected autoinflammatory diseases. *NOD2* variants are linked to Blau and Yao syndromes. Notably, common coding and non-coding *NOD2* variants and single nucleotide polymorphisms (SNPs) are more prevalent in patients with autoinflammatory symptoms, underscoring the need to understand *NOD2*'s role for better diagnosis and treatment. **Objective:** This study reviews *NOD2* SNP prevalence related to Yao syndrome in genome sequencing data from individuals with suspected autoinflammatory syndromes whose initial clinical sequencing was non-diagnostic. **Methods:** The Rheumatology Department at Mayo Clinic collaborated with the Center for Individualized Medicine to create a tailored genetic testing service for autoinflammatory diseases. Utilizing a novel Genetic Testing and Counseling (GTAC) model, individuals referred for autoinflammatory symptoms were evaluated. A custom genome panel targeting 120 genes was employed. Post-test genetic counseling and medical management were provided, with follow-up facilitated by the referring rheumatologist. **Results**: From September 2022 through March 2024, the GTAC team identified 73 patients (46 adults and 27 children, 51 female and 22 male) with non-diagnostic initial panel results. Among them, 56% experienced symptoms as adults, while 38% (28/73) had symptoms starting in childhood. Manual review of genome sequencing data revealed that 32% (23/73) had at least one *NOD2* variant not reported by initial clinical testing. Specifically, 30% (22/73) had the *NOD2* IVS8+158 variant, 8% (6/73) had the *NOD2* R702W variant, and 7% (5/73) had both variants. **Discussion:** Our cohort exhibited a higher rate of *NOD2* variants (32%) compared to the general population (5.1% for IVS8+158 and 1.4% for R702W). These findings suggest *NOD2* SNPs may be common in autoinflammatory patients. IL-1 and IL-6 antagonists like canakinumab and tocilizumab may benefit these patients, underscoring the relevance of reporting these variants. Genetic counseling and testing for autoinflammatory conditions are rare, highlighting the need for collaboration for improved patient management. **Conclusions:** Unraveling the genetic basis of autoinflammatory diseases enhances diagnostic precision and fosters personalized treatment. Our findings demonstrate the significance of *NOD2* SNPs in solving unexplained autoinflammatory cases, guiding personalized treatment, and recognizing initial clinical test report limitations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5043F: Excess of rare variants in MODY genes in T2D families from Endogamous Ethnic Groups in India

Authors:

M. Rout, D. Sanghera, On behalf of the investigators of the AIDHS/SDS and INDIGENIUS consortium; Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Abstract:

Type 2 diabetes (T2D) etiology is highly complex due to its multiple roots of origin. Polygenic risk scores (PRS) based on genome-wide association studies (GWAS) can only partially explain the T2D risk. Asian Indians have up to six times higher risk of developing T2D than Europeans. However, the underlying causes of this disparity are unknown. Here, we have performed targeted sequencing of ten T2D GWAS/candidate regions using endogamous Punjabi Sikh families and replication studies using unrelated Sikhs and families from three other Indian endogamous ethnic groups (EEGs). We detected several rare and ultra-rare variants (RVs) in *KCNJ11-ABCC8* and *HNF4A* (MODY genes) cosegregated with late-onset T2D. In addition, we identified RV enrichment in two novel genes, *SLC38A11* and *ANPEP*, associated with T2D. Gene-burden analysis revealed the highest RV burden contributed by *HNF4A* (p=0.0003), followed by *KCNJ11/ABCC8* (p=0.0061) and *SLC38A11* (p=0.03). Some RVs detected in Sikhs were also found in Agarwals from Jaipur, both from Northern India. Still, they were monomorphic in the other two EEGs from South India. Despite carrying a high burden of T2D and RVs, most members of Sikh families showed a low burden of PRS. These findings may uncover new rare forms of T2D and should reveal new therapeutic targets. Genetic testing in such families will benefit young, unaffected offspring to mitigate the T2D burden by early intervention or lifestyle changes.
Funding This study was supported by grants from the Indian Council of Medical Research [ICMR] (India) Project: No. 55/6/2/Indo-US/ 2014-NCD-II and the National Institute of Diabetes and Digestive and Kidney Diseases [NIDDK], US National Institutes of Health [NIH] grant R21 DK105913. The genotyping of AIDHS/SDS cohorts was also partly supported by NIDDK R01 DK082766, R01DK118427, Dr. Geoffrey Altshuler Children Hospital Foundation Endowment funds, and the Presbyterian Health Foundation grants.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5045F: Expression and Association study of microRNA 192 levels in an urban *Punjabi* sub-population group with Hypercholesterolemia

Authors:

M. Monu, G. Gandhi; Guru Nanak Dev Univ., Amritsar, India

Abstract:

Low density lipoprotein- cholesterol hypercholesterolemia by formation of atherosclerotic plaques can increase the risk for cardiovascular diseases. As changes in the profile of circulating microRNAs have been detected in diverse pathological conditions, miRNAs have potential as novel biomarkers for diagnosis and therapy. In this study, the expression levels of microRNA 192 were assessed in serum samples of 200 patients with hypercholesterolemia and compared with the levels in 200 healthy controls stratified by gender, lipid profile and obesity. Expression profiling was carried out by using the AriaMx Agilent RT-PCR by standard kits from Agilent in stringent conditions. The levels of miR192 (4.69X) were upregulated in patients as compared to controls ($p<0.001$). Levels were elevated in obese patients while female patients had significantly higher central obesity and miRNA expression levels. Patients on Atorvastatin treatment significantly had lowered levels of low density lipoprotein- cholesterol and atherogenic index. A significant association of gender was observed with xanthomas ($r=0.208$, $p=0.003$), miR192 ($r=0.151$, $p=0.032$). Low density lipoprotein- cholesterol levels, age-at-onset of disease and obesity parameters were also significantly associated with miR192 expression. Body mass index and waist to height ratio were significant predictors of miR192. In conclusion, as increased microRNA expression levels showed association with obesity and low density lipoprotein- cholesterol levels, microRNAs may evolve as non-invasive biomarkers for detection and prognosis of cardiovascular diseases at an early age.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5046F: Expression of *APOE4* is greater than *APOE3* but differs between Ancestries.

Authors:

S. Moura[1], K. Celis[1], L. Bertholim Nasciben[1], F. Rajabli[1], J. Rivero[1], K. Hamilton-Nelson[1], P. Whitehead[1], M. Gearing[2], D. Bennett[3], S. Weintraub[4], C. Geula[4], W. Scott[1], D. Davis[1], R. Vontell[1], T. Schuck[5], D. Dykxhoorn[1], M. Pericak-Vance[1], A. Griswold[1], J. Young[1], J. Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Emory Univ., Atkanta, GA, [3]Rush Univ., Chicago, IL, [4]Northwestern Univ., Chicago, IL, [5]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: *APOE4* is the strongest genetic risk factor for Alzheimer's disease (AD), the most common form of dementia, but different populations have differing risk. Specifically, Non-Hispanic White *APOE4* carriers have a higher risk of developing AD than African American *APOE4* carriers. The local ancestry (LA) surrounding the *APOE4* region was previously shown to be the primary factor in this risk difference, with *APOE4* carriers of European LA (ELA)

having increased chromatin accessibility and expression of *APOE4* compared to individuals of African LA (ALA). We investigated if the *APOE* LA associates with differential expression when within the *APOE3* allele. Differences between alleles could provide insight into ancestry-specific regulatory areas controlling *APOE4* expression. **Methods:** We identified AD autopsy samples by genotyping (GSA), all homozygous for *APOE3* and LA (ALA or ELA) from 4 ADRC brain banks: Miami, Emory, Duke, and Rush. We performed single nuclei RNA sequencing (snRNA-seq) on eight (4 ELA and 4 ALA) frozen frontal cortex (BA9) samples using 10x Genomics. Data analysis was carried out using Seurat and differential expression analyses were performed using MAST. **Results:** We integrated our *APOE3* data with previously published data from our group (Griswold, A *et al,* 2021) on *APOE4* carriers of the same ancestries. A total of 145,551 nuclei from 25 brains (14 ELA and 11 ALA) were analyzed. We identified 35 distinct cell clusters at a resolution of 0.4. The proportion of cells per cluster between ALA and ELA samples was similar for all clusters. Our data shows that contrary to our previous observations in *APOE4* carriers, *APOE3* carriers of ALA express higher *APOE* in astrocytes (cluster 2) and Microglia (cluster 6) than ELA carriers. Further, when comparing *APOE3* vs *APOE4* carriers for the same LA, we observe that *APOE3* have significantly lower gene expression than *APOE4* alleles. **Conclusions:** Our data suggest the LA region surrounding *APOE* has an *APOE* allele-specific regulatory effect on *APOE* expression. Whether this specificity is driven by the *APOE* allele itself or by varying regulatory elements present in the allele haplotypes remains unknown. Further, within the same LA, we observed that the different cell types express less *APOE* in *APOE3* carriers compared to *APOE4* carriers. Given that greater expression of *APOE4* is associated with increased risk, the lower expression of *APOE3* may contribute to *APOE3* lower risk. Altogether, this may provide additional insights into the regulatory mechanisms affecting expression of the different *APOE* alleles which can be useful for therapeutic efforts.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5047F: Functional dissection of heart enhancers implicated in inherited cardiomyopathies

Authors:

Z. Weber[1], T. O. Monroe[2], C. Holgren[2], R. Mitchell[1], E. M. McNally[3], M. A. Nobrega[1]; [1]Univ. of Chicago, Chicago, IL, [2]Northwestern Univ., Chicago, IL, [3]Northwestern Univ, Chicago, IL

Abstract:

Purpose: Inherited cardiomyopathies are considered Mendelian disorders, driven by highly penetrant, rare coding variants, albeit with variable clinical penetrance and expressivity. Recently, data from biobank-scale genetic cohorts have suggested a more oligogenic trait liability underlying the expressivity of these diseases. Additionally, novel contributing loci have been implicated in cardiomyopathy GWAS, however, most of these loci have not been experimentally validated, and many of the signals map within noncoding regions, creating challenges for interpretation and mechanistic follow-up studies.
Methods: To better characterize the genetic architecture underlying this oligogenic architecture, we employed human induced pluripotent stem cell (hIPSC)-derived cardiomyocytes in engineered heart tissues (EHTs), an organoid model of myocardium consisting of differentiated cardiomyocytes and cardiac fibroblasts in a 3D collagen-based scaffold.
Results: From these EHTs, we generated genome-wide maps of open chromatin regions (OCRs), comprising more than 400,000 peaks, resolved to the major cell type fractions. These open chromatin regions correlate well with publicly available OCRs from adult left ventricle tissue. Additionally, single-cell RNA-seq of EHTs shows well defined cardiomyocyte, fibroblast, myofibroblast, endothelial, and neuronal-like populations, recapitulating major species in myocardial tissue. Leveraging these OCRs, we fine-mapped GWAS summary statistics for inherited cardiomyopathies and related measures of heart function such as left ventricle ejection fraction (LVEF). From these data, we prioritized a locus at chr3p25.1 harboring genetic signal in both dilated cardiomyopathy and LVEF, with 33 genome-wide significant variants in LD over an intergenic region upstream of the taurine-transporter *SLC6A6*, and downstream of the developmental factor *WNT7A*. Using haplotype-resolved reporter assays, we observed activity consistent with distal enhancer sequences, and significant allele-specific effects on the activity of several overlapping OCRs. We also generated deletion mutants in hIPSCs for the same region to study enhancer activity on the neighboring genes and evaluate EHT contractile function.
Conclusions: These combined genomics and experimental approaches allow us to evaluate the noncoding, modifying contributions to inherited cardiomyopathies and heart function, and build a more complete picture of heart failure liability.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5048F: Gene Discovery and Biological Insights into Anxiety Disorders from a Multi-Ancestry Genome-wide Association Study of >1.2 Million Participants ⭐

Authors:

E. Friligkou[1], S. Løkhammer[2], B. Cabrera-Mendoza[1], J. SHEN[3], J. He[4], G. Deianna[5], M. Zanoaga[6], Z. Asgel[7], A. Pilcher[1], L. Di Lascio[8], A. Makharashvili[1], D. Koller[9], D. Tylee[10], G. Pathak[11], R. Polimanti[11]; [1]Dept. of Psychiatry, Yale Sch. of Med., New Haven, CT, [2]Dept. of Clinical Sci., Univ. of Bergen, Bergen, Norway, [3]Yale Sch. of Med., New Haven, CT, [4]Yale Univ. school of medicine, New Haven, CT, [5]Ctr. for NeuroSci., Pharmacology Unit, Sch. of Pharmacy, Univ. of Camerino, Camerino, Italy, [6]Dept. of Computational Biology, Univ. of Lausanne, Lausanne, Switzerland, [7]Dept. of Child and Adolescent Psychiatry, NYU Langone Hlth., New York Metropolitan Area, NY, [8]IRCCS Istituto Clinico Humanitas, Milan, Italy, [9]Yale Sch. of Med., West Haven, CT, [10]Yale Univ., New Haven, CT, [11]Yale Univ., West Haven, CT

Abstract:

Background: The genetics of anxiety are not well understood, particularly across population groups.
Methods: We conducted genome-wide association studies (GWASs) of anxiety disorders in All of Us (AoU) for individuals of European (EUR), African (AFR), Admixed-American (AMR), East Asian (EAS), and South Asian ancestries, and meta-analyzed the results with FinnGen, iPsych, the Million Veteran Program (MVP), the PGC, and the UK Biobank (UKB), reaching a sample size of 1.2 million. A common genetic factor (ANX) was constructed from EUR cohorts with genomic structural equation modeling. Due to a smaller sample size, other populations were meta-analyzed with METAL. PRS-cs estimated polygenic risk in AoU. Multi-tissue transcriptome-wide association studies were conducted using S-MultiXcan. Dorsolateral prefrontal cortex proteome-wide associations of ANX were sought with FUSION. The number of shared causal variants and local genetic correlations with other psychiatric traits were identified using MixeR and LAVA respectively. Genetic correlations of ANX were identified with LD-score regression among all 11,175 FinnGen, MVP, and UKB phenotypes, and the causality of the significant associations was explored with the Latent Causal Variable method.
Results: Ancestry-specific and cross-ancestry GWASs identified 51 loci, 39 of which are novel. The most significant novel association was rs6689226 (beta=0.01, $p=4.8x10^{-12}$) in EUR and rs2510682 (z=5.92, $p=3.31x10^{-9}$) across all ancestries. The EUR-derived polygenic risk was associated with anxiety (fifth vs. first quintile of PRS distribution OR; 95%CI) in AFR (1.23; 1.12-1.36), AMR (1.55; 1.39-1.74), and EAS (1.69; 1.16-2.49). ANX heritability was enriched in the limbic system, cerebral cortex, cerebellum, metencephalon, entorhinal cortex, and brain stem. Of the 115 transcriptome- and proteome-wide significant genes, *CTNND1, CNNM2, RAB27B, KHK, BTN2A1,* and *NEK4* showed both transcriptomic and proteomic associations. MDD shared 11.3K causal loci with ANX (rg=90%), SCZ shared 9.3 (rg=45%), and BIP 7.7 (rg=48%). ANX was genetically correlated with ADHD, BIP, MDD, PTSD, and SCZ on 47 loci with locus 11q23.2 demonstrating the strongest genetic correlation with MDD (rho=0.97, $p=2x10^{-6}$) and the highest number of cross-disorder associations (N=7). Among the top 100 phenome-wide genetic

correlations, 77 traits were psychiatric, 10 were gastrointestinal, and 4 were pain-related phenotypes. 24 traits were causal for ANX, including 3 psychiatric and 5 cardiovascular conditions.
Discussion: This study expands our understanding of genetic risk and pathogenesis of anxiety disorders in worldwide populations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5049F: Gene dosage architecture of Autism and neurodevelopmental related traits

Authors:

C. Poulain[1,2], V. Bourque[1,2], P. Bergeret[3], S. Kazem[4,5], J-L. Martineau[3], Z. Saci[2], L. Almasy[6], D. Glahn[7], G. Huguet[8], S. Jacquemont[9]; [1]Montréal Univ., Montreal, QC, Canada, [2]CHU Sainte Justine, Montréal, QC, Canada, [3]CHU Sainte Justine, Montreal, QC, Canada, [4]CHU Sainte-Justine, Montreal, QC, Canada, [5]Montréal Univ., Montréal, QC, Canada, [6]Univ. of Pennsylvania, Philadelphia, PA, [7]Boston Children s Hosp. & Harvard, Boston, MA, [8]CHU Ste justine, Montréal, QC, Canada, [9]CHU Sainte justine Res. Ctr., Montreal, QC, Canada

Abstract:

Background: Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder (NDD) characterized by atypical social communication, specific interests and repetitive behaviors. Individuals with ASD often experience co-occurring comorbidities such as intellectual disability (ID), hyperactivity or sleep disorder. Copy Number Variants are routinely screened for in the neurodevelopmental clinics and their contributions to ASD liability are well-documented. **Knowledge gap**: The pleiotropic gene-dosage effects on ASD and comorbid traits have not yet been systematically quantified. **Aim** : To characterize gene dosage architecture of autism and related phenotypes. **Methodology**: We conducted CNV-GWAS using a dataset aggregating 9 cohorts (~ 500k individuals, 37k ASD). We tested CNV associations with ASD, 17 phenotypes among autism patients (eg. learning disability, language delay) , and 18 phenotypes in the general population. **Results**: We identified 6 new loci associated with ASD. Approximately 30% of the known and novel associations remain significant with similar effect sizes after adjusting for the effects of CNVs on general cognitive ability (IQ). We identified significant associations in all except 6 of the traits investigated. The genomic covariance across the 35 traits reveals several clusters correlated with ASD including early developmental milestones, cognition, mood, anxiety, irritability and loneliness.Deletion and duplication effects sizes were negatively correlated across most traits but not for ASD suggesting that any deviation for normal

gene dosage increases liability for this condition. Based on one of the largest association studies to date, we provide a comprehensive characterization of the gene dosage architecture of ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5050F: Gene prioritization analyses on the multi-ancestry Neuroticism GWAS point to novel risk loci.

Authors:

A. Topaloudi, M. Kaka, G. Chen, P. Paschou; Purdue Univ., West Lafayette, IN

Abstract:

Neuroticism is a personality trait characterized by negative emotions and is a risk factor for psychiatric disorders such as anxiety, depression, and schizophrenia. So far, several genetic studies on Neuroticism identified multiple risk loci, however, the majority of them are conducted on European samples. Here, we are exploring the genes involved in Neuroticism, using the results of our multi-ethnic Genome-Wide Association Study (GWAS) that includes individuals from European, African, South and East Asian ancestries. For this study, we initially performed a gene-based analysis using MAGMA with a multi-ancestry LD-reference where we detected 127 significant associations. This was followed by the gene-based analysis incorporating Hi-C data from the adult brain (h-MAGMA) that identified 238 significant genes. Next, we performed a Transcriptome-Wide Association Study (TWAS) through S-MultiXcan to combine the information across the 49 GTEx tissues we tested, where we detected 159 genes whose transcript expression is significantly associated with the trait. We further explored the causally associated genes with Neuroticism by implementing summary data-based Mendelian Randomization (SMR), as well as the causal tissues through Tissue co-regulation score regression (TCSC). Lastly, using the prioritized genes, we conducted gene-set analyses to shed light on the enriched pathways. Overall, the post-GWAS analyses we performed on the multi-ethnic GWAS summary statistics of Neuroticism revealed multiple gene associations, including new ones not present in the previous European GWASs.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5051F: Genetic analysis of variable number of tandem repeats (VNTR) in the Japanese population

Authors:

X. Liu[1], A. Wada[1], K. Tomizuka[1], K. Hikino[1], Y. Ishikawa[1], K. Matsuda[2], BioBank Japan, R. Mukamel[3], P-R. Loh[4], C. Terao[1,5,6]; [1]RIKEN IMS, Yokohama, Japan, [2]Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan, [3]Harvard Med. Sch., Boston, MA, [4]Brigham and Women s Hosp., Boston, MA, [5]Univ. of Shizuoka, Shizuoka, Japan, [6]Shizuoka Gen. Hosp., Shizuoka, Japan

Abstract:

Variable number tandem repeats (VNTRs) have been linked to various diseases and disorders but had been largely excluded from genome-wide association studies (GWAS) due to their sequence complexity and the challenges associated with imputation. Recent advancements in computational methodologies, alongside analyses of the UK Biobank (UKB) data, have identified VNTRs significantly associated with common human phenotypes. Some of these are among the strongest genetic associations observed to date. However, previous studies have predominantly focused on European populations. In this study, we estimated VNTR in the Japanese population by analyzing 16K putative VNTRs in 4,000 Japanese whole genome sequences (WGS) and subsequently imputing them into 180K samples from the Biobank Japan (BBJ) to examine their association with 234 phenotypes. We identified 581 genome-wide significant associations, with 22 VNTRs implicated as causal variants through fine-mapping. We replicated previously identified VNTRs in the UKB for corresponding phenotypes and observed unreported new phenotypic associations. In particular, we observed causal associations for a well-documented VNTR in *MUC1* with multiple phenotypes, including blood urea nitrogen, gastric cancer, uric acid levels, and estimated glomerular filtration rate (eGFR). Notably, the association with gastric cancer (OR = 2.5, P = 7.5E-50) is the most significant when compared to single nucleotide variants (SNVs) from GWAS on the genome-wide scale. Additionally, we identified novel VNTR associations such as *SLC12A7* with mean cell hemoglobin concentration (MCHC). In summary, our findings not only corroborate but also extend previous VNTR analyses in European cohorts, offering new insights into the genetic variability of VNTRs and their impact on human traits.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5052F: Genetic architecture of white matter hyperintensity volume in diverse Hispanic/Latino adults

Authors:

M. Fornage[1], T. Sofer[2], C. R. Isasi[3], R. B. Lipton[3], A. Stickel[4], W. Tarraf[5], H. M. González[6], C. DeCarli[7]; [1]Univ. of Texas Hlth.Sci. Ctr. at Houst, Houston, TX, [2]Beth Israel Deaconess Med. Ctr. / Harvard Med. Sch., Boston, MA, [3]Albert Einstein Coll. of Med., Bronx, NY, [4]San Diego State Univ.,

San Diego, CA, [5]Wayne State Univ., Detroit, MI, [6]Univ. of California San Diego, San Diego, CA, [7]Univ. of California Davis, Sacramento, CA

Abstract:

Cerebral white matter hyperintensities (WMH) on magnetic resonance imaging are part of the spectrum of brain vascular injury accompanying aging and are associated with an increased risk of stroke and dementia. WMH burden is greater among Hispanic and African American than non-Hispanics white (NHW) adults. WMH is highly heritable and genome-wide association studies (GWAS) have identified several genetic loci. The largest GWAS published to date included 50,970 older community-dwelling individuals of European and African ancestry. However, the genetic architecture of WMH in Hispanic/Latino adults remains largely uncharacterized.We conducted GWAS and admixture mapping analyses of WMH volume in a sample of 2159 diverse Hispanic and Latino adults (mean age: 62.4 years; 66% women). We investigated associations at 27 previously-identified WMH loci in our Hispanic/Latino cohort. To identify additional loci for WMH, we meta-analyzed our GWAS results with those of the largest GWAS published to date.The proportion of variance in WMH volume attributable to common and low frequency variants in our sample was 28.9%, similar to a previous estimate in NHW individuals (29%). No association reached genome-wide significance. The strongest association was for rs60690281 ($P=9.3x10^{-8}$). A significant correlation between effect estimates in our and the published GWAS for variants mapping within 300kb of the lead variant was observed at 18 of the 27 previously reported loci. A weighted polygenic risk score constructed from the top SNPs at these 27 loci was significantly associated with WMH burden in our cohort (P<1x10-4; variance explained = 0.85%). Admixture mapping identified a novel locus on chr14q13.2, where counts of European ancestry at that locus were significantly associated with WMH volume ($P=4.9x10^{-7}$). Multi-ancestry meta-analysis combining our GWAS results with those from GWAS of WMH in populations of European and African ancestry identified an additional locus on chr12q22 (rs10859915, $P=1.8x10^{-8}$).Our study suggests that WMH loci previously identified in NHW individuals are relevant to Hispanic / Latino adults and identifies novel loci influencing WMH burden. The locus on ch14q13.2 identified by admixture mapping, spans an 800 kb region containing RALGAPA1, encoding a Ral GTPase activating protein with known impact on neuronal function and brain development; and INSM2, encoding a transcriptional repressor influencing glucose metabolism. The locus on chr12q22, identified by GWAS meta-analysis, is tagged by rs10859915, an expression QTL of AMDHD1. Variants at this locus have been associated with serum levels of Vitamin D and histidine.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5053F: Genetic architecture of white matter revealed by Deep unsupervised learning derived endophenotypes from DTI fractional anisotropy brain images in UK Biobank

Authors:

Z. Xie, W. He, K. Patel, M. Fornage, D. Zhi; The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract:

White matter is critical for facilitating communication among various brain regions, playing a key role in supporting cognitive function. Genetic studies on white matter are vital to understand brain functions and potentially enhance our insights into neurological conditions such as multiple sclerosis and dementia. Previous studies have applied unsupervised learning to extract phenotypes from fundus images and T1/T2-FLAIR brain MRIs, which can be used for genome-wide association analyses. Extending this research, we trained an autoencoder to reconstruct brain fractional anisotropy (FA) from diffusion tensor imaging (DTI) data. Earlier research has demonstrated that this approach enables the extraction of unbiased and comprehensive descriptors of the input data, and importantly, FA has been shown to be a crucial metric of the white matter tracts in DTI images and a predictor of several brain-related diseases. The autoencoder was trained on a subset of 4,597 subjects to compress the linearly registered FA map of DTI into a compact, low-dimensional representation, and then reconstruct the FA map from this representation. We trained autoencoders with bottleneck dimensions ranging from 64 to 512 and then conducted GWAS on these latent representations using a linear model with 17,440 UK Biobank subjects. The covariates included age, sex, 10 principal components, head position, scanner table position, assessment center and attending date. We used a Bonferroni-corrected p-value threshold of 5e-8, divided by the number of phenotype dimensions (i.e., 64, 128, 256, 512), to determine the significance of the associations. We clumped the significant SNPs into loci using FUMA and found that the number of significant loci generally increases with the dimension of the representation. The autoencoder model with 512 bottleneck dimensions identified 29 loci, all of which were previously associated with white matter microstructure across different studies, as found by querying the GWAS Catalog. Many white matter related genes were identified, including EPHA3, GMNC, CENPW, AMZ1, GNA12. The results suggest that the autoencoder model can be used to extract meaningful phenotypes from DTI data and identify genetic associations with these phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5054F: Genetic association analysis of 24-hour rest-activity rhythm

Authors:

H. Wang[1], T. Faquih[1], P. Nagarajan[1], R. Noordam[2], M. K. Rutter[3], S. Redline[1], R. Saxena[4], Q. Xiao[5]; [1]Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [2]Leiden Univ. Med. Ctr., Leiden, Netherlands, [3]Univ. of Manchester, Manchester, United Kingdom, [4]Mass Gen. Hosp., Harvard Med. Sch., Boston, MA, [5]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract:

Rest-activity rhythm, an indicator of circadian rhythm influenced by behavioral and social-environmental factors, is associated with multiple health outcomes. However, the genetic basis for rest-activity rhythms is unclear. We conducted genome-wide common variant and exome sequencing analyses on fifteen parametric and nonparametric 24-hour rest-activity rhythm metrics derived from 7-day accelerometry data in 85,784 individuals of European ancestry from the UK Biobank. Common variant analysis identified 15 genome-wide significant loci ($p<5e-8$), with three loci (*MESIS1, BTBD9, RGS16*) surpassing a more stringent significance threshold ($p<3.3e-9$). The mapped genes are associated with various sleep, activity, cardiometabolic, and psychiatric traits, and have functions in sleep-wake control, neuronal, and metabolic pathways. Genetic correlation showed shared genetic background between more favorable rest-activity rhythms (higher waketime activity-nighttime rest amplitude and more robustness patterns) with morningness chronotype and better cardiometabolic, cognition, and psychiatric conditions. Exome-wide association analysis of 24-hour rest-activity rhythm metrices is ongoing. Together, our findings will provide novel insight into the molecular pathways of rest-activity rhythms and their links to health conditions.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5055F: Genetic Determinants of Lung function Decline: A Multi-Level Analysis of Gene Expression

Authors:

Z. Elhusseini[1], O. Rafique[1], M. Hyung Ryu[1], P. Castaldi[2], I. Ruczinski[3], C. Hersh[4]; [1]Brigham and Women Hosp., Boston, MA, [2]Brigham and Women s Hosp., Cambridge, MA, [3]Johns Hopkins Univ., Baltimore, MD, [4]brigham and women's Hosp., Boston, MA

Abstract:

Background: Chronic obstructive pulmonary disease (COPD) is a progressive respiratory condition characterized by airflow limitation and decline in lung function, often assessed by forced expiratory volume in one second (FEV_1). Understanding the genetic basis of FEV_1 decline is crucial for unraveling the pathophysiology of COPD and developing effective therapies. We hypothesized that gene expression patterns in inflammatory pathways would be associated with FEV_1 decline in COPD. We investigated this hypothesis using whole blood RNA-sequencing data from 5 and 10-year follow-up visits in the COPDGene Study.
Methods: We employed a series of linear regression models (N=435) to assess the association between gene expression and FEV_1, adjusted for age, sex, race, smoking history and white blood cell percentages, in different levels;
1. Cross sectional level:
• **Predictor:** Gene expression levels at 5-year and 10-year visits separately
• **Outcome:** FEV1 measurements at the respective time points
2. Longitudinal phenotype level:
• **Predictor:** Gene expression levels at the 5-year visit
• **Outcome:** Change in FEV1 between 5 and 10-year visits within individuals
3. Total Longitudinal level:
• **Predictor:** Change in gene expression levels between 5 and 10-year visits
• **Outcome:** Change in FEV1 between 5 and 10-year visits within individuals
Furthermore, we generated an FEV1 decline gene signature by analyzing the RNA-seq data from the 5-years visit on the longitudinal phenotype level (Model 2) from three-time intervals (baseline to 5-year N=3850, 5-year to 10-year N=2043, and baseline to 10-year N=2035).
Results: Our analyses revealed distinct gene expression results for each approach, highlighting the importance of considering different perspectives when studying FEV_1 decline. After functional annotation, we found that the genes from the total longitudinal study (Model 3) are enriched in pathway terms such as inflammasome and leukotriene, which are known to be relevant for lung function decline and the development of COPD. The FEV1 decline gene signature showed associations with COPD progression, and rapid FEV_1 decline.
Conclusions: Overall, our findings contribute to a better understanding of the genetic determinants of respiratory health. Functional annotation revealed enrichment of genes involved in inflammasome and leukotriene pathways, known to be critical for lung function. The gene signature showed associations with COPD development, and rapid FEV_1 decline. These findings offer valuable insights that can inform both population-wide public health strategies and individualized clinical interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5056F: Genetic diversity & epidemiology of human adenovirus among hospitalized children with severe acute respiratory tract infection in Guangzhou, China

Authors:

H. Li, Z. Deng; Jinan Univ., Guangzhou, China

Abstract:

Human adenoviruses (HAdVs) can cause infections at any age but most commonly in pediatric population, especially in young children and infants. Adenoviruses can cause many symptoms similar to common cold, including rhinorrhea, fever, cough, and sore throat. Lower respiratory infections such as bronchitis, bronchiolitis, and pneumonia can be severe and even fatal. HAdVs are nonenveloped particles containing a double-stranded linear DNA genome of approximately 34,000-36,000 base pairs, which belong to the family *Adenoviridae* and the genus *Mastadenovirus*. Now HAdVs are grouped into species A-G. To explore the epidemiological characteristics of adenoviruses in children with acute respiratory tract infection(ARTI)in Guangzhou,China,from 2018 to 2020, the nasopharyngeal(NP)swabs of hospitalized children with ARTI collected from Guangdong Panyu Maternity and Children's Health Care Hospital were tested by quantitative polymerase chain reactions to measure adenovirus load. Clinical isolates were obtained by inoculating into cell culture,their growth characteristics studied,and the clinical characteristics of infected children documented. The percentage of HAdV ▪ positive samples was 4.80% in 500 NP(24/500)swabs collected. The susceptible population was children under 5 years of age. Phylogenetic analyses showed that the main genotypes were B ▪ species HAdV ▪ 3(40.34%)and HAdV▪7(15.27%). Twenty▪one cell cultures with sample inoculation were amplified and found to be positive for adenovirus DNA. Of all the genotypes obtained,HAdV▪7 had the most copies of the viral genome. Analyses of the clinical characteristics of infected children showed that HAdV ▪ 3 was prone to increase the number of neutrophils and lymphocytopenia. Our study suggested that HAdV ▪ 3 and HAdV ▪ 7 were the predominant genotypes in hospitalized children with an ARTI in Guangzhou. Viral genes had highly conserved evolution. Among all the clinical isolates, HAdV▪7 has high growth rate. Our findings expand knowledge of HAdV infections' clinical spectrum.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5057F: Genetic Modifiers of Motion Sickness: Whole Genome Sequencing Analysis of Individuals with a Significant History of Motion Sickness Participating in Clinical Studies

Authors:

V. Polymeropoulos, S. Smieszek, H. Bai, C. Xiao, C. M. Polymeropoulos, G. Birznieks, M. H. Polymeropoulos; Vanda Pharmaceuticals, Washington, DC

Abstract:

Motion sickness is a prevalent disorder affecting up to 30% of the general population during normal travel conditions. There are several known risk factors of motion sickness, including a familial history of the disorder. However, our understanding of its genetic background remains limited. We conducted randomized, double-blind, placebo-controlled clinical studies exploring the genetic background of motion sickness through whole genome sequencing. We focused on the variants previously reported in Hromatka et al (2015), the largest GWAS thus far on imputed microarray data. Association analysis was conducted on PC clustered EUR, including 598 cases and 1901 controls to match the original design. The strongest variant replicated is rs10970305 (OR=0.69, p-value=3.52e-6) with consistent direction of effect, an intergenic variant in vicinity of *ACO1* - a gene involved in iron homeostasis. Rs10970305 is a common variant with a global MAF of 0.41 (gnomAD), and from our analysis it is located within a cluster of significant variants. Our analysis also replicated effects from other variants including *SDK1* (OR=1.34, p-value=1.25e-3), *GPR26* (OR=1.27, p-value=4.36e-3), *NR2F2* (OR=0.83, p-value=0.03), and *NLGN1* (OR=0.76, p-value=0.03). As *SDK1* signal was replicating across several variants, we focused on this gene in subset analyses. Interestingly, rs34912216 in *SDK1* was strongly associated with severity of motion sickness symptoms. This difference was statistically significant and more pronounced in homozygous individuals. Additionally, we investigated the association with average nausea as captured by the nausea severity scale. This effect was also significant for *SDK1* variants, especially in a model corrected for age, sex, BMI, PC1, PC2, and ARM. Sidekick-1, encoded by *SDK1*, is a cell adhesion molecule that localizes to synapses - the largest among IgSF membrane proteins. Prior reports have shown that *SDK1* helps to locate lamina neurons, arrange them into columns, and sort photoreceptor axons into lamina cartridges, thereby establishing correct visual motion detection circuits. Based on prior GWAS, *SDK1* polymorphism is implicated in autism spectrum disorders, attention-deficit hyperactivity disorder and motion sickness. We believe further understanding of the genetic architecture of this complex phenotype will allow for in-depth understanding of the pathophysiology ultimately leading to improved treatments.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5058F: Genetic relationships and biobehavioral pathways between suicidal behaviors and comorbid mental disorders

Authors:

Y. Huang[1,2], S. Tao[1], Z. Meng[3], Y. Liu[1], A. M. Levin[3], B. K. Ahmedani[4,5], Q. Wang[1], H. Gui[4,5,6]; [1]Mental Hlth.Ctr. and Psychiatric Lab., State Key Lab. of Biotherapy, West China Hosp. of Sichuan Univ., Chengdu, China, [2]Broad Inst., stanley center, Cambridge, MA, [3]Dept. of Publ. Hlth.Sci., Henry Ford Hlth., Detroit, MI, [4]Behavioral Hlth.Services and Psychiatry Res., Henry Ford Hlth., Detroit, MI, [5]Ctr. for Hlth.Policy and Hlth.Services Res., Henry Ford Hlth., Detroit, MI, [6]Dept. of Psychiatry, Michigan State Univ., East Lansing, MI

Abstract:

Background: The close relationships between mental disorders and suicidality are frequently seen in epidemiology. However, the shared genetic liabilities underlying these comorbidities remain unclear. Here we aimed to investigate the genetic associations between multiple mental disorders and different levels of suicidality.
Methods: We used genome-wide association study (GWAS) summary statistics to estimate global and regional genetic correlation, and to identify pleiotropic genes or gene sets using a combination of statistical genetics tools. Summary statistics of phenotypes clinically associated with suicidality were accessed from Psychiatric Genomics Consortium (major depressive disorder [MDD], bipolar disorder [BD], anxiety disorders [ANX], obsessive-compulsive disorder [OCD], anorexia nervosa [AN], autism spectrum disorder [ASD], attention deficit hyperactivity disorder [ADHD], schizophrenia, cannabis use disorder, and post-traumatic stress disorder), and Million Veterans Program (alcohol use disorder [AUD] and opioid use disorder), and suicidality from their joint analysis. Then we verified polygenic relationships between 12 mental disorders and gradient scales of suicidality (ideation, attempt and death) using structural equation models by UK Biobank (UKB) raw data.
Results: We observed positive genome-wide genetic and local correlations between 12 mental disorders and suicidality (rg ranging from 0·25 [29 loci] to 0·68 [347 loci], $p<0.001$). Across pairs of suicidality and other mental disorders, we identified 32 functional genes (including 26 novel ones to suicidality like BPTF, NOL11 and CACNG5) shared by 6 or more pairs, and enriched in two Gene Ontology (GO) sets: developmental process and regulation of biological process, and 2 protein-protein interaction networks. We also identified unique genes within each pair enriched in distinct GO sets (e.g., glutamatergic synapse for suicidality-MDD, negative regulation of biological process for suicidality-BD, and actin cytoskeleton for suicidality-AUD). For UKB, eight out 12 mental disorders (MDD, BD, ANX, AUD, OCD, AN, ASD and ADHD) showed

significant correlation with suicidality (OR =4.76 at phenotypic level and OR=1.07 at polygenic level, p<0.001). Discussion: This study underscores the urgent need to address the genetic architecture of suicide behaviors, among individuals with multiple mental disorders. The combination of disease-ascertained GWAS data and population-level biobanks are warranted to enhance our understanding of this complex phenomenon. Our findings will guide future suicide prevention among individuals with major mental disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5059F: Genetic risk stratification in the Leuven familial pulmonary fibrosis cohort

Authors:

N. Veyt[1], T. Goos[1], M. Vermant[1], S. Gogaert[2], E. Van Hoof[3], A. Corveleyn[3], I. Cleynen[2], S. Vermeer[1], W. A. Wuyts[1]; [1]UZ Leuven/KU Leuven, Leuven, Belgium, [2]KU Leuven, Leuven, Belgium, [3]UZ Leuven, Leuven, Belgium

Abstract:

Background: Familial pulmonary fibrosis (FPF), defined as the presence of fibrotic interstitial lung disease (ILD) in at least two first or second degree relatives, makes up about 20% of all idiopathic ILDs. Both rare and common variants contribute to the risk of FPF, but true genetic risk stratification accounting for both types of variants is lacking. **Objective**: We aim to offer accurate risk prediction for family members of FPF patients and personalize the screening approach. **Methods**: We assembled a cohort of 337 individuals: 140 patients and 197 relatives. Patients diagnosed with idiopathic pulmonary fibrosis (IPF) and a positive family history (at least one additional first degree relative) were asked to participate, and their first degree relatives were recruited. Gene panel analysis and telomere length assessment was performed for FPF patients. Relatives took part in a multimodal screening program (pulmonary function test, high-resolution CT-scan, genetic testing and telomere length assessment using Flow-FISH). Furthermore, the cohort was genotyped using the Infinium Global Screening Array (Illumina) to perform polygenic risk score (PRS) analysis. **Results**: The Leuven FPF cohort consists of 115 families, with a median number of included individuals per family of 2 (range: 1-18). Gene panel analysis was negative in 69/112 families, detected a variant of unknown significance in 24/112 families and found a (likely) pathogenic (LP/P) variant in 19/112 families. LP/P variants were mostly detected in telomere-related genes (*TERT*, *TERC*, *RTEL1*, *PARN*), in 1 family we found a LP variant in a surfactant-related gene (*SFTPC*). In about 13% of relatives we observed signs of preclinical ILD on CT-scan as part of the screening program. Strikingly, in 3 families where a

variant in a telomere-related gene was found, (preclinical) ILD was seen in non-carriers. **Future directions**: Here we present the baseline genetic characteristics of the Leuven FPF cohort. Further research within this cohort will assess the explanatory value of common variants to the risk of FPF by PRS analysis.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5060F: Genetic study of Crohn's disease in the US population

Authors:

A. Asabere[1], K. Schiabor Barrett[1], M. Levy[1], N. Telis[1], N. Washington[1], D. Stoller[2], C. N. Chapman[3], A. A. Chahal[4], D. P. Judge[5], D. A. Olson[6], J. J. Grzymski[7], E. T. Cirulli[1], A. Bolze[1]; [1]Helix, San Mateo, CA, [2]Univ. of Nebraska Med. Ctr., Omaha, NE, [3]St. Luke's Univ. Hlth.Network, Bethlehem, PA, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Med. Univ. of South Carolina, Charleston, SC, [6]Hlth.Partners, Minneapolis, MN, [7]Univ. of Nevada, Reno, Sch. of Med., Reno, NV

Abstract:

Background: Understanding the genetics of Crohn's disease allows us to identify mechanisms of pathogenesis and potential drug targets. Genetics is also used to identify shared mechanisms between multiple chronic disorders.
Objective: Identify new genetic associations with Crohn's disease. Assess whether variants and genes associated with Crohn's disease in our cohort associate with other disorders, and whether there is evidence of genetic pleiotropy. Test effectiveness of a polygenic risk score model to predict who is at risk of Crohn's disease in our cohort.
Methods: We sequenced and analyzed a cohort of 80,000 adult participants from six health systems in the US, including 452 patients diagnosed with Crohn's disease. Importantly, these cohorts were not part of other efforts to identify genetic associations with inflammatory bowel disease. Clinical information from electronic health records included diagnoses codes, drug prescriptions as well as laboratory test results. We performed 3 types of genetic analyses: GWAS, rare-variant gene collapsing tests and HLA-WAS.
Results: As a control, we first checked the main associations found in large studies from the UK Biobank and the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). In our cohorts, we replicated the known associations with Crohn's disease at the following loci: *NOD2*, *IL23R*, *ATG16L1* and *OCTN1*. Moreover, we showed that an IBD polygenic risk score model based on external summary statistics could help identify a subpopulation in our cohort at higher risk of Crohn's disease, as well as a subpopulation at lower risk of Crohn's disease.
Conclusion: Our genetic study of Crohn's disease in an unselected cohort from 6 health systems

provides an independent and additional dataset to test and validate genetic associations with Crohn's disease.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5061F: Genetic subtypes of type 2 diabetes are distinguished through the lens of abdominal MRI ⭐

Authors:

E. Sorokin[1], M. Cule[1], M. Thanaj[2], N. Basty[2], B. Whitcher[2], E. Thomas[2], J. Bell[2], H. Yaghootkar[3]; [1]Calico Life Sci. LLC, South San Francisco, CA, [2]Res. Ctr. for Optimal Hlth., Sch. of Life Sci., Univ. of Westminster, London, United Kingdom, [3]Coll. of Hlth.and Sci., Univ. of Lincoln, Lincoln, United Kingdom

Abstract:

Recent genetic studies on type 2 diabetes identified eight distinct subtypes of the disease: three linked to insulin deficiency and five to insulin resistance (Suzuki et al, 2024). These subtypes elucidate diverse pathophysiological processes that may influence clinical presentation, including disease severity and risk of complications. We aimed to differentiate these genetic subtypes using magnetic resonance imaging (MRI). We extracted 21 MRI-derived phenotypes at scale using deep learning in the UK Biobank imaging cohort (n = 43,491). These phenotypes reflected fat depots (visceral, subcutaneous, liver, pancreas, thigh, paraspinal muscle, and bone marrow), pancreas volume, muscle size and quality, and fatty acid compositions in visceral and abdominal subcutaneous fat. We created a polygenic risk score for each type 2 diabetes subtype and tested them against each phenotype. Insulin deficiency subtypes had a similar pattern of association with MRI-derived phenotypes, with a few differences. The insulin deficiency without proinsulin subtype was associated with smaller pancreas volume (multiple test corrected p value=2.1e-2), whereas the insulin deficiency with proinsulin subtype was not. The residual glycemic subtype resembled the other insulin deficiency subtypes, except for associations with higher muscle indices. Insulin deficiency subtypes were associated with higher saturated fatty acids and lower monounsaturated fatty acids in subcutaneous fat. By contrast, the insulin resistance subtypes displayed distinctive body composition profiles. Body fat and obesity subtypes were associated with larger fat depots in all compartments, but body fat was associated with lower muscle indices while obesity was associated with higher muscle indices. The metabolic syndrome subtype was associated with higher fat in liver (p<2e-8), pancreas (7e-4), bone marrow (4e-5) and visceral adipose tissue (7e-4). The lipodystrophy subtype was associated with lower subcutaneous fat and higher liver fat (p<1e-14). The liver/lipid

metabolism subtype was associated with increased liver fat (p<1e-50). The insulin resistance subtypes were associated with higher monounsaturated fatty acids and lower saturated fatty acids in visceral adipose tissue, except the lipodystrophy subtype, which showed a pattern of association similar to insulin deficiency subtypes. These differential associations provide a more nuanced characterization for type 2 diabetes subtypes beyond anthropometric traits and highlight the potential of precision phenotyping using MRI to enhance our understanding of type 2 diabetes heterogeneity.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5062F: Genetic Underpinnings of Coronary Heart Disease in Hispanic/Latino Populations

Authors:

Y. Tu[1], G. Chittoor[2], A. Justice[2], Z. Wang[3], A. Pereira[4], E. Frankel[5], J. Below[6], K. North[7], M. Graff[8], L. Fernandez-Rhodes[9]; [1]penn state Univ., UNIVERSITY PARK, PA, [2]Geisinger, Danville, PA, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Harvard Med. Sch., Boston, MA, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Vanderbilt Univ Med Ctr., Nashville, TN, [7]Univ North Carolina, Chapel Hill, NC, [8]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [9]Penn State Univ., State College, PA

Abstract:

Hispanic/Latino (HL) populations are highly under-represented in studies of coronary heart disease (CHD) risk factors in many electronic health record (EHR)-based biobanks and cohorts, even though they are the largest ethnic minority in the US. To ensure their representation in genomic studies of CHD, we assembled over 130,000 Hispanic/Latino adults with genome-wide array genotyping or sequencing and CHD phenotype information from more than ten studies. CHD cases were defined by at least two occurrences of specific ICD-9/10 conditions or the presence of revascularization or cardiothoracic surgery in EHRs from biobanks including BioME, All of Us, and MyCode Community Health Initiative, or study-specific indicators in cohort studies, including Hispanic Community Health Study/Study of Latinos (HCHS/SOL), and Million Veteran Program (MVP). Data included 133345 HL adults (≥18 years; 51.3% female) including 19308 cases and 114037 controls. Genome-wide association results were adjusted by sex, age, and principal components (PCs), and filtered (MAF > 0.01, r^2 > 0.7). Meta-analysis identified 6 loci that reached genome-wide significance (GWS; p<5e-8), one of which is a novel locus on chromosome 18 (*rs75679416*). It alters several binding motifs and is located upstream of a long non-coding RNA related to *SALL3*. The other five loci reported previously include *rs55730499* in

gene *LPA* and *rs9349379* in gene *PHACTR1*. These two genes have been associated with CHD in non-Hispanic/Latino populations. Additionally, we observed 76 loci suggestively significant (p<5e-6), of which 53 have not been reported previously. We are adding additional samples from Latin America and the US, seeking independent replication, and applying fine mapping and conditional analyses to further inform the etiology of CHD in Hispanic/Latino adults. With the largest sample of Hispanic/Latino adults to date, we have generalized and identified potentially novel CHD loci, and future work in this vein will bolster cardiovascular health equity and HL community-based interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5063F: Genetic variants associated with diurnal intraocular pressure fluctuation in primary open-angle glaucoma

Authors:

F. Mabuchi[1], N. Tanaka-Mabuchi[2], Y. Sakurada[1], S. Yoneyama[1], Z. Yamagata[3], K. Kashiwagi[1], Japan Glaucoma Society Omics Group (JGS-OG); [1]Dept. of Ophthalmology, Univ. of Yamanashi, Yamanashi, Japan, [2]Dept. of Rheumatology, Univ. of Yamanashi, Yamanashi, Japan, [3]Dept. of Hlth. Sci.s, Univ. of Yamanashi, Yamanashi, Japan

Abstract:

Purpose: To investigate the genetic variants associated with diurnal intraocular pressure (IOP) fluctuation in primary open-angle glaucoma (POAG).
Methods: One hundred seventy patients with POAG were genotyped for 26 genetic variants predisposing to POAG. Diurnal variation of IOP was measured with a Goldmann applanation tonometry in a sitting position at 3 hour intervals for 24 hours. To find out the genetic variants associated with a diurnal IOP fluctuation (peak-trough), a multiple linear regression analysis was performed with the diurnal IOP fluctuation as a dependent variable and age, sex, and the genotyped variants as independent variables.
Results: The mean age of the enrolled subjects was 61.7±13.1 (mean±standard deviation) years, with 74 males and 96 females. The mean of diurnal IOP fluctuation was 5.8±2.2 mmHg. There was a significant association (Beta=0.92, 95% confidence interval=0.24-1.61, P=0.0081) between the effect allele of rs11710139 (near gene: *LOC107986141/LOC107986142*) and diurnal IOP fluctuation. A significant association (Beta=0.79, 95% confidence interval=0.080-1.49, P=0.029) was also found between the effect allele of rs6732795 (*ANTXR1*) and diurnal IOP fluctuation. As the number of effect allele of these genetic variants increased, the diurnal IOP fluctuation increased.

Conclusion: These results indicate that genetic variants associated with diurnal IOP fluctuation may exist in POAG.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5064F: Genetic variants in known genes in diverse Frontotemporal Dementia cohort, enriched for Caribbean Hispanic patients

Authors:

T. Gu[1], A. Martinez[1], M. Illanes-Manrique[2], P. Mena[1], H. Acosta[3], C. Camargo[1], X. Sun[1], M. Marrero[1], B. Baumel[1], M. Cornejo-Olivas[2], S. Castro-Suarez[2], J. Vance[1], M. Cuccaro[1], M. Pericak-Vance[1], K. Nuytemans[1]; [1]Univ. of Miami, Miami, FL, [2]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [3]Carribean Ctr. for the study of memory and cognition, San Juan, Brazil

Abstract:

Background: Most biomedical data currently available for frontotemporal dementia (FTD) is derived from studies in non-Hispanic white (NHW) populations. To increase representation in biomedical research, we set out to enroll and characterize a diverse FTD patient cohort enriched for Caribbean Hispanic patients. Method: Our current cohort consists of 135 FTD patients (71% Hispanic, 21% NHW, 6.5% Black/African American), with continuing enrollment from University of Miami, Caribbean Center for the Study of Memory and Cognition (PR) and Instituto Nacional de Ciencias Neurologicas (Peru), and samples from non-NHW individuals deposited from ADC centers and ALLFTD (NCRAD). All patients were evaluated using NACC Uniform DataSet or equivalent in their preferred language. We generated Illumina GDA+Neurobooster genotyping data and whole genome sequencing to query variants in the known FTD genes. Results: Initial genetic analyses showed none of the Hispanic patients are carriers of the C9orf72 repeat expansion or reported pathogenic variants in GRN. We observed one carrier of a known pathogenic variant (Hispanic patient) and 4 carriers of putative pathogenic variants (all Hispanic patients) in MAPT, as well as two carriers of novel variants in TBK1 (Hispanic and NHW patient) as well as three carriers of (putative) pathogenic VCP variants (all Hispanic patients). Conclusion: Genetic analyses of FTD in underrepresented population groups is necessary as information from research in NHW is not always generalizable across race/ethnicity. Studies to determine genetic ancestry of identified variants are currently ongoing. The biomedical characterization of FTD across race/ethnicity will help the understanding of disease mechanisms in all patients ultimately preventing further health disparities.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5065F: Genetically-informed drug repurposing analysis identifies hydralazine as a therapeutic candidate for uterine fibroids

Authors:

J. Kim[1], N. K. Khankari[2], J. A. DeCorte[1], J. N. Hellwege[2], D. R. Velez Edwards[2], T. L. Edwards[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Uterine leiomyomata, or fibroids, are the most common benign tumors affecting people with a uterus and have a cumulative incidence of up to 80% by age 50. Currently, hysterectomy is the only definitive treatment for fibroids. Other surgical treatments like myomectomy and uterine artery embolization have multiple contraindications and high recurrence rates (15 to 33%). Additionally, current medications show modest and unclear long-term efficacy. We explored novel, alternative treatments for fibroids by employing drug repurposing methods. We used summary statistics from a large-scale multi-ancestry fibroid genome-wide association study (GWAS) of 74,294 cases and 465,810 controls to conduct predicted gene expression analyses with S-PrediXcan and COLOC. We identified 81 genes that colocalized and were significantly associated through predicted gene expression with fibroid risk. We queried multiple drug-gene interaction databases to identify 67 approved medications that could potentially target these genes to reduce fibroid risk. One candidate gene-drug pair was *tumor protein P53* (*TP53,* gene) and hydralazine (drug). Hydralazine is an anti-hypertensive that induces TP53 activity. We used independent multi-ancestry GWAS summary statistics for systolic blood pressure (SBP) and conducted S-PrediXcan (GTEx v8 tissues). Predicted gene expression of *TP53* was associated with SBP (p-value < 0.05) in thyroid, skin (sun-exposed lower leg and not sun-exposed suprapubic), and minor salivary gland tissue. These tissue-specific gene expression summary statistics were used in a two-sample Mendelian randomization (MR) analysis and were harmonized to reflect hydralazine's effect of increased *TP53* expression. We utilized the fibroid S-PrediXcan *TP53* gene expression summary statistics in the same tissues for the outcome. The inverse variance weighted MR results indicated that hydralazine use, as proxied by increased *TP53* expression, may also reduce fibroid risk by 42% (odds ratio = 0.58, 95% confidence interval 0.49 - 0.67, p-value=1.43×10^{-12}). Leveraging drug repurposing methodology, we show that hydralazine may be a drug candidate to investigate further for reducing fibroid risk. Hypertension is a known fibroids risk factor, and repurposing a medication to target both could provide an efficient therapeutic approach for targeting these comorbid conditions.

Additional studies will be required to verify efficacy, establish dosages, and identify mechanisms of action.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5066F: Genome wide assessment of homozygosity in Parkinson's disease across diverse ancestral populations

Authors:

K. Step[1], C. F. H. Astudillo[2], E. Eltaraifee[3], Z-H. Fang[4], A. J. Hernández-Medrano[5], P-J. Kung[6], M. Ostrožovičová[7], A. Zirra[8], S. Bandres-Ciga[9]; [1]Stellenbosch Univ., Cape Town, South Africa, [2]Univ. del Desarrollo, Santiago, Chile, [3]Univ. of Khartoum, IEND, Khartoum, Sudan, [4]The German center for Neurodegenerative diseases, Tuebingen, Germany, [5]Natl. Inst. of Neurology and Neurosurgery, Mexico City, Mexico, [6]Natl. Taiwan Univ. and Academia Sinica, Taipei, Taiwan, [7]P.J. Safarik Univ., Kosice, Slovakia, [8]Queen Mary Univ. of London, London, United Kingdom, [9]NIH, Bethesda, MD

Abstract:

Background: Long regions of consecutive homozygous genotypes, known as runs of homozygosity (ROHs), result from inbreeding and recessive inheritance. Despite arising from common ancestors, ROHs are found in outbred populations, with potential links to complex phenotypes through recessive patterns of inheritance.
Objective: Our objective is to undertake the most extensive screening of the impact of homozygosity in Parkinson's disease (PD) etiology across diverse populations. We aim to investigate to what extent recessive allelic architecture might define a portion of PD heritability by exploring excess of homozygosity in cases relative to controls in a global context. Here we estimated the degree of inbreeding across diverse populations, and we prioritized regions with potential recessive inheritance where rare variants may be influencing the genetic architecture of disease.
Methods: We explored ROHs leveraging genotyping imputed data from the sixth release of the Global Parkinson's Genetics Program (GP2) across nine ancestral populations, comprising 16,599 PD cases and 13,585 healthy controls. Whole-genome sequencing (WGS) data from GP2 release 6 included 1,956 individuals, comprising 1,647 PD cases and 309 healthy controls. The parameters investigated included the total length of ROHs, the count of ROHs, the average length of ROHs, and inbreeding coefficient estimates. Additionally, we intersected ROHs within PD recessive known genes as well as PD genes within known risk loci and assessed differences in cases versus controls.

Results: Here, we provide a research framework to look for ROHs in PD in a large scale and unbiased manner. Our results showed that PD cases harbor enriched homozygosity, suggesting that unknown and novel recessive effects may explain a proportion of PD missing heritability. **Conclusions:** The identification of autosomal recessive genes for PD to date has been mostly based on the identification of specific consanguineous families in European populations. Yet, the genetic architecture of disease differs across populations. The present work expands our current knowledge of the genetic aetiology of PD in a global context. Moreover, a comprehensive expansion of homozygosity analysis to include WGS data promises to provide a more profound insight into the intricate genetics of PD.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5067F: Genome-wide admixture and association studies of allergic rhinitis in African American population

Authors:

Y. Gautam, T. Mersha; Cincinnati Children s Hosp. Med. Ctr., Mason, OH 45040, USA, OH

Abstract:

Allergic rhinitis (AR) is an inflammatory respiratory disorder mediated by the Immunoglobulin E (IgE)-mediated response to exposure to allergens. To elucidate the genetic risk of AR in African American (AA) population, we genotyped 890 AR cases and 868 non-AR controls from AA population using Multi-Ethnic Genotyping Array (MEGA). Genotype imputation was carried out with the CAAPA reference panel using the Michigan Imputation Server. Global and local ancestry inference was carried out using RFMix v2. Admixture mapping (AM), genome-wide association (GWAS) and Bayesian Mixed (BMIX) analysis were performed to detect ancestry, genotype and joint ancestry-genotype effects on AR, respectively. Fine mapping and functional annotation of AM loci were conducted to identify putative casual variants with ancestry effect. After standard quality control filtering, over 6,000,000 variants were tested by logistic regression adjusted for sex, age, and global ancestry. AM identified 13 significant (2 African and 11 European ancestry specific loci (with p-value < 4.03E-5). GWAS identified 9 loci associated to AR with p-value < 1E-5, among which, an intronic variant rs141896208 (OR = 0.277, Gene = *RP11-143N13.1*) at 16p13.2 reached the genome-wide significance level P-value = 2.35E-8. Three variants, rs138959862 (*HNF1A-AS1*), rs74030425 (*SLCO3A1*), and rs138111094 (*ARHGAP44*) are African ancestry specific. Joint analysis of ancestry-genotype effects replicated the GWAS variant rs141896208. Fine mapping of AM loci using statistical and functional annotations prioritized SNPs rs133376 (*NAGA—FAM109B*), rs13057991 (*SREBF2*), and rs133328 (*WBP2NL*) in 22q13.2,

rs1893592 (*UBASH3A*) and rs9981388 (*PDE9A*) in 21q22.3, and rs201767 (*DLEU1*) in 13q14.3 as the potential AR risk variants. To our knowledge, this is the first comprehensive genetic study of AR that showed both African and European ancestry contribute towards AR-susceptibility, highlighting the importance of ancestry-based analysis in mixed ancestry populations including African Americans.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5068F: Genome-wide association analysis of blood pressure traits in up to 79,402 Hispanic/Latino participants ⭐

Authors:

X. Zhang[1], S. Okello[1], H-H. Chen[2,3], W. Zhu[3], Z. Wang[4], M. Stanislawski[5], L. Fernández-Rhodes[6], A. Justice[7], G. Chittoor[7], E. Blanco[8,9,10], R. Burrows[11], C. Albala[11], J. L. Santos[12], B. Angel[11,13], B. Lozoff[14], S. Gahagan[10], J. Gaziano[15,16,17,18], K. Cho[15,16,17,18,19], J. McCormick[20], S. Fisher-Hoch[20], C. Kooperberg[21], U. Peters[21], S. Buyske[22], T. Sofer[23], A. Mottl[24], K. L. Young[1], M. Graff[1], H. M. Highland[1], J. E. Below[3], K. E. North[1]; [1]Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Academia Sinica, Taipei, Taiwan, [3]Vanderbilt Genetics Inst. and Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Dept. of BioMed. Informatics, Sch. of Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [6]Dept. of Biobehavioral Hlth., Coll. of Hlth.and Human Dev., Pennsylvania State Univ., University Park, PA, [7]Population Hlth.Sci., Geisinger, Danville, PA, [8]Centro de Investigación en Sociedad y Salud, Univ. Mayor, Santiago, Chile, [9]Núcleo de Socio Med., Santiago, Chile, [10]Div. of Academic Gen. Pediatrics, Child Dev. and Community Hlth.at the Ctr. for Community Hlth., Univ. of California at San Diego, San Diego, CA, [11]Dept. of Publ. Hlth.Nutrition, Inst. of Nutrition and Food Technology (INTA), Univ. of Chile, Santiago, Chile, [12]Dept. of Nutrition, Diabetes and Metabolism, Sch. of Med., Pontificia Univ. Católica de Chile, Santiago, Chile, [13]Faculty of Hlth.Care Sci., San Sebastián Univ., Santiago, Chile, [14]Dept. of Pediatrics, Univ. of Michigan, Ann Arbor, MI, [15]VA Boston Hlth.care System, Boston, MA, [16]Million Veteran Program (MVP) Coordinating Ctr., VA Boston, Boston, MA, [17]Div. of Aging, Dept. of Med., Brigham and Women's Hosp., Boston, MA, [18]Dept. of Med., Harvard Med. Sch., Boston, MA, [19]Centralized Interactive Phenomics Resource (CIPHER), Office of Res. and Dev., Veterans Hlth.Admin., Washington, D.C., DC, [20]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Sch. of Publ. Hlth., Brownsville, TX, [21]Div. of Publ. Hlth.Sci., Fred Hutchinson Cancer Ctr., Seattle, WA, [22]Dept. of Statistics, Sch. of Arts & Sci., Rutgers Univ., Piscataway, NJ, [23]Beth Israel Deaconess Med. Ctr. / Harvard Med. Sch., Boston, MA, [24]UNC Kidney Ctr., UNC Sch. of Med., The Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Genome-wide association analysis of blood pressure traits in up to 79,402 Hispanic/Latino participants

Background: Hypertension is a major risk factor for cardiovascular disease (CVD), with graded associations reported between higher systolic blood pressure (SBP) and diastolic blood pressure (DBP) and increased CVD risk. Hispanic/Latino (H/L) populations are the fastest growing group of ethnicities in the United States (U.S.), projected to reach 100M by 2050. Between 2015 and 2018, 45.6% of the H/L adults in the U.S. had hypertension. Genome-wide association studies (GWAS) on BP have been dominated by discoveries made in populations of European descent, leading to poor translation of findings to other populations and jeopardizing downstream applications such as polygenic risk score development and other precision medicine advances. This study addresses these gaps by examining the genetic architecture of BP exclusively in H/L individuals.

Methods: We performed GWAS on five BP traits (SBP, DBP, pulse pressure, mean arterial pressure, and hypertension) in up to 29,810 H/L participants from the Cameron County Hispanic Cohort (CCHC), the Population Architecture using Genomics and Epidemiology (PAGE) Study, and Santiago Longitudinal Study (SLS). Using linear regression models, we adopted a two-step approach: first adjusting for age, sex, and BMI, then rank-normalizing the residuals, and finally adjusting for genetic principal components. We performed inverse-variance weighted meta-analysis using METAL, incorporating summary statistics from Hispanic participants in the Million Veteran Program (MVP) Release 4, totaling up to 79,402 participants.

Results: We identified 21 loci associated with at least one BP trait at $P < 5e\text{-}9$, including four novel loci associated with SBP. Newly identified genes include *GRM4*, *PXT1*, *SRSF3*, *MFSD13A*, *PCD11*, *CALHM2*, *HOXB8*, *HOXB-AS4*, *RGS9*, and *AXIN2*. Specifically, some of these newly identified genes are implicated in ocular hypertension, glomerular filtration rate, coronary heart disease, and cation channel activity. Additionally, we replicated multiple previously reported associations with BP reported in populations other than H/L.

Conclusions: Our results demonstrate the importance of ancestrally diverse populations in the identification of novel loci, highlighting the need for investment of non-European participants in genomic research. Collectively, our findings expand our current understanding of genetic susceptibility underlying hypertension, potentially leading to actionable therapeutic targets with biological relevance.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5069F: Genome-wide association study for abnormal forms of Covid-19

Authors:

L. Zha, B. Borsari, Y. Liu, G. Wang, J. Liu, M. Gerstein; Yale Univ., New Haven, CT

Abstract:

Long COVID is a debilitating, multisystem illness that occurs in at least 10% of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections. Previous studies have performed multiple genome wide association studies (GWAS) to identify significant loci associated with COVID-19 critical illness, hospitalization and SARS-CoV-2 reported infection. We performed a genome-wide association study with data from the UK Biobank to determine genetic determinants among long, severe, recovered COVID-19 patients, as well as those with mild infections, and healthy controls. We identified gene PARK7 for the comparison between long covid (LC) + severe covid (SC) vs. recovered covid (RC), gene APOC1 for [LC+SC] vs mild covid and SC vs. mild covid, and gene ARL17B for SC vs. RC. We also identified rs540702737 to be the lead variant of locus chr17:43463493-44865498 for the SC vs. RC GWAS. The locus contains several genes including KANSL1 previously implicated in Covid19 symptoms measurements and covid-19 critical illness/hospitalization vs population/mild infection. Two random forest models were performed and the first only includes the covariates (age, sex, BMI, smoking status, social economic status for England, 10 principal components), while the second includes all the covariates and the suggestive loci (p value < 10^-5). There is notable better model performance to distinguish patients with outlier forms of COVID (LC/SC) from the mildly infected patients compared to recovered patients. However, the genetic biomarkers identified by the GWAS are useful when discriminating against LC/SC patients against RC patients. These results indicate that there is some shared genetic architecture between abnormal (i.e., long and severe forms) forms of Covid19, with three significant loci identified from GWAS for abnormal COVID-19. Genetic biomarkers also demonstrate a better improvement in distinguishing (LC+SC) from mild/healthy/RC than demographic covariates alone.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5071F: Genome-wide association study meta-analysis reveals novel loci associated with polyps of the female genital tract

Authors:

N. Ingold[1], E. Moro[1,2], T. O'Mara[1]; [1]QIMR Berghofer, Brisbane, Australia, [2]Queensland Univ. of Technology, Brisbane, Australia

Abstract:

Polyps of the female genital tract are benign overgrowths of blood vessels, stroma and glands that can occur in women of all ages. Genital polyps have been associated with abnormal uterine bleeding; however, the majority (up to 82%) of genital polyps are believed to be asymptomatic. It has been estimated that half of all women will have genital polyps at some point in their life, and of these polyps, up to 13% may become malignant. Research surrounding the common genetic drivers of female genital polyps and the extent to which these polyps cause female reproductive cancers, including endometrial cancer, remains understudied.
We performed a genome-wide association study (GWAS) for female genital polyps using UK Biobank data (ICD10 N84 : cases = 13,386, controls = 223,939). We meta-analysed the genetic effect estimates of this GWAS with GWAS summary statistics for female genital polyps from FinnGen (freeze 10) and the Michigan Genomics Initiative (freeze 3) resulting in a total sample size of 30,219 cases and 360,522 controls, the largest such study of female genital polyps to date. After removing single nucleotide polymorphisms with an allele frequency less than 1%, our GWAS meta-analysis identified 16 genome-wide significant ($P < 5\times10\text{-}8$) independent genetic loci associated with risk of female genital polyps. The most statistically significant locus was found on chromosome 3, close to gene *EEFSEC* which has previously been implicated in endometrial cancer risk. We also identified several loci harboring genes associated with estrogen signaling, including *ESR1* and *TRPS1*, an estrogen receptor-associated transcription repressor. We are currently performing downstream analyses, such as Mendelian randomisation, to further our understanding of the genetic underpinnings of female genital polyps, its relationship to cancer development and opportunities for targeted treatment of endometrial cancer.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5072F: Genome-wide association study of childhood dental caries in a multi-ethnic sample.

Authors:

K. K. Ryan[1], J. R. Shaffer[1,2], N. Mukhopadhyay[2], E. Yanca[3], G. Eckert[4], J. Ren[4], P. F. Castelluccio[4], B. P. Katz[4], A. Hara[5], M. A. Keels[6], S. Levy[7,8], M. L. Marazita[2,1,9], M. Fontana[3]; [1]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Biology, Sch. of Dental Med., Univ. of Pittsburgh,

Pittsburgh, PA, [3]Dept. of Cariology, Restorative Sci. and Endodontics, Sch. of Dentistry, Univ. of Michigan, Ann Arbor, MI, [4]Dept. of Biostatistics and Hlth.Data Sci., Sch. of Med. and Richard M. Fairbanks Sch. of Publ. Hlth., Indiana Univ., Indianapolis, IN, [5]Dept. of Cariology and Operative Dentistry, Indiana Univ., Indianapolis, IN, [6]Dept. of Pediatrics, Sch. of Med., Duke Univ., Durham, NC, [7]Dept. of Preventive & Community Dentistry, Coll. of Dentistry and Dental Clinics, Univ. of Iowa, Iowa City, IA, [8]Dept. of Epidemiology, Coll. of Pubic Hlth., Univ. of Iowa, Iowa City, IA, [9]Clinical and Translational Sci., Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Early childhood dental caries is a complex disease involving both genetic and non-genetic risk factors. Despite its high prevalence, there have been few genome-wide studies specifically of early childhood caries, leaving gaps in our understanding of its genetic architecture. In this pilot genome-wide association study (GWAS), we assessed early childhood dental caries in a multiethnic sample (14% Hispanic, 44% Black, 41% White, and 2% Other) of 543 children recruited from Indiana, Iowa, and North Carolina. Dental caries was assessed at ages 4±0.5 years using the International Caries Detection and Assessment System (ICDAS) criteria to generate a semi-quantitative phenotype representing the number of teeth showing evidence of treated or untreated decay (ICDAS code of 2 or more) or missing (i.e., extracted) due to decay. Participants were genotyped using the Illumina Global Diversity Array and imputed to the TOPMed consortium reference panel (version 2). Genetic associations were tested using linear models while adjusting for sex, self-reported ethnicity, self-reported race, and the first eight principal components of ancestry in PLINK. We identified association signals with early childhood caries: SNP rs73911671 ($p = 8.91 \times 10^{-9}$, $\beta = 2.423$) on chromosome 2 downstream of EIPR1 and SNP rs9916199 ($p = 1.46 \times 10^{-8}$, $\beta = 2.888$) on chromosome 17 upstream of AXIN2, a gene linked to tooth agenesis and colorectal cancer. Additionally, SNPs rs8055901 ($p = 2.67 \times 10^{-8}$, $\beta = 1.469$) on chromosome 16 and rs73975067 ($p = 4.71 \times 10^{-8}$, $\beta = 2.320$) on chromosome 17 are located in introns of protein-coding genes LITAF and AIPL1, respectively. Our GWAS offers evidence supportive of genetic factors contributing to complex etiology of early childhood dental caries. We plan to further investigate and meta-analyze these results to compare them to similar studies.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5073F: Genome-wide mapping of the *de novo* index ratio: quantifying and translating intolerance to gene dosage into neurodevelopmental effect sizes ⭐

Authors:

J. England[1], G. Huguet[2], K. Jizi[3], C. Poulain[4], Z. Saci[5], C. Maftei[6], E. Lemyre[7], S. Jacquemont[8]; [1]Université de Montréal, Montreal, QC, Canada, [2]CHU Ste justine, Montréal, QC, Canada, [3]CHU Sainte-Justine, Montreal, QC, Canada, [4]Montréal Univ., Montreal, QC, Canada, [5]CHU Sainte-Justine research center, Montreal, QC, Canada, [6]Ste Justine Hosp., Montreal, Montreal, QC, Canada, [7]CHU Ste justine, Montreal, QC, Canada, [8]CHU Sainte justine Res. Ctr., Montreal, QC, Canada

Abstract:

Genome-wide detection of copy number variants (CNVs) is a first-tier diagnostic test for neurodevelopmental (ND) disorders. Diagnostic laboratories classify CNVs as benign or pathogenic, but this doesn't inform on their effect size on cognitive traits. Moreover, many CNVs are classified as of unknown significance. Additive models based on constraint scores can predict the effect of CNVs on cognition but constraint metrics (LOEUF) are based on LoF SNVs (as opposed to gene dosage), and their predictive value remains limited especially for small genes and oligogenic CNVs.

Our goal is to improve CNV effect prediction using the *de novo* to inherited ratio which is a new metric highly correlated with the effect size of CNVs on cognition. To do so, we calculated this ratio using six datasets from ND and general population cohorts as well as diagnostic laboratories.

We computed a *de novo* index ratio (DIR) for each gene, as the ratio of *de novo* CNVs on the sum of *de novo* and inherited CNVs covering each gene. For multigenic CNVs, the ratio was adjusted for additional genes present in the CNV.

We analyzed the transmission of 25 234 CNVs and computed a DIR for 6 409 and 4 825 genes for deletions and duplications respectively. To obtain interpretable results we excluded multigenic CNVs with a burden over 50 points of 1 on LOEUF and computed a DIR for genes covered by at least 3 CNVs. We first focussed on 519 known ND genes which had the highest DIR compared to intolerant and tolerant genes (based on LOEUF). Many ND genes showed low DIR, meaning that not all ND genes have a high impact on neurodevelopment. Genes that have a DIR in the higher quartile of the distribution of the ND genes can represent new candidate genes for intellectual disability (ID). We identified 48 and 104 candidate genes for deletion and duplication that haven't been reported and would be looked upon because they are considered tolerant. *ARHGAP15* (10 observations in deletion, DIR 0.43) and *XKR8* (3 observations in

duplication, DIR 0.76) are promising candidates because of their respective role in regulating *RAC1* that causes autosomal dominant ID and in pruning of axons in the developing brain. We then computed the effect size of DIR on cognitive ability and it allowed us to have a better prediction through all different categories of LOEUF. The mean effect size of one unit of 1 on LOEUF is 3.6 points of IQ, but with the DIR we increase the granularity and can estimate effect sizes up to 11.3 points of IQ.
This *de novo* genomic map will be available as an online tool for clinicians to facilitate genetic counseling and the interpretation of CNVs to assess whether additional genetic investigations are warranted.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5074F: Genome-wide survival study identified variants associated with disease progression in IgA nephropathy

Authors:

L. Xu, **X-j. Zhou**, J. Lv, H. Zhang; Peking Univ. first Hosp., Beijing, China

Abstract:

Background IgA nephropathy (IgAN) has a poor prognosis and recent data suggested that almost all of patients may progress to kidney failure within their lifetime. Genetic variations intricately shape disease risks. However, genetic markers associated with the progression of IgA nephropathy have been quite limited.
Methods We performed genome-wide survival analyses in two independent cohorts and meta-analysis, the PKU-IgAN follow-up cohort containing 1859 IgAN patients and the TESTING cohort containing 279 patients from China. The primary outcome was end-stage kidney disease (ESKD), and the secondary outcome was the combination of ESKD or ≥40% reduction in estimated glomerular filtration rate (eGFR) after diagnostic kidney biopsy. Genome-wide survival analyses were performed on the two cohorts separately, and variants were selected as candidate variants when they had $P < 1\times10^{-5}$ in one cohort and $P < 0.05$ in the other. Functional annotations of candidate variants, expression quantitative trait loci (eQTL) studies, exploration of epigenetic architecture, differential expression of the candidate genes, and the association between genotype and clinical characteristics were conducted.
Results For the primary outcome, a variant of *SFMBT2* achieved genome-wide significance in the meta-analysis (HR 3.01, $P = 8.24\times10^{-9}$). Two loci, *SLC8A1* (HR 1.90, $P = 2.06\times10^{-7}$) and *MIR548B* (HR 3.19, $P = 5.04\times10^{-7}$) were considered as the candidate loci. For the secondary outcome, three loci, *ESRRG* (HR 3.68, $P = 7.26\times10^{-8}$), *FAM19A5* (HR 1.76, $P = 4.43\times10^{-7}$),

and *LINC00583* (HR 3.43, $P = 6.07\times10^{-7}$) were considered as the candidate loci. *SLC8A1*, *SFMBT2*, and *FAM19A5* variants were also associated with the eGFR slope and time-average proteinuria. In addition, the risk allele of *FAM19A5* was associated with the glomerular sclerosis ($P = 4\times10^{-3}$) and the tubulointerstitial damage ($P = 0.03$).
Conclusion Six candidate loci were significantly associated with prognosis of IgAN. These findings may help characterize molecular mechanisms of progression of IgAN, contributing to the identification of patients at high risk of progression and the development of new therapeutic targets to slow disease progression.
Keywords IgA nephropathy, end-stage kidney disease, human genetics, disease progression, genome-wide survival study

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5075F: Genomic Rare Variant Mechanisms for Congenital Cardiac Laterality Defect: A Digenic Model Approach.

Authors:

A. Rai[1], J. Klonowski[2], B. Yuan[3], K. Coveler[3], Z. Dardas[3], P. Lupo[3], A. Agopian[1], D. Kostka[2], C. Lo[2], S. Yi[4], B. Gelb[5], C. Seidman[6], E. Boerwinkle[7], J. Posey[3], R. Gibbs[3], J. Lupski[3], S. Morris[3], Z. Coban Akdemir[1]; [1]UTHlth.Sch. of Publ. Hlth., Houston, TX, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Baylor Coll. of Med., Houston, TX, [4]Dept. of Oncology and Dept. of BioMed. Engineering, Univ. of Texas at Austin, Austin, TX, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Harvard Med. Sch., Boston, MA, [7]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract:

Laterality defects (LD) are defined by perturbations in the usual left-right (LR) asymmetry of organs in the body. The genetic underpinnings of LD are complex, and the rarity, together with the phenotypic heterogeneity of these defects, challenge the elucidation of complex disease models. Herein, we used digenic model approach for the identification of contributing variants in known LD genes among exome/genome sequencing (ES/GS) data from three congenital heart disease (CHD) cohorts: Baylor College of Medicine Research Center-Genomic Research to Elucidate the Genetics of Rare Diseases (BCM-GREGoR; N=271 probands ES), Gabriella Miller Kids First Pediatric Research program (Kids First; N=228 case-parents trios GS), and Pediatric Cardiac Genomics Consortium (PCGC; N=500 trios ES). We conducted analysis on genes with known LD associations (N=115). Each personal genome was scanned for rare LD genes variants with a deleterious in-silico predictions. Seven unrelated patients (2.58%) in the BCM-GREGoR cohort inherited trans-heterozygous digenic hits, all with unaffected parents. In an extended

family of 8 members, a female proband inherited digenic hits, both alleles from an affected father and these alleles not found together in 6 unaffected family members. In Kids First and PCGC cohorts, 17 (7.45%) and 31 (6.2%) patients were inherited trans-heterozygous digenic hits respectively. Out of a total 999 families with LD in these three cohorts, 55 families (5.5%) presented with trans-heterozygous digenic hits and, one proband inherited both digenic hits from affected father. Out of 56 total digenic pairs, 42 (75%) were present in motile ciliary genes (p = 3e-3). 15 out of 42 ciliary digenic pairs were in the same structural complex of motile cilia, including *DNAH1/DNAH11* (outer dynein arm), and *DRC1/GAS8, CCDC39/CCDC40* (nexin-dynein regulatory complex). Variants in these 42 digenic pairs were likely to be potential disease causing based on their pathway similarity, phenotype similarity, co-expression rank, protein-protein interaction distance, and pathway distance (p = 2.2e-16). To assess the frequency of healthy individuals presenting with digenic events, GS data from 602 control trios provided by the 1000 Genome Project were analyzed. Two individuals had trans-heterozygous digenic events. This proportion was significantly less than that observed in CHD cohorts (BCM-GREGoR vs 1000G; p = 8e-4, Kids First vs 1000G; p = 1.16e-6, and PCGC vs 1000G; p = 6.7e-9). These findings strengthen the hypothesis of digenic inheritance among LD candidate genes and provide further evidence that digenic epistatic interaction can contribute to the complex genetics of LD.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5076F: Genomic Risk Profiles for 11 Common Diseases in the electronic MEdical Records and GEnomics (eMERGE) Network Phase IV Study

Authors:

H. Bangash[1], R. Sharp[1], N. Lennon[2], C. Weng[3], K. Kiryluk[3], C. Prows[4], J. Connolly[5], H. Hakonarson[5], L. Orlando[6], N. Abul-Husn[7], E. Kenny[7], S. Aguilar[8], S. Aradhya[8], S. Murphy[9], R. Rowley[10], T. Manolio[10], R. Chisholm[11], L. Rasmussen-Torvik[11], J. Cortopassi[12], N. Limdi[12], J. Cimino[12], D. Crosslin[13], G. Jarvik[13], D. Roden[14], M. He[14], J. Morse[14], J. Peterson[14], I. Kullo[1], J. Jackson[14], the eMERGE network; [1]Mayo Clinic, Rochester, MN, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Columbia Univ., New York, NY, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Duke Univ., Durham, NC, [7]Icahn Sch. of Med. at Mount Sinai, New York, NY, [8]Invitae, San Francisco, CA, [9]Massachusetts Gen. Hosp., Boston, MA, [10]Natl. Human Genome Res. Inst., Bethesda, MD, [11]Northwestern Univ., Chicago, IL, [12]Univ. of Alabama at Birmingham, Birmingham, AL, [13]Univ. of Washington Med. Ctr., Seattle, WA, [14]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Purpose Inclusion of genomic risk factors (monogenic and polygenic risk and family history [FHx]) improves accuracy of risk prediction for many common diseases. In the eMERGE-IV study, a genome-informed risk assessment (GIRA) for 11 common diseases is being returned to a diverse cohort of 25,000 participants. We performed an interim analysis to assess prevalence of genomic risk factors for the conditions of interest.
Methods Conditions being evaluated in adults are atrial fibrillation (AF), breast cancer (ca), chronic kidney disease (CKD), coronary heart disease (CHD), colon ca, familial hypercholesterolemia, obesity, prostate ca, and type 2 diabetes, and in children are asthma, obesity, and type 1 and 2 diabetes. Participants complete surveys including FHx and provide a specimen for genotyping and sequencing (for adults only). A GIRA which includes monogenic risk, polygenic risk scores (PRS), FHx, clinical risk factors and/or integrated risk scores (IRS) is subsequently generated and placed in the electronic health record (EHR) for disclosure to participants and clinicians. GIRAs are categorized as 'high-risk' based on condition-specific criteria such as positive FHx, presence of a pathogenic/likely pathogenic monogenic variant, a high PRS per the condition specific threshold (upper 2-10%) and/or a high IRS.
Results Of 25,664 participants,17% are pediatric, 48% racial/ethnic minorities, 65% female at birth, 12% sexual or gender minority and 11% with a disability. Survey data are available for 87%, PRS data for 70%, monogenic data for 81% and FHx data for 70%. To date, 12,282 GIRAs have been generated and 11,327 have been returned to participants (9188 adults, 2139 pediatric) and placed in the EHR. Of the GIRAs returned, 3840 (33.9%) were categorized as 'high-risk' and all network sites prioritized the return of these results. Of the high-risk GIRAs, 2102 (54.7%) were for FHx, 229 (6.0%) for monogenic risk, 1961 (51.1%) for PRS, and 182 (4.7%) for BOADICEA ≥ 25% (breast ca IRS). Of the high risk GIRAs, 634 (16.5%) reports had ≥2 types of risk present. Of the individuals at high risk, 2,938 (76.5%) were at high risk for a single condition, 769 (20.0%) for 2 conditions, 121 (3.2%) for 3 conditions, 11 (0.3%) for 4 conditions and 1 (0.03%) for 5 conditions. The 3 conditions with most high-risk GIRAs returned were CHD, CKD, and AF.
Conclusion In this interim analysis of the eMERGE IV study, ~1/3 of participants who had a GIRA returned were categorized as high-risk based on presence of at least one genomic risk factor. The high-risk results are actionable and analyses using participant surveys and the EHR are underway to test the hypothesis that return of a GIRA alters outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5077F: GWAS meta-analyses of fasting glucose, fasting insulin, and HOMA-IR in Samoans

Authors:

L. Spor[1], J. C. Carlson[1], J. Wehr[1], E. M. Russell[1], M. Krishnan[1,2], S. Liu[1], H. Cheng[3], T. Naseri[4,5], M. Reupena[6], S. Viali[7,8,9], J. Tuitele[10], E. E. Kershaw[1], R. Deka[3], N. L. Hawley[5,9], S. T. McGarvey[5], D. E. Weeks[1], R. L. Minster[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univ. of Cincinnati, Cincinnati, OH, [4]Naseri & Associates Publ. Hlth.Consultancy Firm & Family Hlth.Clinic, Apia, Samoa, [5]Brown Univ., Providence, RI, [6]Lutia i Puava ʻae Mapu i Fagalele, Apia, Samoa, [7]Oceanian Univ. of Med., Apia, Samoa, [8]Natl. Univ. of Samoa, Apia, Samoa, [9]Yale Univ., New Haven, CT, [10]Dept. of Publ. Hlth., LBJ Tropical Med. Ctr., Fagaʻalu, American Samoa

Abstract:

Introduction: Globally, 462 million (1 in 16) people have type 2 diabetes (T2D), and the prevalence is higher in Samoa, affecting 1 in 4 Samoans. Polynesians are underrepresented in biomedical research, yet they may have population-specific variants important for understanding T2D. To better understand the genetics of metabolic traits relevant to T2D in Samoans, we performed meta-analysis GWASs of fasting glucose, fasting insulin, and homeostatic model assessment for insulin resistance (HOMA-IR).
Methods: We recruited 3 cohorts from Samoa and/or American Samoa in 1990, 2002, and 2010, totaling 5,174 participants. We genotyped the participants with the Illumina Global Screening Array or Affymetrix 6.0 array and performed genome-wide imputation using a Samoan-specific haplotype reference panel. Fasting glucose and fasting insulin levels were determined, and HOMA-IR was calculated from fasting glucose and insulin values. Participants who reported using diabetes medications were excluded. We performed genome-wide association testing on the remaining 4,333 participants using linear mixed models adjusted for the fixed effects of age, sex, and ancestral principal components derived from PC-AiR, and for random effects of kinship derived from PC-Relate. The results from each of the 3 cohorts were combined with p-value-based meta-analysis with METAL.
Results: No SNVs were statistically significant ($p < 5\times10^{-8}$) in the 3 GWASs. The most significant variant associated with fasting glucose was rs373863828 ($p = 8.2\times10^{-8}$, MAF = 27%), a missense variant in *CREBRF*, which is known to be associated with higher BMI and lower odds of T2D in Samoans. Following the *CREBRF* finding, fasting glucose had a suggestive association with an intronic variant in *LINC00877* (rs902318467, $p = 1.9\times10^{-7}$, MAF = 2.4%), a lncRNA gene which contains many enhancer-like signals and is located < 500kb upstream of *PROK2* and *FOXP1*, two genes involved in glucose homeostasis. The first two introns of *PARD3B* contained 86 variants that were statistically suggestive for an association with fasting insulin. Lastly, the peak variant

associated with HOMA-IR was in *AKR1C2*, a gene involved in aldehyde and ketone metabolic pathways.
Conclusion: We identified several potential genetic determinants of metabolic traits in Samoans, a population at high risk for T2D, through GWASs of fasting glucose, fasting insulin, and HOMA-IR. While the *CREBRF* finding is concordant with prior research, the others require further investigation. The findings from this study may shed light on the etiology of metabolic disorders and highlight the importance of including underrepresented groups in biomedical research.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5078F: GWAS of overweight, not obesity, identifies new genes associated with overweight in two datasets

Authors:

H. Duan[1], **K. Arbeev**[1], O. Bagley[1], D. Wu[1], R. Holmes[1], A. Yashin[2], S. Ukraintseva[1]; [1]Duke Univ., Durham, NC, [2]Duke Univ., Chapel Hill, NC

Abstract:

Introduction: Our previous study found that being overweight in late life is associated with increased chances of extreme longevity and decreased risk of Alzheimer's Disease (AD), compared to normal weight, underweight, or obesity. Approaches, such as Mendelian Randomization and Causal Mediation Analysis can be used to explore causality underlying these associations. These approaches need to use genetic variants that have been reliably associated with overweight in late life. To find such variants, we performed genome-wide association studies (GWAS) of overweight using data from the Framingham Heart Study (FHS) original cohort and the Health and Retirement Study (HRS). One should note that previous genetic association studies rarely addressed 'overweight', especially in older adults, and predominantly focused on finding candidate genes for obesity. Our recent research, however, suggested that overweight and obesity may differently influence longevity, as well as AD. So, there is a possibility that genes associated with overweight and obesity are also different. **Method:** Using the FHS and HRS data, we performed GWAS of overweight defined as a binary variable (1 - 'overweight' BMI 25-30 vs. 0 - 'normal' BMI 18.5-24.9) at ages 65-75. Covariates included sex (1 - male, 2 - female), year of birth, education level (0 - below high school, 1 - high school, 2 - above high school), smoking status (0 - never smoked, 1 - ever smoked), and the first five principal components. **Result:** SNPs in three genes (*INSIG2*, *PSEN2*, *RYR2*) were significantly (p-values less than 1E-3) associated with overweight in both FHS and HRS data. *INSIG2* may be linked to overweight through its connection to lipid levels and insulin. *PSEN2* is a well-known

gene associated with early-onset AD. *RYR2* was previously linked to obesity and glucose homeostasis in mice. **Conclusion**: Our GWAS of overweight versus normal weight identified new genes associated with overweight (not obesity) in older adults aged 65-75 years in FHS and HRS. The SNPs in these genes (at the individual and gene level) will be used in a causal analysis of the protective effect of overweight on AD onset in later life.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5079F: High PRS for hypercholesterolemia is equivalent to Familial Hypercholesterolemia risk in a large prospective study in primary care settings

Authors:

S. Suckiel[1,2], E. R. Soper[1,2], L. Scarimbolo[1], J. A. Odgis[1], P. N. Marathe[1], K. E. Bonini[1], S. Nirenberg[3], E. J. Gallagher[2,4,5], A. R. Kontorovich[1,2,6,7], M. F. Murray[1,2], E. E. Kenny[1,2,6], N. S. Abul-Husn[1,2,6]; [1]The Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Scientific Computing and Data Div., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Div. of Endocrinology, Diabetes and Bone Disease, Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Tisch Cancer Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Cardiovascular Res. Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Hypercholesterolemia (HCL) is common yet frequently unrecognized in routine care. HCL stems from genomic (monogenic and/or polygenic) and environmental factors, and identifying individuals with genomic risk can enable effective clinical interventions and improved health outcomes. The eMERGE IV study is implementing a risk assessment for HCL, including cross-ancestry polygenic risk score (PRS), monogenic results for familial hypercholesterolemia (FH), family history, and clinical risk factors. We assessed the impact of having a PRS in the top 3%-ile for HCL (PRS+) or a positive monogenic result for FH (FH+) compared to not having high genetic risk (not-high-risk) on HCL-related risk factors and health outcomes.
Our analysis included 11,166 adults enrolled in eMERGE IV. Participant demographics, clinical data, and family history were collected through surveys or extracted from electronic health records. We used Anova and post-hoc Tukey tests to compare continuous variables and Chi-squared tests for dichotomous variables. Participants' mean age was 51 (range 18-75), 67% were female, and most self-reported as White (51.1%), Hispanic/Latino(a) (19.8%), or Black

(14.8%). There were 231 (2.1%) PRS+, 77 (0.7%) FH+, and 10,856 (97.2%) not-high-risk participants. Two (0.02%) were both PRS+ and FH+ and were excluded from analysis.
There were no significant differences in age, sex, self-reported group, BMI, physical activity, smoking history, type 2 diabetes, or chronic kidney disease among the three result groups. Between PRS+ and FH+ participants, there were no differences in LDL-C (mean 121.5 mg/dL and 123.6 mg/dL, respectively), family history, or self-reported HCL diagnosis. PRS+ had 19.4 mg/dL (95% CI 13.7-25.1 mg/dL; $p<3.4\times10^{-11}$) higher and FH+ had 21.5 mg/dL (95% CI 11.7-31.2mg/dL; $p<7.2\times10^{-7}$) higher LDL-C compared to not-high-risk participants (mean LDL-C 102.1 mg/dL). PRS+ ($p<1.1\times10^{-4}$) and FH+ ($p<7.8\times10^{-4}$) participants were more likely to have a positive family history (≥1 first degree relative with HCL). PRS+ ($p<1.5\times10^{-10}$) and FH+ ($p<2.7\times10^{-9}$) participants were more likely to report HCL diagnoses compared to not-high-risk individuals. However, of those with HCL diagnoses, FH+ were more likely to be on lipid-lowering medication compared to PRS+ participants ($p<0.046$).
These results suggest that individuals with high PRS for HCL demonstrate comparable levels of LDL-C and rates of HCL diagnoses to those with genomic FH risk, but have lower treatment rates. Implementation of PRS for HCL in primary care could potentially identify four-fold more individuals at high risk, and opportunities to optimize HCL management.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5080F: Higher African ancestry proportion attenuates the increase of serum phosphorylated tau181 in a longitudinal study of aging African Americans

Authors:

L. Wang[1], H. Xu[1], Y. Sun[1], Y. Lu[1], M-K. Lei[1], M. Ong[1], R. Philibert[2], R. Simons[1], S. Beach[1], K. Ye[1]; [1]Univ. of Georgia, Athens, GA, [2]Univ Iowa, Iowa City, IA

Abstract:

Background: Individuals of African ancestry have been shown to suffer from a higher risk of Alzheimer's disease (AD) than those of European ancestry. This study leverages an admixed African American cohort to investigate the relationship between the proportion of African ancestry and five plasma biomarkers for AD and dementia: phosphorylated tau181 (p-tau181), amyloid beta 40 (Abeta40), amyloid beta 42 (Abeta42), neurofilament light (NFL), and glial fibrillary acidic protein (GFAP).
Methods: The Family and Community Health Study (FACHS) started in 1996 and recruited 5th graders and their caregivers from several hundred African American families. This cohort was

followed, and eight waves of data and biological samples have been collected. Plasma biomarker levels were measured in 565 caregivers at least once in samples collected at either Wave 5 (2008) or Wave 8 (2019). A subset of 223 participants has measurements from both waves. Using the 1000 Genomes Project as a reference dataset, we defined the genetic ancestry of FACHS participants via principal component analysis and estimated genetic ancestry proportions using ADMIXTURE analysis. We tested associations between African ancestry proportion and biomarker levels in all 565 participants. Additionally, we examined associations with changes in biomarker levels in participants with two measurements.
Results: All biomarkers were positively associated with age, and serum GFAP levels were lower in women. No significant associations were found between African ancestry proportion and the five biomarkers. In a subset analysis comparing individuals with >90% African ancestry to those with > 90% European ancestry (n = 85, 9, respectively), we found that the high-African-ancestry group was associated with lower serum levels of Abeta40 and Abeta42. In participants with longitudinal data over 11 years, African ancestry proportion was associated with a slower rate of increase, no increase, or even a decrease in plasma p-tau181 levels. Moreover, when utilizing Polygenic Scores (PGS) and APOE4 status, which are highly associated with the biomarkers and Alzheimer's pathology in European cohorts, we found no associations with the biomarkers in our African American cohort.
Conclusion: We found suggestive evidence that higher African ancestry proportion is associated with an attenuated increase of p-tau181 levels over time. More research is needed to examine the population-specific pathology of AD and dementia and to identify genetic associations in African American populations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5081F: Host genome-wide association study of the Swedish gut microbiome

Authors:

K. F. Dekkers[1], K. Pertiwi[1], P. Lundmark[1], U. Hammar[1], G. Baldanzi[1], M. Miari[2], D. Nguyen[1], S. Sayols-Baixeras[1,3,4], A. C. Eklund[5], J. B. Holm[5], H. B. Nielsen[5], G. Méric[1,6,7,8,9], C. Ohlsson[10,11], K. Hveem[12,13,14], O. Melander[2,15], P. M. Nilsson[2], G. Engström[2], K. Michaëlsson[3], J. Ärnlöv[16,17], M. Orho-Melander[2], T. Fall[1]; [1]Molecular Epidemiology, Dept. of Med. Sci., Uppsala Univ., Uppsala, Sweden, [2]Dept. of Clinical Sci. in Malmö, Lund Univ., Malmö, Sweden, [3]Med. Epidemiology, Dept. of Surgical Sci., Uppsala Univ., Uppsala, Sweden, [4]CIBER Cardiovascular diseases (CIBERCV), Inst. de Salud Carlos III, Madrid, Spain, [5]Clinical Microbiomics A/S, Copenhagen, Denmark, [6]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Inst.,

Melbourne, Australia, [7]Dept. of Cardiometabolic Hlth., Univ. of Melbourne, Melbourne, Australia, [8]Central Clinical Sch., Monash Univ., Melbourne, Australia, [9]Dept. of Cardiovascular Res., Translation, and Implementation, La Trobe Univ., Melbourne, Australia, [10]Dept. of Internal Med. and Clinical Nutrition, Inst. of Med., Sahlgrenska Osteoporosis Ctr., Ctr. for Bone and Arthritis Res. at the Sahlgrenska Academy, Univ. of Gothenburg, Gothenburg, Sweden, [11]Region Västra Götaland, Sahlgrenska Univ. Hosp., Dept. of Drug Treatment, Gothenburg, Sweden, [12]HUNT Ctr. for Molecular and Clinical Epidemiology, Dept. of Publ. Hlth.and Nursing, Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [13]HUNT Res. Ctr., NTNU, Levanger, Norway, [14]Levanger Hosp., Nord-Trøndelag Hosp. Trust, Levanger, Norway, [15]Dept. of Internal Med., Skåne Univ. Hosp., Malmö, Sweden, [16]Div. of Family Med. and Primary Care, Dept. of Neurobiology, Care Sci. and Society, Karolinska Inst.t, Huddinge, Sweden, [17]Sch. of Hlth.and Social Studies, Dalarna Univ., Falun, Sweden

Abstract:

Objective: An improved understanding of host genetic determinants of the gut microbiome may provide insights into physiological mechanisms. Here, we present results from the largest metagenomic sequencing-based genome-wide association studies (GWAS) of the human gut microbiome in 16,040 participants aged 18-96 from the population-based Swedish studies of SCAPIS, SIMPLER and MOS.
Methods: After processing all metagenomes with the same pipeline and imputing the human genotypes against the HRC1.1 reference panel, a series of GWAS were conducted for 921 microbial species and the alpha diversity measures Shannon diversity, inverse Simpson and species richness using REGENIE. We assessed presence or absence for species with 5-50% prevalence, and rank-inverse normalized relative abundance for species with prevalence ≥50% and alpha diversity. Covariates were sex, age, age^2, metagenomic DNA extraction plate and genetic principal components. Cohort-specific results were meta-analyzed using METAL.
Results: We identified eight study-wide significant loci associated with up to four microbial species (p-value < $5x10^{-8}$/921) and one study-wide significant loci associated with species richness (p-value < $5x10^{-8}$/3). Besides the previously replicated *ABO* and *LCT* loci, we identified study-wide significant loci that have previously been reported at genome-wide significance, such as *FUT2*, and novel loci associated with the gut microbiome, including a locus between genes *FUT3* and *FUT6*. Results were consistent across cohorts.
Conclusions: The large sample size and harmonized data allowed for the discovery of new host genetic loci associated with human gut microbial species and richness, including more blood group and secretor status loci. We plan to replicate these findings further and share the complete GWAS summary statistics with the research community. Of note, we are currently performing GWAS for genes and gene function.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5082F: Host genomics is associated with longitudinal infant gut microbial changes in childhood asthma and atopy

Authors:

S. Sivanathan[1], S. A. Stickley[2], Z. Y. Fang[1], A. Ambalavanan[1], E. George[1], C. Petersen[3,4,5], D. Dai[3,4,5], P. Mandhane[6], E. Simons[7], T. J. Moraes[8,9], M. G. Surette[10], M. B. Azad[7,11], S. E. Turvey[3,4,5], P. Subbarao[8,9], Q. Duan[1]; [1]Queen's Univ., Kingston, ON, Canada, [2]Queen s Univ., Kingston, ON, Canada, [3]Univ. of British Columbia, Vancouver, BC, Canada, [4]British Columbia Children's Hosp., Vancouver, BC, Canada, [5]Child and Family Res. Inst., Vancouver, BC, Canada, [6]Univ. of Alberta, Edmonton, AB, Canada, [7]Univ. of Manitoba, Winnipeg, MB, Canada, [8]Hosp. for Sick Children, Toronto, ON, Canada, [9]Univ. of Toronto, Toronto, ON, Canada, [10]McMaster Univ., Hamilton, ON, Canada, [11]Children's Hosp. Res. Inst. of Manitoba, Winnipeg, MB, Canada

Abstract:

Earlier studies of infant gut microbiota reported associations with childhood asthma and atopy. These studies, however, investigated microbial abundances at a single time point, which do not consider if and how microbial dysbiosis over time may impact risk of atopic diseases. Therefore, our study investigates the main effects of longitudinal microbial changes during the first year of life on risk of childhood asthma and atopy, as well as their interaction effects with host genomics on disease prevalence. First, we calculated changes in abundances of 29 microbes that were present at 3 months and 1 year of age (e.g., slope), which were derived from 16S rRNA sequencing of stool samples collected from infants in the CHILD Study (N=584) . Next, univariate regression analyses identified microbial changes associated with health outcomes in pre-school aged children (e.g., asthma , atopic dermatitis, and food/inhalant sensitization by age 5 years) . We determined that changes in *Flavonifractor sp.* ($P_{Benjamini-Hochberg(BH)}$ =0.015), *Anaerostipes sp.8* (P_{BH}=0.011) and *Clostridium innocuum sp.* (P_{BH}=0.011) abundances were associated with increased prevalence of food sensitization by 3 years of age. Moreover, we performed gene-environment interaction (GxE) analyses using infants' polygenic risk scores (PRS) for disease (G) and the changes in microbial abundances as environment exposures (E). This analysis identified additional associations, such as *Flavonifractor sp.* and increased asthma prevalence by age 5 years (P_{BH}=0.028), only when simultaneously considering the infants' individual polygenic risk for lung function. Finally, single nucleotide polymorphisms (SNPs) genotyped from the Illumina HumanCoreExome Bead Chip were used in genome-wide association studies (GWAS) to determine genetic factors contributing to changes in gut microbial composition in the first year of life. Interestingly, GWAS detected SNPs, such as

rs1487669249 in the *KCNN2* gene, encoding a calcium-activated potassium channel, which were associated with the slope of *Bacteroides sp. 6* (P=1.4e-08). In conclusion, our findings suggest host genomics could impact gut microbiota changes during the first year of life, which may contribute to childhood asthma and atopic disease susceptibility. These changes may represent potential microbial biomarkers for disease-related dysbiosis which can be further investigated for therapeutic and preventive applications.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5083F: Identification of candidate genes associated with congenital hypopituitarism through *PROP1* phylogenetic tree and transcriptomic analysis

Authors:

N. Trigueiro[1], D. M. Adamoski[2], L. Carvalho[1]; [1]Hosp. das Clínicas da Faculdade de Med. da Univ.e de São Paulo, São Paulo, Brazil, [2]Centro Natl. de Pesquisa em Energia e Materiais, Laboratório Natl. de Biociências., Campinas, Brazil

Abstract:

Congenital hypopituitarism (CH) is a rare endocrine disorder characterized by the deficiency of one or more pituitary hormones. Molecular diagnostics of CH reveal a low detection rate of 12.7%, with 11.2% sporadic cases and 63% familial in which most cases are attributed to variants in the *PROP1* gene, a key regulator of pituitary development and function. *PROP1*, a PRD Homeobox domain-containing gene, is crucial for pituitary stem cell homeostasis, differentiation, and migration. A phenomenon known as transcriptional adaptation has been observed where the loss of function in certain genes can be compensated by homologous genes with similar function. This study explores the potential of *PROP1* homologs in causing CH. To identify *PROP1* homologs, a BLASTP analysis was performed using the *prop1* from *Danio rerio* against all Deuterostome classes. This species was chosen for its evolutionary position linking older and more derived vertebrates. Sequence alignment was made in MUSCLE and the tree built using VeryFastTree. Differential expression analysis was conducted using GTEx bulk RNA sequencing data to identify upregulated genes in the pituitary compared to 53 other tissues using DESeq2. Genes identified in the phylogenetic analysis were then cross-referenced with differential expression analysis. Finally, Whole Exome Sequencing (WES) data of 52 CH patients without molecular diagnoses from HCFMUSP were analyzed using the Ensembl Variant Effect Predictor to identify genetic variants. The phylogenetic analysis identified 55 gene families comprising 178 genes. Some of these gene families, such as SHOX, OTX, LIM, PAX and PITX, have

been previously associated with CH or other endocrine disorders. Differential expression analysis revealed 35 genes that were upregulated in the pituitary gland compared to at least 27 other tissues. Among these, 24 genes, including *RAX2*, *ALX3* and *DMBX1*, had no previously known association with pituitary development or CH, according to OMIM. WES analysis of CH patient cohort identified 217,214 missense variants, 216,012 synonymous variants and 5,468 frameshifts. Several variants were identified in *PROP1* homologs that require functional characterization. The integration of phylogenetic analysis, transcriptomic and exomic data has revealed new candidate genes potentially involved in CH. The upregulation of *PROP1* homologs in the pituitary and presence of variants in these genes in CH patients underscores the need for comprehensive functional studies to elucidate their roles in CH. These findings, provide new foundation for understanding CH and pituitary development.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5084F: Identification of differentially expressed genes in postmortem tissue from African American Alzheimer disease cases and controls

Authors:

M. Logue[1,2], A. Labadorf[3], N. O'Neil[2], D. Dickson[4], M. Flanagan[5], E. Franklin[6], A. Frosch[7], M. Gearing[8], L-W. Jin[9], J. Kofler[10], R. Mayeux[11], A. McKee[12], C. Miller[13], M. Murray[14], P. Nelson[15], R. Perrin[6], J. Schneider[6], T. Stein[16], A. Teich[17], J. Tronsco[18], S-H. Wang[19], J. Mez[2], L. Farrer[2]; [1]Natl. Ctr. for PTSD at VA Boston, Jamaica Plain, MA, [2]Med. (BioMed. Genetics), Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [3]Bioinformatics Hub, Boston Univ., Boston, MA, [4]Dept. of NeuroSci., Mayo Clinic Florida, Jacksonville, FL, [5]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [6]Knight ADRC, Washington Univ. Sch. of Med., St. Louis, MO, [7]C.S. Kubik Lab. for Neuropathology, Massachusetts Gen. Hosp. & Harvard Med. Sch., Boston, MA, [8]GADRC and Emory Univ. Dept. of Pathology & Lab. Med. and Dept. of Neurology, Atlanta, GA, [9]Univ. of California Davis, Sacramento, CA, [10]Univ. of Pittsburgh, Pittsburgh, PA, [11]Columbia Univ, New York, NY, [12]Boston Univ. Alzheimer's Disease Ctr. and CTE Ctr., Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [13]Dept. of Pathology, Univ. of Southern California, Los Angeles, CA, [14]Dept. of NeuroSci., Mayo Clinic, Jacksonville, FL, [15]Univ. of Kentucky, Lexington, KY, [16]Dept. of Pathology and Lab. Med., Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [17]Dept. of Pathology and Cell Biology, Columbia Univ., New York, NY, [18]Dept. of Pathology, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [19]Duke Univ. Med. Ctr., Durham, NC

Abstract:

Although the rate of Alzheimer's disease (AD) in African-ancestry (AA) Americans is higher than that of persons from European-ancestry (EA) populations, AA participants have been underrepresented in AD neuropathological studies. Utilizing the AD Research Centers (ADRC) infrastructure, we obtained AA donor pre-frontal cortex (PFC) tissue from brain repositories of 12 ADRC and generated bulk RNA sequencing (RNA-seq) data for 179 samples that met QC and inclusion criteria. Previously generated PFC RNAseq data were obtained for 28 additional AA donors from the Columbia University ADRC. Differential gene expression was evaluated among 125 donors with a neuropathological diagnosis of AD (NIA-Reagan intermediate or high likelihood) and 82 neuropathologically confirmed controls using regression models including covariates for age at death, sex, cell-type frequencies, and RNA integrity number (RIN) calculated with Limma. FDR-corrected p-values (p_{adj}) were calculated to control for the 33,611 genes examined. A total of 482 genes surpassed the multiple-testing threshold. The most significant, *ADAMTS2* ($p=2.96x10^{-8}$, $p_{adj}=0.001$), showed increased expression in AD cases. We note that *ADAMTS2* was differentially expressed in a prior EA study of neuropathologically confirmed AD cases and controls (Panitch et al. *Molecular Psychiatry* 2021). Additionally, a recent analysis of cognitive resilience in EA neuropathological AD cases identified a strong association with *ADAMTS2*. Of the differentially expressed genes observed in the Panitch et al. EA study, 385 (35%) were nominally significant,65 (5.8%) were corrected significant, and most (89%) of these genes had the same effect direction in the AA cohort. Some of the observed associations appear to be AA specific (e.g., *EFR3B*, *IRS4*, and *CA12*). Additionally, we found nominally significant ($p<0.05$) associations with expression of *APOE* (Log2 fold change [L2FC=-0.20, p=0.012) and several other established AD genes including *SORL1* (L2FC=-0.10, p=0.014) and *IGF1R* (L2FC=0.11, p=0.0013). This largest-ever (to our knowledge) gene expression study of AD in postmortem brain tissue from AA donors implicates many more genes as having a role in AD in this population than previously identified in genome-wide association studies and provides insight into trans-ancestry differences risk for AD.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5085F: Identifying functional type 1 diabetes risk variants in T cells and beta cells using a massively parallel reporter assay

Authors:

K. A. Spaulding[1,2], S. N. Lavalle[1], A. Fonseca[1], C. W. McGrail[2], K. J. Gaulton[2], G. McVicker[1,2]; [1]Salk Inst. for Biological Studies, La Jolla, CA, [2]Univ. of California, San Diego, La Jolla, CA

Abstract:

Type 1 diabetes (T1D) is an autoimmune disorder characterized by T cell-mediated destruction of insulin-producing beta cells. Genetic variants associated with T1D are predominantly found in non-coding regions of the genome, and are presumed to affect gene regulation. These variants are difficult to interpret because linkage disequilibrium obscures the identities of causal variants and we don't know which variants affect gene regulation in specific cell types. Deciphering which variants influence gene regulation in T cells and beta cells is a crucial step for elucidating the mechanisms underlying T1D susceptibility. This study aims to identify T1D-associated risk variants that affect regulatory activity in T cells and pancreatic beta cells using a massively parallel reporter assay (MPRA).
We used SuSiE to fine-map 102 T1D-associated loci using summary statistics from a new meta-analysis of 817,718 samples (20,355 T1D cases and 797,363 non-diabetes controls). We identified 138 independent signals and a total of 8,592 risk variants with a posterior inclusion probability (PIP) >0.001. We designed a custom oligo library that centered each risk variant's reference and alternative allele within 250 bp of flanking genomic sequence, and also included 2,430 positive and negative controls. The oligos were cloned into an MPRA vector, each tagged with a unique barcode at the 3'-UTR of the reporter. This plasmid library will be transfected into primary T cells and EndoC-βH1 cells. Following transfection, transcripts will be isolated from these cells, and relative expression will be quantified using Illumina deep sequencing. Allele-specific read counts from the MPRA will be analyzed to determine which variants affect enhancer function.
This analysis is expected to reveal T1D-associated variants that influence enhancer activity in T cells and beta cells. The findings will provide critical insights into the genetic basis of T1D, highlight potential therapeutic targets, and pave the way for future research into the regulatory mechanisms of autoimmune diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5087F: Impact of BMI-driven proteins on nonalcoholic fatty liver disease (NAFLD)

Authors:

S. Cao, G. Butler-Laporte; McGill Univ., Montreal, QC, Canada

Abstract:

Nonalcoholic Fatty Liver Disease (NAFLD) is a major public health challenge, closely linked to the epidemic of obesity and metabolic syndrome. The progression of NAFLD from simple steatosis to nonalcoholic steatohepatitis and to cirrhosis or liver cancer poses significant morbidity and

mortality risks. While the association between obesity and NAFLD is well-documented, the specific molecular mechanisms driven by obesity that led to NAFLD progression are not well understood. This gap persists despite advances in proteomics, which have revealed potential biomarkers and therapeutic targets. Currently, the application of Mendelian randomization (MR) to explore these molecular pathways presents an opportunity to discern the causal effects of specific BMI-influenced proteins on NAFLD, using genetic variants to address confounding and reverse causation, important limitations in observational studies. Initially, our study employs two-sample MR to estimate the effect of BMI on circulating protein levels and the causal effects of the identified proteins on NAFLD patients, leveraging extensive proteomic data from deCODE genetics and Genome-Wide Association Studies (GWAS) data from UK Biobank. Then, we will validate the causality between prioritized proteins and obesity, and exclude the potential bias from linkage disequilibrium (LD) by performing separate MR analyses with body fat percentage (another proxy measure for obesity) as exposure and colocalization analyses to evaluate whether the cis-pQTLs of the identified proteins and NAFLD outcomes shared a single causal variant. Furthermore, to gain cellular functional insights of identified proteins in NAFLD patients' livers, we will integrate single-cell RNA-sequencing data from liver biopsy samples from Single-Cell Portal of the Broad Institute. Ultimately, this research seeks to illuminate the pathogenesis of NAFLD and offer new avenues for therapeutic intervention, improving diagnosis and management of this prevalent condition.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5088F: Improved Diagnostic Accuracy by Trio Whole Exome Sequencing in Autism Spectrum Disorder - Experience from a single quaternary care centre in India

Authors:

C. Ghati, G. A. Udupi, A. Kotambail, R. Sukrutha, H. Padmanabha, R. K, S. Nashi, J. P. John, R. K M, R. Mundlamuri; NIMHANS, Bangalore, India

Abstract:

Background / Objectives: Globally 1 in 100 children is estimated to be suffering from Autism Spectrum Disorder [ASD]. Prevalence in India is estimated to be about 1 in 500 [0.20%]. Males are affected at least three times more frequently than females. Aetiologically ASD has been heterogeneous, with over 736 genes [loci] known to be associated with. Parting a few well characterized syndromes with known pattern of inheritance, for majority of cases of ASD, it is difficult to establish a genetic diagnosis clinically. Even after extensive genetic testing, genetic

aetiology for a majority of cases remains obscure [17.1-23.7% positivity]. We describe the experience of genomic investigation into ASD from a single quaternary care centre.
Methods: Among our cohort of 82 cases diagnosed with ASD from India underwent Extensive aetiological evaluation. Targeted Exome Sequencing in 46 cases and singleton testing by Whole Exome Sequencing [WES] was performed in 26 cases while Trio WES was performed for 10 cases. The study was initiated after appropriate consenting process. Genotype phenotype correlation was performed to evaluate the significance of the likely causative variants.
Results: Parting the core manifestations of ASD such as impairment of social interaction, expressive language deficits, stereotypies and other behavioural abnormalities, the cohort had epilepsy in 24/82 cases and dysmorphism in 30/82. Rett/Rett like features were noted in 31/82 cases. WES and targeted sequencing in the probands revealed pathogenic, likely pathogenic variants and variants of uncertain significance in 21, 13 and 15 cases, respectively and copy number variants in 2 cases of which 20 were ultra-rare and 30 novel variants. Thus, achieving a clear diagnosis in 35/82 cases and a diagnostic yield of 42.6%. While MECP2 related Rett were relatively easily detectable by Targeted Exome Sequencing, WES helped to identify likely causative variants in less implicated genes like CTNNB1, HNRNPH2, FOXP1, SETD2, and CC2D1A with a broader phenotype spectrum. However, Trio WES sequencing improved variant detection than WES and targeted exome with a diagnostic yield of 60% and was especially beneficial in the evaluation of variants in implicated genes with variable penetrance such as X-linked genes like TMLHE.
Conclusion: As NGS studies are increasingly identifying rare variants in novel genes associated with ASD, Trio WES should be considered as a first line diagnostic test that could improve diagnostic yield in ASD cases with an uncharacteristic phenotypic pattern.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5089F: Improved prediction of hepatic steatosis in the UK Biobank using multimodal predictors

Authors:

Y. Alyousifi, C. Raut, A. Oliveri, V. Chen, E. Speliotes; Univ. of Michigan, Ann Arbor, MI

Abstract:

Current risk analysis tools for hepatic steatosis rely on imaging techniques, which are not readily available. Non-invasive prediction models that can be used more widely are being developed and refined. We identified 17 genetic loci that are associated with metabolic dysfunction associated steatohepatitis (PMID 37709864). Here we study the influence of those 17 genetic

variants on improving the conventional steatosis indices for hepatic steatosis prediction. We analyzed data from 29,632 participants in the UK Biobank who had liver fat quantified using MRI-PDFF. The ages of these individuals varied from 40 to 70 years, and 52% were female. Previous predictors of hepatic steatosis include the Dallas Steatosis Index (DSI) (PMID 31982611), Hepatic Steatosis Index (HSI) (PMID 19766548), Framingham Steatosis Index (FSI) (PMID 27046482). These scores use various combinations of diabetes status, genetic ancestry, hypertension status, genetic sex, glucose levels, alanine aminotransferase (ALT), body mass index (BMI), age at recruitment, aspartate aminotransferase (AST), triglycerides to predict steatosis. We assessed the performance of these indices in predicting an MRI PDFF ≥ 5% in the UK Biobank. The data were split into a training set (23,705) and a testing set (5,927). We employed binary multivariate logistic regression models and determined their accuracy in predicting PDFF values ≥ 5% using the area under the receiver operating characteristic curve (AUC). We also created our own non-invasive score by training a model in the UK biobank data directly using the previously mentioned predictors both with and without specific SNPs from study PMID 37709864. MRI PDFF in the UK Biobank has been predicted with AUC of 0.81, 0.80, and 0.82 using DSI, HSI, and FSI respectively. Our UK Biobank non-genetic baseline model had an AUC of 0.84. When we added unrelated individual genetic predictors from PMID 37709864, the model's AUC increased to 0.86. This indicates that our models could adequately predict hepatic steatosis and outperformed the existing risk analysis tools. We developed a genetic-based steatosis model to predict MRI PDFF-measured liver steatosis. Including genetic information significantly improved the hepatic steatosis prediction, demonstrating superiority over alternative models. With this score, we can improve and personalize early diagnosis of hepatic steatosis, decreasing disparities in time to intervention and diagnosis, which will enhance outcomes and reduce the burden of metabolic liver disease.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5090F: Improving Polygenic Risk Prediction of Kidney Function by removing Biomarker Specific Genetics

Authors:

J. Du[1], Q. Sun[1], N. Franceschini[1], N. Cox[2], Y. Li[1], M. Shuey[2]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Glomerular Filtration Rate (GFR), is a critical parameter for assessing renal function and renal disease risk, however, it is not routinely measured and instead estimated (eGFR) using

biomarkers. There are 5 different eGFR equations using serum creatinine and/or cystatin-C and demographic variables (eGFRcys, eGFRcr^{2021}, eGFRcr^{2009}, eGFRcr-cys^{2021}, eGFRcr-cys^{2012}) that are widely variable across calculations. We observed inconsistencies in eGFR values and GWAS results in UK Biobank (UKB) and hypothesized that such discrepancies arise from the genetics of the individual biomarkers and that removal of these biomarker specific effects may improve genetic prediction of renal function. Local genetic correlation analysis identified ~18k HapMap3 variants in 278 shared genetic regions among the 5 eGFR estimates. Variants in these shared regions (in-region) demonstrate up to 148% higher correlation in -$\log_{10}$ p-values among the 5 eGFR GWAS compared to all variants (benchmark). We then constructed 3 Polygenic Risk Scores (PRS) using PRS-CS with (1) in-region variants (m = 17,992), (2) out-region variants excluding LD buddies ($R^2 > 0.8$) (m = 1,098,350), and (3) all variants available (m = 1,117,353). The UKB Europeans (N = 441,906) were split into training (N = 350,000) and testing (N = 91,906) populations. We then used covariate adjusted association analyses to test how well the 3 PRSs predict 5 eGFR values or 4 kidney-related clinical endpoints. Compared to the benchmark PRS, our in-region PRS explained 60% more variance of eGFRcr^{2021} (the most widely used measurement). We observed less explained variance for eGFRcys by the in-region PRS, 37.4% of the R^2 compared to the benchmark, which is likely due to cystatin-C levels being impacted more often by extrarenal factors which are included in the out-region variants. This observation was supported by the addition of body mass index (BMI), a known extrarenal influence of cystatin-C, and BMI-PRS as covariates. When these are included we observed an increase of 1.7% R^2 in the in-region PRS but a decrease of 4.3% R^2 in the benchmark PRS for eGFRcys, with no change for eGFRcr^{2021}. Similarly, our in-region PRS achieved 74% of the benchmark PRS performance for ICD-defined Chronic Renal Failure and 62% for Unspecified Renal Failure. Initial validation of eGFRcr^{2021} in BioVU confirmed that the in-region PRS outperformed the benchmark with an improvement in R^2 of 5.2%. In conclusion, we identified in-region variants that reflect true GFR and by restricting genetic estimates of renal function to these variants achieved higher or similar performance than those constructed using all genome-wide variants.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5091F: Improving the predictability of polygenic risk score

Authors:

H. Xu; Augusta Univ., Augusta, GA

Abstract:

Polygenic risk scores (PRSs) have been widely used for predicting the risk of developing complex human diseases. However, the predictive power of PRS-based models has been rather limited, especially when they are used across human populations. In this study, we used the whole genome sequence data from the All of Us project for patients with type 2 diabetes. We performed genome-wide association studies of type 2 diabetes in White and African populations. PRS was constructed by using the top 100 most significant SNPs. To further improve the predictive power, we added demographic variables and life-style variables to the PRS model. We evaluated the predictive accuracy of our model using cross validation. It is found that our models have improved predictive powers over the original PRS in both White and African populations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5092F: INHBC influences the effect of Type 2 diabetes on chronic kidney disease

Authors:

K. Liang[1], T. Sasako[2], Y. Ilboudo[3], Y. Chen[1], S. Yoshiji[4], G. Butler-Laporte[1], T. Zheng[5], D. Tan[1], B. Richards[6]; [1]McGill Univ., Montreal, QC, Canada, [2]The Univ. of Tokyo Hosp., Tokyo, Japan, Tokyo, Japan, [3]Jewish Gen. Hosp., Montreal, QC, Canada, [4]McGill Univ./Broad Inst., Montreal, QC, Canada, [5]McGill, Montreal, QC, Canada, [6]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada

Abstract:

Diabetic nephropathy is a major complication among individuals living with type 2 diabetes (T2D), occurring in up to 40% of the people with diabetes, and is the leading cause of end-stage renal failure. The cause of diabetic nephropathy is not fully understood and few interventions influence chronic kidney disease in diabetics. Here, we use large-scale proteomics data to identify proteins that mediate the relationship between T2D and kidney disorders. To do so, we first used Mendelian randomization (MR) to identify 69 circulating proteins whose levels are altered by genetic predisposition to T2D in 35,447 individuals with protein measurements and up to 1,004,040 participants with measures of kidney function. Then we used cis genetic variants to proxy the causal effect of these T2D-influenced circulating proteins and found that 5 of the 69 T2D altered proteins influenced three kidney-related traits (blood urea nitrogen [BUN], estimated glomerular filtration rate [eGFR], and serum urate) and one kidney disease (nephrolithiasis). Of the five proteins, we found that increased circulating level of the inhibin beta C chain (INHBC) protein is associated with decreased eGFR and increased BUN, and serum urate. We next found that increased INHBC levels were associated with more deterioration of

kidney function in 38,782 UK Biobank participants of European ancestry. Taken together, these findings suggest that increased INHBC levels may increase the risk of diabetic nephropathy.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5093F: Insights into causes and consequences of repeat expansions from biobank sequencing data

Authors:

M. Hujoel[1,2], R. Handsaker[3,2], N. Kamitaki[1,2], R. Mukamel[1,2], S. Rubinacci[1,2], S. McCarroll[3,2], P-R. Loh[1,2]; [1]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Harvard Med. Sch., Boston, MA

Abstract:

Expansions and contractions of tandem repeats have produced many repeat-length polymorphisms in the human genome. Some expanded repeats cause inherited disorders previously studied in patient and family cohorts, and some of these are also somatically unstable. Here we analyzed blood-derived DNA sequencing data from >700,000 participants in UK Biobank (UKB) and *All of Us* to draw new insights into repeat instability. We focused on 15 CAG trinucleotide repeat loci at which >5 UKB participants carried expanded alleles (≥45 CAGs). We first explored the mutability of common alleles with ≤30 CAG repeats, observing contrasting propensities of different repeats to mutate in the germline versus somatically in blood. We estimated allele-specific intergenerational expansion and contraction rates by analyzing discordances among alleles inherited identical-by-descent. Most repeats expanded more frequently than they contracted, and mutation rates increased with allele length, rising to 0.5% per generation for the most mutable repeats (in *ATXN8OS*, *DMPK*, and *GLS*).
For four of the 15 repeats (in *TCF4*, *GLS*, *DMPK*, and *ATN1*), we further detected somatic repeat instability in blood. Repeats in *TCF4* were the most somatically unstable: in carriers of 30-CAG alleles, we estimated that 1% of blood cells contained a somatic expansion. Longer *TCF4* alleles (≥45 repeats; EUR AF=4%) were even more unstable and appeared to lengthen with age.
We performed a GWAS on a *TCF4* somatic-expansion phenotype (quantified from read alignments in the DNA repeat) and identified seven loci at which inherited variants modulate *TCF4* repeat instability. Three loci - *MSH3* ($p=2x10^{-52}$), *FAN1* ($p=9x10^{-29}$), and *PMS2* ($p=3x10^{-8}$) - overlapped mismatch repair genes known to modulate Huntington's disease age at onset, which is driven by somatic expansion of the *HTT* repeat in neurons. Surprisingly, at both *MSH3* and *FAN1*, common modifier variants had opposite effects on *TCF4* instability in blood versus *HTT* instability in brain, suggesting locus- or tissue-

specific *cis*-regulation. The four new modifier loci included DNA damage response genes *ATAD5*($p=5x10^{-12}$) and *GADD45A* ($p=3x10^{-8}$).
Phenome-wide analysis of expanded repeats uncovered pathogenic effects of *GLS* 5' UTR expansions on renal and liver biomarkers ($p=1x10^{-9}$ for cystatin C; $p=1x10^{-18}$ for GGT). Carriers had increased risk for stage 5 chronic kidney disease (OR=22.0 [5.8-82.7]) and liver diseases (OR=9.3 [2.0-44.2]). These associations appeared to be driven by toxicity of expanded alleles, as loss-of-function *GLS* alleles did not produce these effects.
These results illustrate the potential for new insights into repeat instability from biobank data sets.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5094F: Insights into mosaic chromosomal alterations from whole-genome sequencing in UK Biobank

Authors:

D. Tang[1], R. Mukamel[2], S. Rubinacci[3], P-R. Loh[4]; [1]Harvard Univ., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA, [4]Brigham and Women's Hosp., Boston, MA

Abstract:

Clonally expanded cell populations carrying mosaic chromosomal alterations (mCAs) are commonly detectable in the blood of elderly individuals. However, previous SNP-array-based mCA analyses have had limited sensitivity to detect mCAs of smaller size or lower clonal fraction. We developed new computational methods to sensitively detect mCAs from whole-genome sequencing (WGS) data available for 490,414 participants in UK Biobank, including an efficient pipeline for WGS-based allelic imbalance analysis (achieving >2-fold increased sensitivity) and methods that make use of WGS read-pair and read-depth information.
These WGS-based analyses showed that focal deletions are >10-fold more common than previously appreciated: short (~1Mb) deletions at *DLEU2*, *DNMT3A*, and *TET2* were each detected in ~1% of individuals, suggesting that such mCAs may account for a substantial fraction of cases of clonal hematopoiesis with previously unknown drivers. Breakpoints of these deletions clustered in hotspots on the scale of 10-100kb, indicating that local genomic or epigenomic features strongly influence generation of mCAs. Shorter deletions at these loci appeared to rise to higher cell fractions more quickly than multi-megabase deletions, suggesting that deletion of larger regions surrounding target genes may be detrimental to clonal fitness. Breakpoint analysis of mosaic deletions genome-wide implicated non-homologous end joining

(NHEJ) as the dominant mutational mechanism producing mCAs and enabled reconstruction of complex somatic structural variants present in apparently healthy individuals.
Detectable B-cell and T-cell clonality in 1,348 and 2,822 individuals with multiple WGS read pairs derived from the same V(D)J recombination enabled assignment of certain mCAs to hematopoietic lineages. B- and T-cell clonality were broadly enriched among mCA carriers (OR=8.0 [7.1-8.9] and OR=2.4 [2.2-2.7] for B- and T-cell clonality, respectively), with the strongest enrichments observed for B-cell clonality in carriers of *DLEU2* deletions (OR=103.5 [89.1-120.3]) and trisomy 12 (OR=21.3 [16.4-27.5]), consistent with these events being drivers of CLL. Other notable mCAs that were enriched in individuals with B-cell clonality included 7q deletions (OR=17.0 [10.7-26.9]) and 17p CN-LOH (OR=17.5 [9.6-31.8]). Mosaic trisomy 21 associated strongly with T-cell clonality (OR=10.9 [8.4-14.3]), possibly indicating T-cell dysregulation, which has been observed in cases of constitutional trisomy 21 (Down syndrome). These results demonstrate the potential for further insights into clonal hematopoiesis from new approaches to analyze large WGS data sets.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5095F: Insulin resistance measure TG:HDL-C ratio adjusted for BMI polygenic risk score associates with metabolic syndrome sequalae in the Michigan Genomics Initiative

Authors:

A. Oliveri, R. J. Rebernick, E. Speliotes; Univ. of Michigan, Ann Arbor, MI

Abstract:

Introduction & Objective: Insulin Resistance (IR) is genetically influenced and known to increase risk of cardiometabolic disease. Genome-wide association studies (GWASes) of IR have been limited in scale, due to limited samples with direct measures of IR. Our study circumvents this limitation by using the surrogate insulin resistance measure triglyceride to high-density lipoprotein cholesterol ratio adjusted for BMI (TG:HDL-C adj BMI). Here we aim to: 1) identify genetic loci linked to IR through TG:HDL-C adj BMI; 2) study the cumulative effect of these loci using a polygenic risk score (PRS).
Methods: To identify genetic loci associated with IR, we performed a GWAS of the TG:HDL-C adj BMI in the UK Biobank (N=401,090). We fit a mixed linear model controlling for sex, age and principal components in SAIGE. We defined 'high-confidence IR-associated loci' (HCIRL) as the SNPs reaching genome-wide significance ($p < 5e-08$) for the TG:HDL-C adj BMI with effects (FDR < 0.05) on MAGIC Consortium fasting insulin, HOMA-IR or insulin sensitivity. Association

between a HCIRL PRS and ICD-defined Phecodes were tested in the Michigan Genomics Initiative (MGI) (N=51,527). Significantly associated traits (Bonferroni $p_{adj} < 0.05$) were identified after controlling for sex, age and principal components.
Results: 343 lead SNPs associated with TG:HDL-C adj BMI at genome-wide significance levels. 119 of these were HCIRL. The HCIRL PRS showed association with phecodes related with IR and metabolic syndrome including hyperglyceridemia, hyperlipidemia, type 2 diabetes, obesity, abnormal glucose, myopia, hypertension, coronary atherosclerosis, chronic liver disease and cirrhosis (all $p_{adj} < 3.43e-02$).
Conclusion: The PRS, composed of IR-associated loci identified in the UK Biobank, demonstrated significant associations with phenotypes consistent with IR and metabolic syndrome in the MGI, an independent cohort. These results underscore the specificity and generalizability of these genetic signals in identifying individuals at elevated risk for IR and related metabolic conditions. Identifying such high risk individuals can help us to target treatments for prevention or treatment of IR related diseases to those most at risk.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5096F: Integrating functional genomic annotations improves prediction performances of colorectal cancer polygenic risk score across different racial and ethnic groups

Authors:

M. Thomas[1], S. Kundu[2], E. Kotler[3], L. C. Sakoda[4], C. Qu[1], R. S. Steinfelder[1], E. A. Rosenthal[5], S. L. Schmit[6], C. Gignoux[7], J. C. Figueiredo[8], J. R. Huyghe[1], L. L. White[9], I. Lansdorp Vogelaar[10], G. P. Jarvik[11], D. A. Corley[12], U. Peters[1], A. Kundaje[13], L. Hsu[1]; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Stanford Univ., Stanford, CA, [3]Stanford Univ., Menlo Park, CA, [4]Kaiser Permanente, Oakland, CA, [5]Univ. of Washington Sch. of Med., Seattle, WA, [6]Cleveland Clinic, Cleveland, OH, [7]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [8]Dept. of Med., Samuel Oschin Comprehensive Cancer Inst., Cedars-Sinai Med. Ctr., Los Angeles, CA, [9]Inst. for Hlth.Res., Kaiser Permanente Colorado, Aurora, CO, [10]Dept. of Publ. Hlth., Erasmus MC, Univ. Med. Ctr., Rotterdam, Netherlands, [11]Univ Washington Med Ctr., Seattle, WA, [12]Div. of Res., Kaiser Permanente Northern California, Oakland, CA, [13]Stanford Univ., Palo Alto, CA

Abstract:

Colorectal cancer (CRC) is a leading cause of cancer-related mortality, yet many CRC deaths are preventable through screening. Polygenic Risk Scores (PRS) offer the prospect of precision prevention by assessing an individual's risk of CRC and tailoring screening recommendations

accordingly. However, the currently available genome-wide CRC-PRS, which are based on training data from primarily European and Asian ancestry individuals (PMID: 36789420), have limited performance in Black or African American and Hispanic/Latino/a/e/x (henceforth referred to as Hispanic) populations. To further enhance PRS performance on these ancestral groups, we integrated functional information in the PRS development using the SBayesRC approach. This approach leverages functional genomic annotations to distinguish likely causal SNPs from non-causal ones in high linkage disequilibrium (LD), thereby improving polygenic prediction. Our training data included over 100,000 CRC cases and 180,000 controls of diverse ancestries, primarily non-Hispanic White and Asian, with approximately 1 million variants. We used 96 publicly available functional genomic annotations, which included binary coding, conserved regions, regulatory elements, LD-related annotations, as well 41 cell type- and cell state-specific annotations generated from single-cell ATAC-seq and RNAs-seq data from colorectal normal and tumor tissue. We evaluated and compared the functionally informed CRC PRS with existing genome-wide CRC PRS, using independent datasets consisting of 641-1,561 cases and 7,241-75,611 controls from Asian, Black/African American, Hispanic, and non-Hispanic White populations. The AUC estimates (95% CI) of this PRS are 0.65(0.63-0.66), 0.60(0.57-0.63), 0.65(0.62-0.67), and 0.65(0.64-0.66) for Asian, Black/African American, Hispanic, and non-Hispanic White, respectively. Racial and ethnic identification in these validation studies were self-reported. Compared to our existing genome-wide PRS, the newly proposed PRS improves the AUC by 0.02 in both Black/African American and Hispanic ancestral groups while performing similarly or slightly better in Asian and non-Hispanic White groups. Our initial results demonstrate that leveraging functional annotation in PRS development modestly improves polygenic prediction accuracy. Studies show that leveraging 7 million imputed common SNPs, compared to ~1 million HapMap3 SNPs, can further improve the prediction performances. In the future, we aim to expand our PRS by incorporating these additional SNPs with functional information to further improve PRS performance.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5097F: Integrating Methylation and Transcriptomic Analyses with Whole Genome Sequencing YieldsNovel Insights into SARS-CoV-2 in African Americans

Authors:

G. Chen, L. DeRoo, G. Goodney, A. Doumatey, j. zhou, A. Adeyemo, C. Rotimi, A. Gaye; Natl. Human Genome Res. Inst., Bethesda, MD

Abstract:

Background: The SARS-CoV-2 which caused the 2019 COVID-19 pandemic, continues to engender inquiries into the role of host genetic factors in disease susceptibility. Despite the identification of over 1000 genes potentially associated with SARS-CoV-2 infection and severity of disease, the intricate mechanisms connecting genotype to phenotype remain elusive and under intense investigation. To shed light on these mechanisms, we undertook an integrated analysis of data from whole-genome sequencing (WGS) of SARS-CoV-2 with methylation and transcriptomic analyses. **Methods**: The study includes 371 African American adults with WGS, 203 with DNA methylation data, and 321 with RNA sequencing data. All generated data originated from whole blood, with 135 samples overlapping across the datasets. Significant loci associated with SARS-CoV-2 were examined within the framework of methylation quantitative trait loci (mQTL) and expression quantitative trait loci (eQTL), enabling mediation analysis to assess the mediating effects between genetic variants and SARS-CoV-2 outcomes. **Results**: Four intronic variants within *PMF1* were identified in association with SARS-CoV-2, after adjusting for covariates including sex, age, Body Mass Index (BMI), hypertension, diabetes, and principal components in WGS analysis. Within the *PMF1* region, we identified 6 cis-eQTL, involving 4 SNPs and 2 mRNA transcripts, as well as 20 cis-mQTL, comprising 19 SNPs and 10 CpG sites, with p-value < 1×10^{-5}. In mediation analysis, no significant mediator, such as mQTL or eQTL were identified between genetic variants and susceptibility to SARS-CoV-2. However, we did observe that certain genetic variants within *PMF1*, exert an influence on SARS-CoV-2 susceptibility through a sequential series of mediators involving CpG sites (five mQTL within *ARHGEF2*, *LMNA*, *TMEM79*, *MEX3A,* and *MEF2D*), and two mRNA transcripts within *PAQR6,* and *GLMP*. **Conclusions**: Through our integrated multi-omics analyses, we have identified a set of genetic loci whose effects on SARS-CoV-2 susceptibility are modulated by changes in DNA methylation and mRNA expression. These findings offer valuable insights into potential mechanistic pathways that merit further exploration.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5098F: Integrating Single-Cell & Bulk Multi-Omics Data to Uncover Novel Drug Targets for Neurological and Psychiatric Disorders

Authors:

Z. Mao, Y. Xia, H. Wang, D. Clarke, M. Gerstein; Yale Univ., New Haven, CT

Abstract:

Background: Neurological and psychiatric disorders are widespread and disabling, yet effective treatments remains a challenge. Genetic research is crucial for understanding these disorders and developing personalized therapies, as two-thirds of FDA-approved drugs backed by human genetic evidence. Linking genetic variation to drug development requires understanding the intricacies of various brain cell types, which single-cell omic data now enables. These data are revolutionizing our understanding of how genetic variants impact brain function, leading to more targeted therapy regimens.
Methods: In this study, we utilized single-cell expression quantitative trait loci (scQTLs) across 17 distinct brain cell types from the PsychENCODE consortium (n=388) as instruments to link genetic variation to altered expression. We conducted two-sample Mendelian randomization to predict the impact of the genes on 21 neurological and psychiatric disorders. This was followed by a cross-estimation using Summary data-based Mendelian Randomization with bulk tissue data from multi-omics studies, including brainmeta eQTLs (n=2,865), brainmeta sQTLs (n=2,865), PsychENCODE bulk eQTLs (n=1,387), brainmeta mQTLs (n=1,160), brain pQTLs (n=793), and brain caQTLs (n=272). We then overlapped these colocalized genes with genes annotated as drug targets in resources such as DrugBank, ChEMBL, STITCH, DGIdb, and Ruiz to identify promising drug targets. For each drug target gene, we performed a type of variance partition analysis to estimate the relative contributions of individual, cell type, and tissue on the target gene's expression variability
Results and Conclusions: Our analyses revealed a potential causal link between the expression of 51 genes and specific disorders, including schizophrenia, bipolar disorder, major depressive disorder, Alzheimer's disease, Parkinson's disease, multiple sclerosis, and amyotrophic lateral sclerosis. We found that the expression variability of these genes exhibit greater contributions from different cell types relative to individual variation. Since these genes exhibit greater expression variability between cell types, therapies can be tailored to affect only the diseased cell types, thereby minimizing off-target effects in healthy cells. We identified 16 genes as potential drug targets for schizophrenia, as well as 3 genes (*ACE*, *ADAM10*, *MAPT*) for Alzheimer's disease, suggesting their suitability for further investigation in functional studies and clinical trials. This study highlights the value of integrating multi-omics data to uncover drug targets for neurological and neuropsychiatric diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5099F: Integration of GWAS and multi-omic QTLs identifies uncharacterized Atopic Dermatitis gene-biotype associations and pinpoint keratinocyte-linked causal candidates

Authors:

M. Oliva[1], M. Sarkar[2], M. March[3], A. Saeidian[3], C-L. Hsieh[1], F. Tang[1], R. Uppala[2], M. Patrick[2], L. Qinmengge[2], R. Bogle[2], M. Kahlenberg[2], D. Watson[3], J. Glessner[3], L. Tsoi[2], H. Hakonarson[4], J. Gudjonsson[2], K. Smith[1], B. Riley-Gillis[1]; [1]AbbVie, North Chicago, IL, [2]Univ. of Michigan, Ann Arbor, MI, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Here, we characterize Atopic dermatitis (AD) genetic architecture by the identification of genomic loci not previously associated with the disease, implicated molecular and cellular contexts, and deep characterization of putative causal genes in keratinocytes. We identified 101 genome-wide significant loci by GWAS meta-analysis on 12 cohorts comprising a total of 56,146 AD cases and 602,280 controls of European, Asian, African, American, and admixed ancestries. We unveiled 15 loci not previously associated with AD and identified population-enriched gene-associated fine-mapped alleles, e.g., for *TESPA1* and *NLRP10* in Finns and East Asians, respectively. This GWAS effort reports the largest number of AD-associated loci to date and highlights the value of cohorts from diverse ancestries to characterize AD genetic architecture. By GWAS-QTL colocalization of 297 expression, splicing, protein and DNAm QTL maps, we identified molecular links for most known and novel AD loci. We scored and prioritized per-locus candidate genes by integrating QTL evidence with coding variants, variant-to-gene predictions and AD phenotype annotations. Among loci not previously associated with AD, we observe top-scoring genes encompassing multiple instances of receptor-ligand interactions related to signaling pathways known to be causally implicated to AD. These include *IL22RA2* for IL-22 signaling, and *IL6, IL6R* and *SOCS3* for IL-6 signaling. These results support the utility of the employed prioritization approach in identifying putative AD-causal genes. Since skin barrier dysfunction is a defining feature of AD, we aimed to identify genes playing a causal role in AD through pathogenic effects in keratinocytes. To functionally characterize 22 gene candidates, we generated a comprehensive array of skin and keratinocyte functional assays, including a cross-body skin scRNA-seq atlas, keratinocyte multi-timepoint differentiation expression data, and keratinocyte-derived eQTL maps and KOs across AD-relevant stimulation conditions in interleukin pathways. The integrated cross-assay results pinpoint *RORA*, *CEBPA*, *AQP3*, and *ANK3* as AD-keratinocyte-linked causal candidates and highlight keratinocyte-subtype and context specificities. Importantly, the array of unique keratinocyte-derived resources generated

herein can be further utilized by the scientific community to better understand AD pathobiology linked to the disruption of the epidermal barrier. Combined, **our work provides new insights into the etiology of AD by harnessing multiple genetic and functional approaches to unveil the mechanisms by which AD-associated variants impact genes and cell types.**

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5100F: Integration of GWAS and multi-omic QTLs identifies uncharacterized COVID-19 gene-biotype and phenotype associations

Authors:

J. Degner[1], J. Resztak[1], R. Hammond[1], J. S. Lee[2], E. King[1], B. Riley-Gillis[1], M. Oliva[3]; [1]Abbvie, Des Plaines, IL, [2]Northwestern Univ. IT Res. Computing and Data Services, Evanston, IL, [3]AbbVie, North Chicago, IL

Abstract:

To better understand COVID-19 pathobiology and to prioritize treatment targets, we sought to identify human genes influencing genetically driven disease risk and severity, and to identify additional organismal-level phenotypes impacted by pleiotropic COVID-19-associated genomic loci. To this end, we performed ancestry-aware, trans-layer, multi-omic analyses by integrating recent COVID-19 Host Genetics Initiative genome-wide association (GWAS) data from six ancestry endpoints - African, Amerindian, South Asian, East Asian, European and meta-ancestry - with quantitative trait loci (QTL) and GWAS endpoints by colocalization analyses. We identified associations for 47 COVID-19 loci with 307 GWAS trait endpoints and observed a highly variable (1-435 endpoint colocalizations) degree of pleiotropy per COVID-19 locus but a high representation of pulmonary traits. For those, directionality of effect mapped to COVID-19 pathological alleles pinpoints to systematic protective effects for COPD, detrimental effects for lung adenocarcinoma, and locus-dependent effects for IPF. Among 64 QTL-COVID-19 colocalized loci, we identified associations with most reported (47/53) and half of unreported (19/38) COVID-19-associated loci, including 9 loci identified in non-European cohorts. We generated colocalization evidence metrics and visualization tools, and integrated pulmonary-specific QTL signal, to aid the identification of putative causal genes and pulmonary cells. For example, among likely causal genes not previously linked to COVID-19, we identify desmoplakin-driven IPF-shared genetic perturbations in alveolar cells. Altogether, we contributed to providing molecular and phenotype links to the genetic architecture of COVID-19 risk and severity phenotypes, by further characterizing previously reported loci and by providing novel insights for uncharacterized loci.

Disclosure Statement: MO, JR, RH, EK, BRG and JD are employees of AbbVie. JSL is an employee at Northwestern University IT Research Computing and Data Services. The design, study conduct, and financial support for this research were provided by AbbVie. AbbVie participated in the interpretation of data, review, and approval of the publication.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5101F: Integrative Dissection of Causal Variants at Four Novel Alzheimer's Disease Risk Loci

Authors:

J. Regan[1], A. Diwadkar[1], L. Carrel[2], D. Liu[1]; [1]Pennsylvania State Univ., Hershey, PA, [2]Pennsylvania State Coll. of Med., Hershey, PA

Abstract:

Alzheimer's disease (AD) is a progressive, incurable disorder, strongly influenced by genetic risk factors. Genome-wide association studies (GWAS) have identified numerous disease risk genetic loci for AD. Most of the associated variants are non-coding and the causal variants and risk genes in the identified loci remain unclear. Existing statistical approaches have limited success in addressing this challenge. Integrating statistical analysis and experimental validation offers a promising approach to improve causal variant identification. As few studies have utilized this approach in AD, there is a need to perform comprehensive functional validation of AD GWAS risk loci to identify true causal regulatory variants. This approach will help elucidate AD pathophysiology and improve drug target discovery.
We developed a novel mixed-effects meta-regression model IMAC (Integrated Meta-Analysis for Cross-trait GWAS) that accommodates phenotype heterogeneity and adjusts for sample overlap. We applied it to publicly available AD GWAS summary statistics from four studies, replicated 56 known loci, and identified 4 novel loci. To prioritize causal variants and target genes at these loci, we used an integrative computational and experimental approach. For computational strategies, we performed co-localization with eQTLs from brain tissues and cell types. Further, we identified active enhancer and promoter regions using promoter capture Hi-C and ATAC-Seq datasets in brain tissue and neural cell types and used them to identify regulatory variants and their target genes. Lastly, we utilized gene prioritization models like Polygenic Priority Score (PoPs). We successfully identified putative causal genes that are supported with multiple lines of evidence, e.g., the lead SNP in our novel locus Chr5q34 successfully co-localized with the *FAM193B* gene in multiple bulk brain tissue eQTL datasets and overlapped with the activate promoter region for the same gene in both microglia cells and dorsolateral prefrontal cortex

(DLPFC) tissue. *FAM193B* has been previously identified as associated with sex-specific genetic risk in AD progression.
We are performing massively parallel reporter assay (MPRA) to screen causal variants in our novel loci variants and assess their regulatory and allelic functions. 190 variants from these loci including SNPs in LD (r^2>0.8) were tested in HMC3 and HEK293 MPRAs to identify variants that had cell specific, allelic, and regulatory effects. Overall, our approach demonstrates robust and comprehensive strategies for validating regulatory disease risk variants.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5102F: Interplay between genetic predisposition and plasma metabolome in relation to dementia risk and healthy aging

Authors:

Y. Liu[1,2,3], Y. Li[2], X. Gu[2], F. Wang[2,3], C. M. Vyas[4], D. Dong[1,3], Y. Li[2], Y. Zhang[2], O. Zeleznik[1], J. H. Kang[1], M. Wang[1,2], O. I. Okereke[1,2,4], A. Eliassen[1,2], P. Kraft[2,5], M. J. Stampfer[1,2], D. D. Wang[1,2,3]; [1]Channing Div. of Network Med., Brigham and Women's Hosp., Boston, MA, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [5]Natl. Cancer Inst., Rockville, MD

Abstract:

Metabolic dysregulation contributes to cognitive aging and dementia. Recent genome-wide association studies (GWAS) have described the genetic architecture of the plasma metabolome, a comprehensive quantification of metabolism, and many aging-related phenotypes. However, our understanding of the interplay between genetics and metabolome in cognitive health and aging remains limited. Here, we conducted an integrated analysis of genetic and metabolomic data from 4,271 participants in the Nurses' Health Study with a prospective follow-up of up to 34 years. We profiled plasma metabolomics (400 metabolites) using liquid chromatography-mass spectrometry, collected data on incident dementia, and assessed cognitive function longitudinally. We identified 63 significant interactions of *APOE4* genotypes or other common Alzheimer's disease (AD) variants with metabolites in relation to dementia risk (false discovery rate $q < 0.05$), e.g., more pronounced positive associations for trimethylamine-N-oxide in *APOE4* noncarriers ($q_{interaction} = 0.04$) and dimethylguanidino valeric acid in rs2154481-C (*APP*) carriers ($q_{interaction} = 0.01$) and inverse association for betaine in *APOE4* noncarriers ($q_{interaction} = 0.04$) with dementia risk. The gene-metabolite interactions were broadly consistent for both dementia risk and cognitive decline. Incorporating metabolomic features moderately improved

the performance for predicting 20-year risk of dementia based on lifestyle and genetic factors (Harrell's C-index increased from 0.84 to 0.85, $p = 0.59$). Utilizing published GWAS summary statistics, we performed Mendelian randomization analyses to examine the causal associations between 802 plasma metabolomic features and 15 aging-related phenotypes (GWAS sample sizes: 6,618-1,126,563). We identified 318 causal relationships for 173 metabolomic features and 15 phenotypes ($q < 0.05$). Notably, we found that a genetically predicted higher ratio of plasma levels of benzoate and oleoyl-linoleoyl-glycerol (18:1/18:2) was casually associated with a lower risk of AD as well as longevity and overall healthy aging. A higher plasma level of kynurenine was associated with increased frailty and poorer healthy aging outcomes. A higher 4-acetamidobutanoate level was linked to poorer cognitive function, while a higher bilirubin level was associated with healthy aging and increased lifespan. Our findings provide insights into the gene-metabolite interplay in relation to dementia and healthy aging, highlighting novel biological mechanisms and the clinical potential of measuring different metabolite biomarkers for individuals with varying genetic predispositions.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5104F: Investigating genetic effects on height shrinkage in the East Asian population

Authors:

K. Kitamura[1], M. Koido[1], Y. He[1], T. Takahashi[1], N. Itokawa[1], N. Asanoma[2,3], N. Shiomi[2,3], A. Fujino[3], Y. Kamatani[1], M. Seyama[4], T. Suzuki[5], Y. Murakami[5,6]; [1]Lab. of Complex Trait Genomics, Dept. of Computational Biology and Med. Sci., Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Tokyo, Japan, [2]NTT Smart Data Sci. Ctr. and NTT Computer and Data Sci. Lab., Nippon Telegraph and Telephone Corp., Tokyo, Japan, [3]Bio-Med. Informatics Res. Ctr. and NTT Basic Res. Lab., Nippon Telegraph and Telephone Corp., Tokyo, Japan, [4]Med. Business Planning Office, Nippon Telegraph and Telephone Corp., Tokyo, Japan, [5]Project Div. of Genomic Med. and Disease Prevention, The Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan, [6]Dept. of Molecular Biology, Inst. for Advanced Med. Sci., Nippon Med. Sch., Tokyo, Japan

Abstract:

Height shrinkage is caused by age-related factors such as osteoporosis and is associated with increased cardiovascular disease risk and mortality. Although the genetic effects on height change within personhave been investigated in European populations, the impact in populations of other ancestries and differences in effects between populations of different ancestries remain unexplored. We examined the genetic influence on height shrinkage in a

working-age cohort comprised of the employees of Nippon Telegraph and Telephone Corporation (NTT) group companies who underwent annual health examinations. This longitudinal cohort consisted of up to 36,389 participants with whole genome genotypes, including data from 1990 to 2022 (range in age: 17 - 67 years). The data of all participants was collected with written informed consent. After stringent quality control, genotype data were imputed using the 1000 Genomes Project reference panel (phase3v5, N=2,504). For longitudinal height data of 17,059 individuals measured at least 10 times after age 40, we performed a regression analysis for each individual to estimate baseline height and linear height change. Associations between the two estimated traits and genetic variants were examined by genome-wide association study (GWAS) using linear mixed models. For baseline height, 24 genome-wide significant genetic loci were found, all of which were known from previous cross-sectional height GWAS. One genome-wide significant genetic locus was found near *TBX5* for height change. This was also a known genetic locus from the cross-sectional height GWAS but was not found in the previous GWAS for height change in the European population. Genomic Restricted Maximum Likelihood (GREML) analysis was conducted to estimate the SNP-based heritability and genetic correlations between baseline height and height change. The SNP-based heritability for baseline height was 0.563 [95%CI: 0.514 - 0.612], comparable to that of 0.524 [95%CI: 0.508 - 0.540] in the European population. The SNP-based heritability for height change was 0.125 [95%CI: 0.078 - 0.172], which was nearly 10-fold higher than that of 0.015 [95%CI: 0.003 - 0.027] in the European population. No significant genetic correlation was found between baseline height and height change (0.115 [95%CI: -0.022 - 0.252], p=0.098), which was consistent with results in the European population (-0.130 [95%CI: -0.271 - 0.011], p=0. 072). In summary, new genetic loci associated with height shrinkage have been discovered, and SNP-based heritabilities for height shrinkage differ between European and East Asian populations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5105F: Investigating sex-specific genetic susceptibility in autoimmune diseases

Authors:

P. Sauer[1], B. Waxse[1], S. Goleva[1], A. Williams[1], T. Ferrara[1], C. Zeng[1], J. Keaton[1], T. Tran[1], D. Shittu[1], A. Awan[1], H. Mo[1], M. Davis[2], J. Denny[1]; [1]NIH, Bethesda, MD, [2]Brigham Young Univ., Provo, UT

Abstract:

Females are four times more likely to develop an autoimmune disease. This increased risk may occur because some genetic risk factors may have more of an impact in females. This study

tested this hypothesis with the All of Us Research Program, a biobank of diverse participants with genome sequencing data linked to Electronic Health Records (EHRs).
We used diagnosis codes and lab values to phenotype eight autoimmune diseases with at least 100 female and male cases each (Celiac disease, Hashimoto's thyroiditis, Multiple sclerosis, Sjogren's syndrome, Graves' disease, Rheumatoid arthritis, Psoriasis, Crohn's disease). These datasets were majority female, with Sjogren's Syndrome having the highest female-to-male ratio (~5:1).
We performed a genome-wide association study (GWAS) for autoimmune status. The strongest associated signals were a single nucleic polymorphism (SNP) in *HLA-DQB1* (p=1.49e-14, OR=1.30) and a SNP intronic to *XG* (p=3.32e-25, OR=0.68). *XG* is an escape gene on the X chromosome that encodes the XG blood group antigen. It's in strong linkage-disequilibrium (LD) with *CD99*, another escape gene that is a known target for controlling T-cell mediated autoimmune diseases.
In a female-only GWAS for autoimmune status, the strongest signal was a SNP in *HCP5* (p=1.10e-9, OR=1.34), another gene in the Human Leukocyte Antigen region. Additionally, this stratification identified a signal in a SNP intronic to *STRN* (p=1.08e-8, OR=1.46) as significant only in females; this SNP (chr2:36925229) is 13 nucleotides away from a splice region. *STRN* encodes a subunit for the enzyme PP2A which interacts with T cell activation and is known to affect sex-specific phenotypes. *STRN* and *HCP5* in males were not significant (*STRN*: p=0.35, OR=0.917; *HCP5*: p=0.03, OR=1.147).
We also performed a GWAS for each autoimmune disease and observed that six of eight diseases had strongly associated signals with genes in chromosome 6 as well as the X chromosome's pseudo-autosomal regions. Comparing the strongest associated signals in the autoimmune status cohort, *XG* was associated with six diseases tested while *HLA-DQB1* was associated with four diseases.
These findings show how sex-stratification may highlight other genes not observable without female-only cohorts. Investigating genetic risk factors both specific to sex chromosomes, as well as different sexes, may highlight new insights on autoimmune disease genetic risk factors.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5106F: Investigating the association between vitamin D levels and uterine fibroids using polygenic scores

Authors:

A. Pigg Akerele[1], J. Piekos[2], J. Hellwege[3], T. Edwards[4], D. Velez Edwards[5]; [1]Meharry Med. Coll., Nashville, TN, [2]Vanderbilt Univ., Nashville, TN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]VUMC, Nashville, TN, [5]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

Uterine leiomyomata (fibroids) are the most common benign tumors in premenopausal individuals, having an estimated cumulative prevalence of up to approximately 70%. These genetically complex tumors are characterized by the presence of fibrous masses growing in and on the smooth muscle of the uterus. Symptoms include dysmenorrhea, heavy or abnormal uterine bleeding, and anemia, pelvic pain, and infertility. Epidemiological studies have identified strong associations with uterine fibroids and various risk factors, including vitamin D deficiency. Vitamin D supplementation is hypothesized to reduce the risk and severity of uterine fibroids and existing evidence suggests that vitamin D blood levels and uterine fibroids are inversely correlated. Furthermore, various studies have explored vitamin D supplementation as a potential therapy for uterine fibroids. The objective of this study is to use polygenic scores to examine the relationship between vitamin D blood levels and uterine fibroids. In this preliminary analysis, we used published vitamin D polygenic scores to investigate if vitamin D levels are associated with uterine fibroids in 1,394 self-reported, White samples from Vanderbilt University Medical Center's biorepository (BioVU). We conducted 10-fold cross-validation logistic regression analyses to test for individual associations of vitamin D and uterine fibroids. The samples were split into two groups where 80% of samples were used for training and 20% were used for validation. Three models were applied to the polygenic score and we evaluated the predictive performance of each polygenic score using area under the curve (AUC) values calculated from receiver operator characteristic (ROC) curves to assess their capabilities of predicting cases of uterine fibroids. The model achieved an AUC value of 0.65 (95% confidence interval [CI] 0.62-0.68)), suggesting moderate predictive ability, but the polygenic score itself was insignificant. Excluding covariates did not change the predictive ability of the model, providing AUC value of 0.65 (95% confidence interval [CI] 0.61-0.68)). Our data, although preliminary, support the hypothesis that vitamin D levels may be associated with uterine fibroids. Future directions of this work will entail the inclusion of additional samples from BioVU and from the Electronic Medical Records and Genomics Network (eMERGE). We will also examine the relationship between Vitamin D and uterine fibroids in more diverse populations, specifically African American and Asian individuals. This expansion aims to improve predictive power and to potentially yield more significant findings.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5108F: Investigating the genetic link between bipolar disorder and gut microbial taxa through polygenic risk score analysis

Authors:

J. Ciarcia[1], A. De Pins[2], The Bipolar Disorder Working Group of the Psychiatric Genomics Consortium (PGC-BIP), J. Xu[1], L. Huckins[1]; [1]Yale Univ., New Haven, CT, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Introduction: In recent years, growing evidence has emerged suggesting a link between gut microbiota and psychiatric disorders including bipolar disorder (BD). While studies have shown differences in the diversity and taxonomic composition of the gut microbiome between BD cases and controls, no study to date has investigated this link via a polygenic risk approach. The aim of this project is to identify gut microbial taxa linked to BD by examining whether polygenic risk scores (PRS) for microbial taxa abundance and diversity are associated with BD status. **Methods:** Using summary statistics from the largest human gut microbiome GWAS (N=18,340) by the MiBioGen Consortium, microbiome PRS for 211 microbial taxa and 2 measures of alpha diversity (Shannon and Simpson indices) were calculated with PRSice software at 10 P-value thresholds ranging from 5×10^{-8} to 1 across 34 PGC-BIP cohorts (N=74,366) and meta-analyzed across cohorts. Genetic principal components were included as covariates to address population stratification. Multiple testing was addressed with FDR correction, and taxa with negative heritability estimates were excluded. **Results:** 77 of the 211 taxa (36.5%) and the Shannon index reached significance at the nominal level for at least one P-value threshold tested. 132/211 taxa had positive heritability estimates, including all 77 of the nominally significant taxa. While no taxon remained significant after FDR correction, several approached FDR significance (P_{FDR}= 0.06629762) using the threshold with the lowest meta-analysis P-value for each taxon, including 1 order-level taxon (Enterobacteriales, P_{meta}=0.00428), 1 family-level (Enterobacteriaceae, P_{meta}=0.00428), and 7 genus-level (Eubacterium ruminantium, P_{meta}=0.00452; Akkermansia, P_{meta}=0.00084; Anaerofilum, P_{meta}=0.00425; Eggerthella, P_{meta}=0.00235; Roseburia, P_{meta}=0.00173; Ruminococcaceae, P_{meta}=0.00334; and Turicibacter, P_{meta}=0.00298). **Conclusion**: Our results suggest a possible host genetic link between BD and microbiota. We plan to re-run these analyses using PRS-CS, analyze across subphenotypes of BD, and explore the potential effects of alcohol or psychotropic medication use. We also aim to run similar PRS analyses using cases and controls for major depressive disorder, schizophrenia, and PTSD, psychiatric disorders that are genetically correlated with BD. These analyses will allow us to address the questions of whether there are any gut microbiome PRS specific to certain psychiatric disorders or shared across disorders, and understanding this genetic link could facilitate future development of microbiota-directed therapeutics.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5109F: Investigating the role of *APOE* and retinal biomarkers for predicting Alzheimer Disease

Authors:

N. Moore[1], Y. Song[1], A. Lynn[1], K. Miskimen[1], W. Wang[1], R. Laux[1], S. Fuzzell[1], S. Hochstetler[1], L. Caywood[2], J. Clouse[2], S. Herington[2], P. Wang[1], Y. Liu[1], M. Nittala[3], S. Sadda[3], D. Stambolian[4], P. Ogrocki[5], A. Lerner[5], J. Vance[2], M. Cuccaro[6], W. Scott[2], M. Pericak-Vance[7], J. Haines[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Miami, Miami, FL, [3]Doheny Eye Inst., Los Angeles, CA, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [6]John P. Hussman Inst. for Human Genomics, Miami, FL, [7]Univ. of Miami Miller Sch. of Med, Miami, FL

Abstract:

Alzheimer Disease (AD), the most common form of dementia, and Age-Related Macular Degeneration (AMD), are degenerative diseases affecting the brain and eye respectively. These diseases share the risk locus *APOE*, where the e2 allele provides protection and e4 confers risk in AD, while the opposite is true in AMD. Similarly, they share amyloid beta aggregates, located in the brain of AD individuals and in the drusen in the retina of AMD individuals. Other retinal biomarkers such as retinal thickness have been associated with AD previously. By evaluating the relationship between retinal biomarkers and AD/dementia in the context of *APOE* we can better understand the potential role of retinal biomarkers as noninvasive risk predictors for AD and dementia. We tested six retinal biomarkers (subfoveal choroidal thickness, presence/absence of drusen, and drusen area and volume in 3mm and 5mm circles) and AD (based on detailed neuropsychological exams) and dementia (defined as a 3MS score of &lt 88). The subfoveal choroidal thickness and drusen datasets consisted of 821 and 1114 Midwestern Amish individuals respectively. Right and left eyes were analyzed in separate analyses. A result was considered significant if it had a $p < 0.05$. Logistic regressions for the five quantitative biomarkers, with the outcome AD/dementia and adjusted for age, sex, and *APOE*, were not significant. However, the logistic regression using only the binary drusen presence/absence as a predictor and no covariates, resulted in an increased risk of AD/dementia when drusen is present (OR = Right: 3.1 & Left: 2.4; p &lt 0.001). Once the regression was adjusted by age, *APOE*-e2 and *APOE*-e4 status the results became non-significant (Right: OR = 1.4 & p = 0.16; Left: OR = 1.0 & p = 0.87), as expected since the risk of AD/dementia increases with increased age which is supported by these results. Upon further investigation of the individuals with drusen, the majority (71.1%) were CU. We examined a potential relationship between

drusen and AD/dementia by analyzing individuals with *APOE* genotype 3|3 (individuals that are not predisposed to or protected from AD or AMD) in a chi-square test, assessing the independence of drusen presence/absence and AD/dementia presence/absence. The results were significant for both eyes suggesting these two variables are not independent (Right: X^2= 27.7 & p = 1.39e-07; Left: X^2= 13.9 & p = 1.88e-04). When this analysis was repeated on individuals with one or two *APOE*-e4 alleles, the results were no longer significant, most likely due to the small sample size of 284 individuals. These results provide evidence that drusen presence could influence the risk of AD/dementia in *APOE* 3|3 individuals.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5110F: Investigating the Role of ETFDH and UQCRC1 Mutations in Hypertrophic Cardiomyopathy Using a Zebrafish Model ⭐

Authors:

D. Manoel[1], K. Al-Shafai[1], A. Al-Naama[1], M. Makhlouf[1], D. Al-Ansari[2], R. Hasnah[1], E. Aboumoussa[1], S. Jacob[1], L. Saraiva[1]; [1]Sidra Med., Doha, Qatar, [2]Natl. Heart and Lung Inst., Imperial Coll., London, United Kingdom

Abstract:

Hypertrophic cardiomyopathy (HCM) is a prevalent familial heart condition characterized by left ventricular myocardial thickening and reduced ventricular space, resulting in impaired cardiac function. Despite extensive research, the genetic and molecular mechanisms underlying HCM pathophysiology remain incompletely understood. Here, we report a case of familial HCM with two homozygous missense mutations identified in the ETFDH and UQCRC1 genes. These mutations are predicted to impact the protein products crucial for mitochondrial metabolic and energetic functions. To investigate the causality of these mutations in HCM and explore potential therapeutic strategies, we utilized genetically modified zebrafish larvae as a model system. Phenotypic assessments were conducted at cardiac, metabolic, and transcriptomic levels. Our findings suggest that the ETFDH mutation observed in the patient contributes to HCM pathogenesis. Cardiac analysis following ETFDH knockdown using morpholinos revealed increased myocardial thickness alongside reduced ventricular area and volume. Metabolomic profiling showed altered acylcarnitine levels resembling Multiple Acyl-CoA Dehydrogenation Deficiency (MADD). Transcriptomic analysis corroborated cardiac and metabolic alterations, highlighting affected pathways and potential therapeutic targets. We investigated the therapeutic potential of Riboflavin, known to be deficient in MADD patients, in our zebrafish models. Results demonstrate Riboflavin's efficacy in reverting metabolic and cardiac phenotypes

caused by ETFDH knockdown in zebrafish larvae, suggesting its potential as a preventive therapy for ETFDH-associated HCM. Overall, our study contributes to personalized diagnostics, potential preventive treatments, and reveals a novel link between mitochondrial function and cardiomyopathy.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5111F: Investigation of the molecular mechanism underlying IBD phenotypes through a large meta-analysis of scRNAseq data

Authors:

P. Pibiri[1], R. Oelen[1], W. T. C. Uniken Venema[1], E. Bigaeva[1], M. G. P. van der Wijst[1], M. D. Paraskevopoulou[2], X. Zhang[2], K. N. Faber[1], L. Franke[1], E. A. M. Festen[1]; [1]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [2]Takeda Pharmaceuticals, Cambridge, MA

Abstract:

Background and Objectives:Inflammatory Bowel Disease (IBD) is one of the most prevalent gastrointestinal diseases worldwide. However, despite the advent of new therapeutic agents, patients deal with debilitating symptoms and high rates of therapy non-response. Single-cell RNA sequencing (scRNAseq) studies, such as Martin et al. (Cell 2019) on Crohn's disease (CD) and Smillie et al. (Cell 2019) on Ulcerative Colitis (UC), have provided insights into the molecular basis of the disease mechanism, yet their limited statistical power due to small cohort sizes impedes a comprehensive understanding. To acquire further insight into the molecular differences of IBD phenotypes, and to understand the mechanisms of non-response to therapy, we propose a large-scale meta-analysis by combining and harmonizing independent IBD scRNAseq and genotype studies.
Methods:We developed a standardized pipeline on gut mucosa scRNAseq and genotype data to align quality control, preprocessing, cell type annotation, differential proportion analysis, differential expression, and eQTL mapping. We tested this pipeline to harmonize 4 smaller IBD studies (2 UC and 2 CD), to be subsequently integrated into a bigger IBD study object.
Results:In our preliminary results, we harmonized and integrated the two CD and the two UC scRNAseq datasets. Our preliminary results demonstrate successful harmonization and integration, yielding extensive datasets comprising over 485,000 cells from more than 150 individuals for CD and over 290,000 cells from more than 30 individuals for UC.
Conclusion:We created a standardized pipeline for harmonizing scRNAseq and genotype data applied to four IBD datasets. By expanding our analysis to include international datasets, we aim

to enhance statistical power and deepen our understanding of IBD pathomechanisms, with the ultimate goal of improving therapeutic strategies for patients.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5112F: Large-scale Multi-Ancestry Epigenome-wide Analysis of BMI in the Million Veteran Program

Authors:

K. Markianos[1], **P. Schreiner**[2,1], B. Gorman[2,1], B. Despard[2,1], M. Francis[3,1], Y. Shi[3], S. Pyarajan[3]; [1]Ctr. for Data and Comp. Sci., VA Boston, Boston, MA, [2]Booz Allen Hamilton, McLean, VA, [3]Ctr. for Data and Comp. Sci., VA Boston, Boston, MA

Abstract:

Background. Obesity is a risk factor for many diseases including diabetes, cardiovascular disease, certain cancers, among other serious health problems. The cause of obesity can be attributed to genetic factors, activity levels, diet, the ability to afford healthy food, access to healthcare, among many other factors. The prevalence of obesity around the world has more than doubled since 1990. As of 2022, 2.5 billion surveyed adults worldwide were identified as at least overweight, corresponding to 43% of adults 18 and older. Disparities in the relative prevalence of obesity are known to exist based on an individual's age, gender, and racial-ethnic background, which have yet to be comprehensively characterized. The Million Veteran Program (MVP) is a mega-biobank initiated by the United States Department of Veteran Affairs in 2011, with combined multi-omic and clinical data in a diverse set of individuals. **Methods.** We present a single cohort of 850K Infinium MethylationEPIC array data for 42,703 individuals in the Million Veteran Program classified in three continental ancestries (EUR, AFR, and HIS) using the HARE method. We performed an epigenome-wide association study using meffil, with p-value correction in bacon. We then aggregated single probe associations from CpG site clusters to find differentially methylated regions (DMRs). **Results.** In analyses stratified by ancestry, we identified 6,384 EUR, 2,368 AFR, and 124 HIS significant individual probe associations with BMI on autosomal chromosomes. Those corresponded to 4,780 DMRs in EUR, 1,755 in AFR, and 112 in HIS that are significantly associated with BMI. In this presentation, we summarize the genomic context of each significantly associated DMR and report the proximity to genes that are likely to be affected by methylation perturbations. We identify pathway associations with BMI for genes implicated in fat mass growth, cytokine signaling, cholesterol metabolism, mitochondrial function, and immune function. Further, we report novel, ancestry specific methylation associations with BMI. **Conclusion.** Our results provide multi ancestry BMI

associations in the largest single cohort currently available. These findings significantly advance our understanding of factors that drive disparities in rates of obesity.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5113F: Late-Stage Skeletal Muscle Transcriptome in Duchenne muscular dystrophy shows a BMP4-Induced Molecular Signature

Authors:

H. Sothers[1], X. Hu[1,2], D. Crossman[1], Y. Si[1], M. Alexander[1,2], M-L. McDonald[1,3], P. King[1,3], M. Lopez[1,2]; [1]Univ. of Alabama - Birmingham, Birmingham, AL, [2]Children's of Alabama, Birmingham, AL, [3]Birmingham VA Med. Ctr., Birmingham, AL

Abstract:

Duchenne muscular dystrophy (DMD) is a fatal genetic condition caused by loss of function mutations in the *DMD* gene, leading to lack of dystrophin, a crucial component for the stabilization of striated muscle cells. In DMD, the muscle progressively deteriorates over time , leading to a myriad of health issues and eventual death by the 20s-30s, despite recent improvements in treatment. Beyond *DMD* gene mutations, biological pathways dysregulated in DMD are poorly understood, especially as it progresses to advanced stages. Among recognized pathways, the TGFβ family of cytokines is implicated in muscle wasting in DMD. The TGF receptor transcription factor, *SMAD9*, has been found to be upregulated in severely affected DMD muscle and is known to be regulated through *BMP4*. Previous work has demonstrated *BMP4* stimulation can induce upregulation of *SMAD9* transcription in a C2C12 cell culture model. We hypothesized that *BMP4* signaling could be a driver of aberrant transcriptomic signatures underlying the pathology of DMD in human skeletal muscle. To test this hypothesis, RNA-sequencing data was generated from human skeletal muscle biopsy samples (n=6: 3 with and 3 without DMD) and C2C12 cell cultures (n = 6 : 3 with and 3 without exogenous BMP4 stimulation). Both datasets were filtered for low-counts and assessed for outliers prior to normalization and differential expression analysis performed using DESeq2. Weighted gene correlation network analysis (WGCNA) was used to discover modules of correlated transcripts and assess module preservation between the independent networks. A total of 3,048 and 5,291 genes were differentially expressed (FDR<0.05) in human muscle and C2C12 cells, respectively. WGCNA found 4 modules significantly associated with the DMD trait in humans and 2 with BMP4 stimulation in the C2C12 cells, respectively. The turquoise module was the only significantly upregulated module in both and module preservation analysis indicated moderate preservation (ZSummary score = 4.7) between the human tissue and mouse

cell networks. This overlapping gene signature contained 1,027 genes: *SERPING1* and *Aff3* were among the hub genes. Other transcripts that were highly upregulated in the DMD muscle and members of the preserved module included *SMAD9* , *ADAMTS3*, and *HCAR2*. Overall, we demonstrate a transcriptome-level similarity between advanced stage DMD and BMP4-exposed muscle cells *in vitro*, suggesting a potential mechanistic link in advanced disease pathology. Next steps will include further investigating the function of *SMAD9* in this pathway as it could improve our understanding of the role of TGF signaling in advancing DMD pathology.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5114F: Leukodystrophy and Cancer in Sisters Suggests Genetic Pathogenesis: Case Series

Authors:

S. Goyalpatel[1], J. Sampson[2], M. Han[2]; [1]Los Altos High Sch., Los Altos Hills, CA, [2]Stanford Univ., Palo Alto, CA

Abstract:

Objective: To describe a case series of two sisters presenting with leukodystrophy and cancer suggesting an underlying genetic cause.
Background: Leukodystrophies are a group of rare genetic disorders affecting the white matter of the brain. Five common genetic disorders account for more than fifty percent of leukodystrophies, but many others have unknown genetic causes. Cancers, on the other hand, are hereditary 5-10% of the time.
Case Description: Case #1: 55 y/o female, history of epilepsy, thyroid cancer, breast cancer (May 2024), seen for incidental findings of confluent progressive white matter lesions on MRI brain. She had no neurological symptoms and neurological examination including a comprehensive neuropsychological testing was normal.
Case #2: 59 y/o female, history of multiple cancers (thyroid cancer 2006, breast cancer 2006, lung cancer 2021), seen in 2007 for incidental findings of white matter lesions on MRI brain during breast cancer screening. Blood tests and CSF evaluation were negative. Repeat MRI showed progression of white matter lesions. She had no neurological symptoms, but had mild abnormalities on neurology examination and neuropsychiatric testing.
Family history revealed a maternal uncle with white matter changes on MRI without cancer, brother with testicular cancer and undiagnosed autism, and multiple family members with cancer.
Genetic Testing: Genetic testing in both patients showed the same single nucleotide nonsense

heterozygous LAMA2 gene mutation c.9192_9193del (p.Phe3065Cysfs*7). Even though LAMA2 mutations can cause white matter lesions, two pathogenic variants are required. Whole Genome Sequencing for Case #1 showed pathogenetic mutation c.562C>T (p.Gln188Ter) which is known to cause familial focal epilepsy. No genetic mutations have been found to date that explain the cancers and leukodystrophy.
Discussion: We present a case series of two sisters with leukodystrophy and cancer, and a family history of cancer, white matter lesions on MRI, and other neurological disorders (autism and epilepsy) suggesting an underlying genetic etiology. Genetic testing, including whole genome sequencing, has been indeterminate to date. Next steps include enrollment into the undiagnosed disease network.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5115F: Leveraging multi-omics data to explore genetic etiology for peripheral artery disease and coronary artery disease

Authors:

J. Hu[1], C. I. Chaar[2], H. Zhao[3], A. Dewan[1]; [1]Yale Sch. of Publ. Hlth., New Haven, CT, [2]Yale Sch. of Med., New Haven, CT, [3]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract:

Genetic factors are important in the development of arterial diseases such as peripheral artery disease (PAD) and coronary artery disease (CAD). However, the variants identified by genome-wide association studies (GWASs) failed to fully explain the disease heritability, as epigenetic factors were not accounted for/ included in the studies. To enhance the detection of disease-associated genetic factors, we utilized transcriptome-wide association studies (TWASs) and proteome-wide association studies (PWASs), which aggregate variant-level effects to strengthen the power of association analyses. Employing GWAS summary statistics and established prediction models, we applied S-PrediXcan—a method that inferred relationships between genetically predicted gene/protein expression levels and disease risk. Our objective was to identify novel disease-associated genes, defined as those located in regions that were not previously linked to diseases in the GWAS Catalog. Additionally, we explored the biological functions of these genes using gene-set enrichment analysis (GSEA) of co-expressed genes and phenome-wide association analysis (PheWAS). We maintained a significance threshold of 0.05 and controlled for multiple comparisons using the False Discovery Rate (FDR) procedure. Our study analyzed 7,155 and 1,289 genes for PAD in TWAS and PWAS respectively, identifying 34 and 6 genes significantly associated with PAD. Upon conditioning on nearby risk SNPs, 28 and 4

genes respectively remained significant. There were no shared genes between TWAS and PWAS results. Notably, four novel PAD risk genes were identified from TWAS. GSEA indicated these genes may be related to the fibrillar collagen system, the regulation of DNA transcription and RNA translation, mitochondrial functions, inflammation, and olfactory functions. PheWAS further found some of these genes were associated with neutrophil-to-lymphocyte ratio (NLR) and systemic immune-inflammation index (SII). For CAD, conditional analysis in TWAS and PWAS yielded 192 and 85 significant genes respectively, with three genes shared between analyses. Among the identified disease-associated genes for CAD, three genes were novel. These genes, based on GSEA, might be related to cytochrome C activity, the ubiquitin-proteasome system, macroautophagy, cell maturation, and the immune system. Our results underscored the utility of leveraging multi-omics data to unravel the genetic etiology of complex diseases. Further analyses will be conducted to explore the overlapping multi-omics risk factors for PAD and CAD.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5116F: Leveraging the *All of Us* Research Program to better understand the genetic architecture of opioid use disorder spectrum behaviors

Authors:

A. Moore[1], B. Webb[2], N. Gaddis[1], O. Corradin[3], B. Maher[4], A. Palmer[5], E. Chesler[6], D. Hancock[1], J. A. Bubier[6], D. Jacobson[7], L. Davis[8], S. Sanchez-Roige[9], V. Troiani[10], E. O. Johnson[1]; [1]RTI Intl., Durham, NC, [2]RTI Intl., Research Triangle Park, NC, [3]MIT, Cambridge, MA, [4]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [5]Univ. of California San Diego, La Jolla, CA, [6]The Jackson Lab., Bar Harbor, ME, [7]Oak Ridge Natl. Lab., Oak Ridge, TN, [8]Vanderbilt Univ Med Ctr., Nashville, TN, [9]UCSD, La Jolla, CA, [10]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Attempts to understand the genetic underpinnings of opioid use disorder are stymied by limited sample size, insufficient genetic diversity, difficulty identifying appropriate control groups, and challenges in characterizing the spectrum of opioid use behaviors. The *All of Us* (AoU) Research Program contains extensive behavioral and clinical data coupled with genotype information, with targeted recruitment of individuals traditionally underrepresented in biomedical research. We used ICD-9/10 codes for opioid abuse or dependence from electronic health records (EHRs) and survey self-report indicating lifetime opioid misuse to identify opioid misuse cases. After requiring high-quality genotyping data and genetically predicted ancestry group determinations, we calculated autosomal SNP-heritability (h^2-SNP) for each of three ancestry groups from AoU:

African (AFR) (n_{cases}=5,647), American (AMR) (n_{cases}=3,680), and European (EUR) (n_{cases}=15,761) using genotype data from the Global Diversity Array. We considered four control definitions, which varied in the stringency of inclusion criteria. Unscreened controls were anyone who was not a case. Screened controls required the presence of EHR data and no opioid misuse in the Lifestyle Survey data. Exposed controls were defined as screened controls who had an opioid medication on at least two separate calendar days in their EHR. The fourth control group excluded high-risk individuals from screened and exposed controls. We defined high-risk opioid medication exposure as at least ten calendar months having an opioid medication within a 12-month period. Control group sizes ranged from 11,637 to 44,342 for AFR, 10,366 to 36,554 for AMR, and 42,907 to 110,927 for EUR. Heritability calculations were performed using the Bolt-REML algorithm implemented in Bolt-LMM software and adjusted for sex at birth, year of birth, and 16 principal components of ancestry. Results comparing opioid misuse to unscreened controls within genetically predicted ancestry groups in AoU yield h^2-SNP of 0.030 (SE=0.010) for AFR, 0.165 (0.011) for AMR, and 0.063 (0.004) for EUR ancestry. As the control group definition became more stringent, regardless of ancestry, h^2-SNP estimates increased to a maximum of 0.094 (0.029) for AFR, 0.255 (0.026) for AMR, and 0.099 (0.008) for EUR, respectively, despite decreasing sample size. Ongoing work will conduct genome-wide association testing in the full whole genome sequencing cohort (up to N = 245,394) and use other data present in AoU to consider further examination of the genetic architecture of opioid misuse, including sex differences, comorbid conditions, and patterns of use.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5117F: Linkage and association of cognitive SuperAging to known protective Alzheimer's locus at *WDR12* in the Midwestern Amish

Authors:

D. Dorfsman[1,2], M. B. Prough[1], A. Gulyayev[1], L. J. Caywood[1], J. E. Clouse[1], S. D. Herington[1], S. H. Slifer[1], L. D. Adams[1], R. A. Laux[3,4], Y. E. Song[3,4], A. Lynn[3,4], S. L. Fuzzell[3,4], S. D. Hochstetler[3,4], K. Miskimen[3,4], L. R. Main[4,5], P. Wang[3], Y. Liu[4], N. Moore[4], P. Ogrocki[6], A. J. Lerner[5], J. M. Vance[1,2], M. L. Cuccaro[1,2], J. L. Haines[3,4], M. A. Pericak-Vance[1,2], W. K. Scott[1,2]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Dr. John T. Macdonald Fndn. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [4]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [5]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [6]Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH

Abstract:

Despite the increasing prevalence of Alzheimer's disease (AD) and dementia with age, some individuals maintain mid-life cognitive performance into their later years. A unique group, termed SuperAgers (SA), are ≥ 80 years old and possess episodic memory capabilities that match or exceed expectations for middle-aged adults. Genetic studies of SA are crucial for uncovering heritable factors underlying exceptional cognition in older adults. This study incorporates linkage and association-based methods in an Amish cohort to identify genetic variants that explain SA.

Adult Amish participants older than 75 (N=529) underwent a comprehensive neuropsychological assessment and were genotyped using Illumina-based platforms. SA (N=83) were defined as those aged ≥ 80 with episodic memory performance ≥ the mean for ages 35-44 and performance in non-episodic memory tasks (e.g., language, executive functioning) within one standard deviation or greater for their age. All subjects were connected in a single pedigree, which was partitioned into 16 subpedigrees suitable for affected-only dominant linkage allowing for heterogeneity (HLOD) and non-parametric (LOD*) linkage analysis using MERLIN. A subset of 7,366 genotyped markers in low linkage disequilibrium was used for linkage analysis. Genetic variants imputed using TopMed and located in linked regions of interest were tested for association with SA using generalized linear mixed models in GENESIS. Association tests adjusting for age, sex, top 2 principal components, and genetic relatedness were performed using SA as the outcome and two separate comparison groups: age-matched cognitively unimpaired non-SA individuals (CU-nonSA, N=157) and individuals with AD (N=40).

In the nonparametric analysis, a LOD* score > 3 was observed on chromosome 2 (peak at 204 Mb). In the parametric linkage analysis, HLOD scores > 3 were found on chromosomes 1 (44 Mb), 2 (203 Mb), 7 (30 Mb), and 20 (17 Mb). The overlapping evidence of linkage to chromosome 2 in both analyses, and the colocalization of the parametric linkage peak with a known locus protective for AD at *WDR12* reported by the ADSP, prioritized variants within this gene for further analysis (n=69 variants). The SA to AD comparison revealed an interval containing 46 highly correlated, nominally associated variants enriched in SA (OR=3.1; 95% CI: 1.1-8.6), spanning the protective locus and demonstrating a consistent effect. A separate nominal association was observed when comparing SA to CU-nonSA (rs112065260, OR=0.51; 95% CI: 0.30-0.84).

This study identified linkage and association of SA to a known AD protective locus. Ongoing studies aim to further elucidate these findings.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5119F: Linking R-loop dysregulation to anxiety-related behaviors in fragile X-associated neuropsychiatric disorders (FXAND)

Authors:

Y. Jin[1], Y. Cao[2], Y. Hou[1], N. Rivera[2], B. Yao[2], P. Jin[2]; [1]Emory Univeristy, Atlanta, GA, [2]Emory Univ., Atlanta, GA

Abstract:

The premutation of the fragile X messenger ribonucleoprotein 1 (*FMR1*) gene is characterized by an expansion of the CGG trinucleotide repeats (55 to 200 CGGs) in the 5' untranslated region and increased levels of *FMR1* mRNA. The elevated risks of psychiatric symptoms have been reported in male and female carriers of PM alleles, which have been termed fragile X-associated neuropsychiatric disorders (FXAND). The molecular mechanisms underlying fragile X premutation involve both CGG repeat RNA and a repeat-associated non-AUG (RAN) translation product called FMRpolyG peptide. However, the molecular link between expanded CGG repeats and neuropsychiatric symptoms remains largely unknown. To understand the molecular etiology of FXAND, we have performed a battery of behavioral testing using two transgenic mouse models of fragile X premutation, the mouse line that could express both fragile X premutation CGG RNA and FMRpolyG protein (FMRpolyG) and the mouse line that only expresses CGG repeats RNA (CGG_RNAonly). In the contextual/cued fear conditioning tests, the CGG_RNAonly mice, but not FMRpolyG, exhibited excessive contextual fear responses and enhanced cued fear memory when exposed to a cue, indicating anxiety-related behavioral changes. At the molecular level, we performed genome-wide profiling of R-loops, an RNA:DNA hybrid structure, in the prefrontal cortex (PFC) of CGG_RNAonly mice and control animals. Our analysis identified over 4,600 misregulated R-loop regions associated with approximately 2,600 genes. Notably, we observed a substantial increase in R-loop formation in CGG_RNAonly mice, many highly associated with genes involved in abnormal fear regulation and anxiety-related behavior pathways. We also performed gene expression profiling and observed significant overlap between genes showing alterations in both R-loop formation and gene expression, indicating that altered R-loops might significantly impact gene expression. Furthermore, we examined the distribution of distances between the altered R-loops and CGG repeats in the mouse genome and found that these gain-of R-loop regions in CGG_RNAonly mice are frequently found near expanded CGG repeats, suggesting a potential direct interaction between expanded CGG repeats and R-loops formation. These results highlight the potential critical roles of R-loops in the molecular etiology of FXAND. The interplay between dysregulated R-loops, expanded CGG repeats, and altered gene expression may contribute to the behavioral phenotypes observed in FXAND.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5121F: Local genetic covariance analysis with lipid traits identifies novel loci for early-onset Alzheimer's Disease

Authors:

N. Ray[1], J. Bradley[2], E. Yilmaz[1], C. Kizil[1], J. T. Kurup[1], E. R. Martin[3], H-U. Klein[1], B. W. Kunkle[3], D. A. Bennett[4], P. L. De Jager[1], Alzheimer's Disease Genetics Consortium, G. W. Beecham[3], C. Cruchaga[2], C. Reitz[1]; [1]Columbia Univ., New York, NY, [2]Washington Univ., Saint Louis, MO, [3]Univ. of Miami, Miami, FL, [4]Rush Alzheimer's Disease Ctr., Chicago, IL

Abstract:

Background: The genetic component of early-onset Alzheimer disease (EOAD), accounting for ~10% of all Alzheimer's disease (AD) cases, is largely unexplained. Recent studies suggest that EOAD may be enriched for variants acting in the lipid pathway. **Objective:** To examine the shared genetic heritability between EOAD and the lipid levels by genome-wide multi-trait genetic covariance analyses. **Methods:** Summary statistics were obtained from the GWAS meta-analyses of EOAD by the Alzheimer's Disease Genetics Consortium (*n*=19,668) and five blood lipid traits (total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, non-high-density lipoprotein cholesterol, and triglycerides) by the Global Lipids Genetics Consortium (*n*=1,320,016), and genetic covariance analyses were performed via SUPERGNOVA. Genes in linkage disequilibrium (LD) with top EOAD hits in identified regions of covariance with lipid traits were scored and ranked for causality by combining evidence from gene-based analysis, AD-risk scores incorporating transcriptomic and proteomic evidence, eQTL data, eQTL colocalization analyses, DNA methylation data, and single-cell RNA sequencing analyses. **Results:** Local genetic covariance analyses identified 5 regions of covariance between EOAD and at least one lipid trait. Gene prioritization nominated six likely causative genes at these loci: *ANKDD1B*, *CUZD1*, *MGAT5*, *MAPK3*, *ALDOA*, and *MS4A64*. **Conclusion:** The current study identified genetic covariance between EOAD and lipids, providing further evidence of shared genetic architecture and mechanistic pathways between the two traits.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5122F: Longitudinal proteomics of body mass index (BMI) in the Cameron County Hispanic Cohort (CCHC) ⭐

Authors:

V. Buchanan[1], H. Highland[2], J. Below[3], J. McCormick[4], S. Fisher-Hoch[5], M. Lee[6], E. Frankel[7], R. Roshani[7], W. Zhu[7], A. Howard[8], K. Young[9], M. Graff[9], K. North[9]; [1]Univ. of North Carolina - Chapel Hill, Westfield, IN, [2]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [3]Vanderbilt Univ Med Ctr., Nashville, TN, [4]UTH SPH Brownsville, Brownsville, TX, [5]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [6]UT Hlth.Houston, Brownsville, TX, [7]Vanderbilt Univ., Nashville, TN, [8]Univ. of North Carolina, Chapel Hill, NC, [9]Univ North Carolina, Chapel Hill, NC

Abstract:

The public health burden of obesity is immense. While proteomics studies of adiposity are increasing, temporal variations in these associations, particularly in underrepresented populations, remain understudied. We aim to identify protein level changes that are associated with body mass index (BMI) changes in the Cameron County Hispanic Cohort (CCHC) of participants living in the border communities of Brownsville and Laredo, TX. Proteomics profiling (Olink Explore) assayed abundance of 2,921 proteins from fasting samples drawn from a subset of CCHC. Of 249 adults ages 18-75 (65% female, mean $BMI_{baseline}$: 32.4), 216 had both BMI and proteomics at 2 or more time points, ~5.2 years apart. We investigated whether protein level changes were associated with BMI changes over time by first using linear mixed effects regression and a 2 degree of freedom (2DF) test comparing the following to a null model that excludes the BMI_{ij}*follow-up $time_{ij}$ interaction term:

$protein_{ij}$ = BMI_{ij} + follow-up $time_{ij}$ + BMI_{ij}*follow-up $time_{ij}$ + sex + $age_{baseline}$ + PEERs1-10_{ij} + PCs1-3 + b_{oi}

where (i) and (j) are indexed for each participant and time point to account for repeated measures, with follow-up time in years, 10 PEER factors, and the first 3 genetic principal components (PCs) of ancestry to limit possible confounding. Although 442 proteins were associated with BMI (FDR $p < 0.05$ in null model), no protein changes were significantly associated with BMI changes over time. We then restricted to BMI gainers and found that change for one protein (leptin, LEP) was significantly associated with BMI change (FDR p for BMI_{ij}*follow-up $time_{ij}$ = 0.047; FDR p for 2DF= 0.027). Leptin is a hormone that plays an important role in energy homeostasis by binding to receptors that activate multiple pathways involved in signaling the sensation of satiety and modulating nutrient absorption. We then performed linear regression to investigate how baseline BMI may influence change in protein levels (baseline level subtracted from follow-up level):

Δprotein = $BMI_{baseline}$ + follow-up time + sex + $age_{baseline}$ + PEERs1-10 + PCs1-3 + $protein_{baseline}$

and found that baseline BMI was significantly (FDR p <0.05) associated with Δprotein for 2 proteins (IL1R and ADM) in the overall group, and 11 proteins (HGF, IL1RN, BCAN, LGALS9, CCL22, ADAM12, RNASE1, FABP4, CRHBP, RNASE4, and YAP1) in BMI gainers. Although many proteins are associated with BMI, these associations may vary over time with body weight changes. These time-dependent variations may improve our understanding of underlying molecular profiles of obesity and how we identify potential modifiable targets for obesity prevention and intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5123F: Mapping multivariate associations among polygenic factors, brain, behaviors and psychopathologies in children

Authors:

Y. Joo[1], E. Lee[2], B-G. Kim[2], G. Kim[2], J. Seo[2], J. Cha[2]; [1]Samsung Med. Ctr. / Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

Preadolescent childhood is a critical period for cognitive and behavioral development, characterized by significant structural and functional brain changes. This study aims to explore the genetic architecture underlying brain development, cognition, and behavior in children by integrating genome-wide polygenic scores (GPSs), brain imaging-derived phenotypes (IDPs), and cognitive-psychological data using advanced multivariate statistical techniques.We analyzed data from over 8,600 multiethnic preadolescent children from the Adolescent Brain and Cognitive Development (ABCD) study, employing sparse generalized canonical correlation analysis (SGCCA) to identify covariation patterns among GPSs of 29 traits, 7 brain imaging modalities (structural MRI, diffusion MRI, resting-state and task-based fMRI), and 266 cognitive and psychological phenotypes. Our findings reveal a low-to-moderate SNP-based heritability in most IDPs, consistent with previous research indicating lower heritability in children's brains compared to adults.Significant positive associations were found between cognition-related GPSs (educational attainment (EA), cognitive performance (CP), intelligence quotient (IQ)) and neuroanatomical features such as total grey matter volume, left ventral diencephalon volume, and right accumbens surface area. Conversely, GPSs for ADHD, depression, and neuroticism showed negative associations with these brain regions. Additionally, educational attainment GPSs were positively associated with brain activation during the N-back task, while BMI GPSs showed a positive correlation with fractional anisotropy (FA) between the cerebellum cortex and amygdala. Notably, GPSs for educational attainment and cannabis use were negatively

associated with FA in the same regions.Our study provides a comprehensive map of the relationships between genetic profiles, neuroanatomical diversity, and cognitive-behavioral traits in preadolescence. The findings highlight the substantial genetic contributions to cognitive variability and the nuanced genetic underpinnings of mental and behavioral phenotypes. By leveraging SGCCA, this research offers new insights into the complex interplay between genetics, brain development, and cognition in children, paving the way for future studies to explore the biological basis of cognitive and psychological development.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5124F: Meta-analysis of Genome Wide Association Studies identifies genetic loci that affect steatosis and steatohepatitis

Authors:

L. Yin[1], Y. Chen[1], C. Raut[1], A. Oliveri[1], C. Still[2], X. Chu[2], G. C. Wood[2], G. Gerhard[3], E. Speliotes[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]The Ctr. for Obesity and Metabolic Res., Geisinger Obesity Inst., Danville, PA, [3]Lewis Katz Sch. of Med. at Temple Univ., Philadelphia, PA

Abstract:

Background: Steatotic liver disease (SLD) affects ~35% of individuals in the United States. Accumulation of fat in the liver can progress to steatohepatitis or fibrosis. Previous GWAS studies have identified common variants that explain about 25-35% of heritability of SLD measured using radiology. Radiology assessment of SLD cannot determine steatohepatitis and fibrosis, but they can be quantitated on histology; genetic analysis of these traits can identify determinants of steatotic disease progression.
Methods: In a Geisinger Bariatric sample, an expert pathologist quantitated hepatic steatosis, ballooning, and lobular inflammation in liver histology samples. We calculated the NASH Activity Score (NAS) from these measures. The samples were genotyped on Infinium CoreExome-24 v1.4 BeadChip array and imputed to TOPMed reference panel r2-1.0.0. We carried out a GWAS analysis of steatosis or NAS with imputed SNPs in an additive genetic model controlling for age, sex, and 10 principal components using linear regression in SAIGE v1.2.0. In the Michigan Genomics Initiative (MGI, PMID:36819667), we carried out text mining on pathology and radiology reports where we identified individuals with liver “steatosis” or “steatohepatitis” or “NASH”. We also identified individuals with an ICD-10 code K75.81 (NASH). Controls were individuals that did not have these diagnoses. MGI samples were genotyped on a customized Illumina Infinium Global Screening Array and imputed to TOPMed. We carried out a GWAS analysis of steatosis or steatohepatitis with imputed SNPs in an additive genetic model

controlling for age, sex, and 10 principal components using logistic regression in SAIGE v1.2.0. We meta-analyzed steatosis and separately steatohepatitis results from Geisinger and MGI using a sample size and direction of effect weighted approach in METAL. Fine mapping of causative SNPs was performed utilizing the sum of single-effects (SuSiE) regression model in susieR.
Results: 55,747 individuals of European ancestry were included in the study. Meta-analysis identified 3 SNPs near the genes *PNPLA3, SUGP1, APOC* that associate with steatosis and 2 SNPs near *PNPLA3, HAPLN4* that associate with steatohepatitis at genome-wide significant levels. Using SuSiE we identified 3 credible sets that include missense variants in *PNPLA3, TM6SF2*, and *APOE*, and 2 credible sets that associate with steatohepatitis that contain missense variants in *PNPLA3* and *TM6SF2*.
Conclusion: We identified genes that drive SLD progression. Future work involves increasing sample size to increase our power to find more genes with similar effects.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5125F: Mitochondrial gene signatures across the clinical spectrum of age-related macular degeneration

Authors:

J. Slone[1], T. Shwani[1], J. Vu[1], W. Liu[1], I. Kim[2], T. Shaikh[3], L. A. Owen[4], M. Deangelis[1]; [1]Univ. at Buffalo, Buffalo, NY, [2]Mass Eye & Ear Infirmary, Boston, MA, [3]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Abstract:

Age-related macular degeneration (AMD) is a polygenic disorder and leading cause of irreversible blindness in the aging population. This progressive neurodegenerative condition impacts the macula, the part of the eye responsible for our central vision. Current therapeutic interventions are limited in their effectiveness, highlighting the need to further elucidate disease mechanism for suitable biomarkers/therapeutics. Prior studies and our recent work profiling the transcriptome across clinical stages of AMD (n=27 unique donor eyes) implicate mitochondria and its self-contained genome (referred to as mitochondrial DNA, or “mtDNA”) in disease pathophysiology. Unlike nuclear DNA, mtDNA is maternally inherited, differs between tissue and cells, and has a high mutation rate. Mutations and variation in mtDNA have been implicated in other complex conditions, including Alzheimer’s disease; however, the precise mechanisms governing its role in AMD are unknown. Thus, we present here preliminary genomic analysis of mtDNA’s role in AMD across intermediate AMD (iAMD) and late stages of disease—geographic atrophy (GA) and neovascular (NEO)—compared to age-matched controls.

Utilizing our well-phenotyped donor eye repository and leveraging our transcriptomic datasets, we focused on: mtDNA sequence heteroplasmy, mtDNA haplogroups, and nuclear-mitochondrial genome interactions. We identified a statistically significant divergence in mtDNA copy number between NEO compared to GA, and normal eyes, suggesting differences in mtDNA dynamics between the different disease states. Moreover, we genotyped peripheral blood samples corresponding to 172 donor eyes using Illumina's HumanOmni2.5-8 BeadChip v1.2 and identified mtDNA haplogroup subclades and variants unique to each clinical subtype of AMD. We then used gene set enrichment analysis (GSEA) to investigate hallmarks related to nuclear-encoded mitochondrial genes (defined by MitoCarta3.0) in our previously published transcriptomic dataset (n=27 eyes). We found high enrichment in iAMD as compared to normal and NEO, with the top hallmarks being "mtDNA maintenance" and "FE-S-containing proteins". Significant genes were validated in a 2nd bulk RNAseq dataset (n=85 donor eyes). Ingenuity Pathway Anlaysis (IPA) revealed interconnections between our 26 overlapping genes, with TP53 emerging as a top upstream regulator. In summary, this is the first study of its kind profiling mtDNA in well-characterized, disease-affected tissues, providing novel insights into mitochondrial dysregulation in AMD.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5126F: Modeling the genomic architecture of anthropometrics and adiposity across the lifespan

Authors:

C. Arehart[1,2], M. Lin[3,4], R. Gibson[1,2], S. Raghavan[3,4], C. R. Gignoux[3,4], M. A. Stanislawski[3,4], A. D. Grotzinger[1,2], L. M. Evans[1,2]; [1]Univ. of Colorado, Boulder, CO, [2]Inst. for Behavioral Genetics, Boulder, CO, [3]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [4]Div. of BioMed. Informatics and Personalized Med., Aurora, CO

Abstract:

Obesity-related conditions are among the leading causes of preventable death and are increasing in prevalence worldwide. However, body size and body composition are challenging phenotypes to characterize due to their heterogeneity between sexes and differing health risks based on adipose distribution. Human body size and body composition are complex traits linked to many environmental, lifestyle, and genetic influences with variation throughout an individual's life course and across individuals in a population. Through an exploratory factor analysis and confirmatory factor analysis, we constructed a 4-factor genomic structural equation model (SEM) using 18 genome wide association studies (GWASs) as indicator variables. This

SEM achieved a reasonable balance between model fit and model parsimony (CFI = 0.94 and an SRMR = 0.11) and unveiled shared and distinct components of the genetic architecture underlying differences in birth size, adult size, weight distribution, and adiposity. First, we used multivariate GWAS and fine-mapping transcriptome wide association study (TWAS) methods to identify genes associated with the factors at the genomic and transcriptomic levels. Second, we characterized the factors' differing patterns of genetic enrichment across physiological systems such that F1, F2, and F3 showed no enrichment in nervous tissues while F4 was solely enriched for gene expression profiles in the brain. Third, our phenome wide association study (pheWAS) of the polygenic risk scores (PRSs) for each factor in an independent sample (Colorado Center for Personalized Medicine Biobank, N = 25,240) identified downstream health outcomes, with F4 predicting many adverse health-associations even when conditioned on body mass index (BMI). Finally, we characterized the factors' genetic correlations with obesity-related traits and examined the druggable genome by constructing a bipartite drug-gene network to identify viable therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5127F: Multi-Factor Analysis of Nicotine Metabolism, Genomics, and Epigenomics in American Indians from the Strong Heart Study

Authors:

K. Ferrier[1], C. R. Musser[1], C. A. Sherman[1], K. M. Aagaard[1], T. H. Wroblewski[2], R. F. Tyndale[3], M. J. Chenoweth[3], N. Franceschini[4], S. A. Cole[5], A. Navas-Acien[6], A. Domingo-Relloso[6], A. J. Kupsco[6], K. G. Claw[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [3]Univ. of Toronto, Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Texas BioMed. Res. Inst., San Antonio, TX, [6]Columbia Univ., New York, NY

Abstract:

Background: American Indians and Alaskan Natives (AIAN) show the highest prevalence of tobacco use compared to all other racial/ethnic groups in the US. Genetic variants of CYP2A6 influence nicotine metabolism and smoking behaviors; slower metabolism is associated with reduced tobacco consumption and easier smoking cessation. However, few studies have been conducted on nicotine metabolite ratio (NMR; a biomarker of CYP2A6 activity) and the potential genomic and epigenomic factors that may contribute to smoking-related behaviors in AIAN communities. Multi-factor analysis (MFA) facilitates high-dimensional data reduction of quantitative and qualitative variables simultaneously. MFA of American Indians (AI) from the

Strong Heart Study (SHS) with phenotypic, genomic, and epigenomic data identified significant relationships among factors that will help guide future investigations of NMR and smoking behaviors in this community. Methods: SHS is a longitudinal cohort study of cardiovascular disease in 12 AI tribes. Genotype, methylation, and NMR data were measured from blood samples. Individuals with all three data types and information on age, biological sex, BMI, and smoking status (n=518) were used in a MFA with: 5 genetic principal components (PCs); 10 methylation PCs; cotinine (COT), trans-3'-hydroxycotinine (3HC) and nicotine concentrations (ng/mL); NMR (COT/3HC); and Houseman estimated blood cell proportions. Results: Dimension (Dim.) 1 of the MFA explains 14.0% of the total variance with the primary contributing factors being COT concentration, smoking status, and geographic region. Dim. 2 (10.5%) is described by proportion of neutrophils and the 2nd methylation PC. Dim. 3's (9.2%) greatest contributing factors were genetic PC1 and geographic location. Dim. 4 (8.9%) is explained by NMR, followed by gender, genetic PC1, and geographic location. Dim. 5 (7.4%) is explained by age and methylation PC1. Overall, this analysis highlighted a confounding imbalance in sample sizes across geographic location and collinearity between blood cell proportions and methylation data that will need to be accounted for in future statistical models. The results also provide evidence of a correlation between NMR and genetic factors that will require further investigation. Conclusions: The results of this exploratory analysis of NMR, smoking-related phenotypes, basic demographic information, and genomic and epigenomic factors helped clarify relationships between different factors that will enable more appropriate statistical modeling in future investigations of NMR and smoking-related phenotypes and behaviors in the SHS cohort of American Indians.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5128F: Myocardial Infarction Metabolomic Signatures in a Middle Eastern Cohort

Authors:

M. Elshrif[1], A. Shaar[1], A. El-Menyar[2,3], M. Al-Maadheed[4], V. Mohamed-Ali[4], M. Saad[1], J. Al Suwaidi[5]; [1]Qatar Computing Res. Inst. (QCRI) - HBKU, Doha, Qatar, [2]Clinical Res., Trauma & Vascular Surgery, Hamad Med. Corp., Doha, Qatar, [3]Dept. of Clinical Med., Weill Cornell Med. Coll., Doha, Qatar, [4]Anti-doping Lab Qatar, Doha, Qatar, [5]Dept. of Cardiology, Heart Hosp., Hamad Med. Corp., Doha, Qatar

Abstract:

Background: Despite considerable advances in the treatment of acute coronary syndrome (ACS), it remains the leading cause of morbidity and mortality worldwide. It is possible to accurately diagnose patients with irreversible injury secondary to myocardial infarction (MI) using several biomarkers. It is estimated that MI occurs in 23.3% of coronary artery disease (CAD) patients. Preventing MI is important for better prognosis and treatment of CAD patients. Metabolomics biomarkers have been shown to play a role in CAD and MI and could be used to identify individuals at risk before diseases occur. For Middle Eastern populations, there is a lack of metabolomics studies for MI despite the large population size in Middle East and the increasing prevalence of CAD and MI risk factors such as obesity and type 2 diabetes.**Methods:** In this study, we investigated 641 metabolites from the Qatar Cardiovascular Biorepository (QCBio) initiative, which comprises of 969 Qatari CAD patients. Within this cohort, 350 patients had MI and xx did not have MI. Moreover, 488 have T2D and 481 did not. Univariate and multivariate analyses were performed to identify metabolites associated with MI adjusting for BMI, sex, and age. Metabolite risk scores were developed to assess the predictive power MI.**Results:** Many metabolites were significantly associated with MI. We identified 58 metabolites with Bonferroni significance levels (0.05/641). This included taurine amino acid (b = -0.69 and $P = 8.96 \times 10^{-10}$), EDTA (b = 1.12 and $P = 1.36 \times 10^{-09}$), 1-(1-enyl-palmitoyl)-GPE lipid (P-16:0)* (b = -1.01 and $P = 1.88 \times 10^{-09}$) , mannose (b = 0.67 and 3.07×10^{-09}) and xanthine Nucleotide (b = -0.21 and $P = 1.53 \times 10^{-08}$). Preliminary results showed a good performance of metabolite risk scores to predict MI.**Conclusions:** Metabolomics profiling has the potential for early detection of metabolic alterations that precede clinical symptoms of MI within CAD patients. This early detection potential allows for timely interventions and improved management strategies for MI.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5129F: Noncoding de novo mutations in SCN2A are associated with autism spectrum disorders

Authors:

Y. Zhang[1], M. Ahsan[2], K. Wang[3]; [1]Children's Hosp. of philadelphia, Philadelphia, PA, [2]Children s Hosp. of Philadelphia, Philadelphia, PA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Coding de novo mutations (DNMs) contribute to the risk for autism spectrum disorders (ASD), but the contribution of noncoding DNMs remains relatively unexplored. Here we use whole genome sequencing data of 12,411 individuals (including 3,508 probands and 2,218 unaffected

siblings) from 3,357 families collected in Simons Foundation Powering Autism Research for Knowledge (SPARK) to detect DNMs associated with ASD, while examining Simons Simplex Collection (SSC) with 6383 individuals from 2274 families to replicate the results. For coding DNMs, *SCN2A* reached exome-wide significance ($p=2.06\times10^{-11}$) in SPARK. The 618 known dominant ASD genes as a group are strongly enriched for coding DNMs in cases than sibling controls (fold change=1.51, $p=1.13\times10^{-5}$ for SPARK; fold change=1.86, $p=2.06\times10^{-9}$ for SSC). For noncoding DNMs, we used two methods to assess statistical significance: a point-based test that analyzes sites with a Combined Annotation Dependent Depletion (CADD) score ≥15, and a segment-based test that analyzes 1kb genomic segments with segment-specific background mutation rates. The point-based test identified SCN2A as marginally significant ($p=6.12\times10^{-4}$) in SPARK, yet segment-based test identified *CSMD1*, *RBFOX1* and *CDH13* as exome-wide significant. When combining evidence from both coding and noncoding DNMs, we found that *SCN2A* with 11 coding and 5 noncoding DNMs exhibited the strongest significance ($p=4.15\times10^{-13}$). In summary, we identified both coding and noncoding DNMs in *SCN2A* associated with ASD, while nominating additional candidates for further examination in future studies.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5130F: Non-human reads in human whole-genome sequencing identify endogenous human herpes virus (HHV)-6B and blood anellovirus virome associated with autoimmune diseases and COVID-19 risk

Authors:

Y. Okada[1], N. Sasa[2], S. Kojima[3], K. Rie[3], T. Hasegawa[4], H. Namkoong[5], T. Hirota[6], W. Rei[2], K. Sonehara[1], R. Edahiro[2], Y. Shirai[2], M. Ishii[7], R. Koike[4], A. Kimura[4], H. Inohara[2], M. Fujimoto[2], S. Ogawa[8], T. Kanai[5], A. Morita[9], F. Matsuda[8], A. Kumanogoh[10], Y. Tanaka[11], K. Ohmura[12], K. Fukunaga[5], M. Tamari[6], S. Imoto[13], S. Miyano[4], N. Parrish[3]; [1]The Univ. Tokyo / Osaka Univ. / RIKEN, Tokyo, Japan, [2]Osaka Univ., Osaka, Japan, [3]RIKEN IMS, Tokyo, Japan, [4]Tokyo Med. and Dental Univ., Tokyo, Japan, [5]Keio Univ., Tokyo, Japan, [6]The Jikei Univ. Sch. of Med., Tokyo, Japan, [7]Nagoya Univ., Nagoya, Japan, [8]Kyoto Univ., Kyoto, Japan, [9]Nagoya City Univ., Nagoya, Japan, [10]Osaka Unviersity, Osaka, Japan, [11]Univ. of Occupational and Environmental Hlth., Kitakyushu, Japan, [12]Kobe City Med. Ctr. Gen. Hosp., Kobe, Japan, [13]The Univ. of Tokyo, Tokyo, Japan

Abstract:

Aberrant immune responses to viral pathogens contribute to pathogenesis. Whether viruses comprising human viromes influence pathological immune responses remains unclear. We re-analyzed non-human reads in human short-read whole-genome sequencing (WGS) datasets of 6,321 Japanese individuals, including those with autoimmune diseases (psoriasis vulgaris, rheumatoid arthritis [RA], systemic lupus erythematosus [SLE], pulmonary alveolar proteinosis [PAP], or multiple sclerosis) and coronavirus disease 2019 (COVID-19), or healthy controls. We systematically quantified two constituents of the blood DNA virome, endogenous HHV-6 (eHHV-6, overall prevalence = 0.60%) and anellovirus (12.3%, 1.0%, and 0.29% for anellovirus-detected, -infected, and -viremic). Subjects with eHHV-6B had higher risks of SLE and PAP (adjusted OR = 6.9, $P = 0.0061$ and OR = 7.0, $P = 0.0075$). Replication studies using All of Us 240k WGS datasets confirmed an elevated SLE risk of eHHV-6B (OR = 2.7, $P = 0.0020$ for European ancestry and OR = 2.4, $P = 0.0012$ for multi-ancestry). eHHV-6B and high SLE Disease Activity Index (SLEDAI) scores had strong correlations (mean SLEDAI = 30.5 and 6.0 for eHHV-6B-positive and -negative patients with SLE; $P = 1.3\times10^{-8}$). GWAS of eHHV-6B positivity (neHHV-6B-positive = 22 and neHHV-6B-negative = 216) and 30× PacBio HiFi long-read WGS of eHHV-6B carriers with SLE ($n = 4$) positioned the endogenization of the HHV-6B genome on chr22q, supporting the endogenous nature of HHV-6B. Plasma antibody epitope mapping (PhIP-seq) to HHV-6 peptides in 16 patients with SLE (neHHV-6B-positive = 4 and neHHV-6B-negative = 12) revealed that eHHV-6B results in distinctive immune induction in patients with SLE. Single-cell RNA sequencing of PBMCs from the same individuals investigated by PhIP-seq showed increased antiviral interferon-stimulated gene expression in eHHV-6B-positive patients with SLE. In addition, strong positive correlations between high anellovirus load and SLE, RA, and COVID-19 were observed (adjusted OR = 58.5, 15.4, and 6.2 for infected status, $P = 2.6\times10^{-8}$, 2.5×10^{-5}, and 0.0042). Although only a small number of individuals carry eHHV-6 or a high load of anellovirus, positivity appears to have a striking impact on disease risk and clinical phenotype and to be a new and potentially clinically-useful biomarker for these diseases. Our work unveils relationships between the human virome and autoimmune and infectious diseases and outlines a path through which the study of the human virome can be integrated into the genotype:phenotype analysis for which human WGS are often collected.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5131F: Novel Chromosome *2* locus associated with post-traumatic osteoarthritis of the knee in 47,396 cases and comparators

Authors:

M-L. McDonald[1,2], J. Richman[3,4], D. Dochtermann[5], C. Boer[6], R. Kalsoum[7], C. Nguyen[3], A. Papadaki[7], T. Barker[8,9], K. Briem[10], K. Hatzikotoulas[11], B. Hollis[12], L. Jostins-Dean[13], S. Lemmela[14], T. Ryhanen[14], FinnGen, R. Magi[15], e. reinmann[16,15], L. Southam[17], J. Zhou[18], U. Styrkarsdottir[19], H. Tiwari[20], Genetics of Osteoarthritis consortium, J. van Meurs[21], E. Zeggini[22], J. Singh[3,2], F. E. Watt[7]; [1]Univ. of Alabama, Birmingham, Birmingham, AL, [2]Birmingham VA Hosp. Canter, Birmingham, AL, [3]Univ. of Alabama at Birmingham, Birmingham, AL, [4]Birmingham VA Hosp. Canter, Bimingham, AL, [5]Ctr. for Data and Computational Sci., VA Boston Hlth.care System, Boston, MA, [6]Erasmus MC,, University Medical Center, Netherlands, [7]Ctr. for Inflammatory Disease, Dept. of Immunology & Inflammation, Imperial Coll. London, London, United Kingdom, [8]Wexner Med. Ctr., The Ohio State Univ., Columbus, OH, [9]Dept. of Orthopaedics, Univ. of Utah, Salt Lake City, UT, [10]Dept. of Physical Therapy, Univ. of Iceland, Reykjavík, Iceland, [11]Inst. of Translational Genomics, Helmholtz Zentrum München, German Res. Ctr. for Environme, Neuherberg, Munich, Germany, [12]Univ. of Cambridge, Cambridge, United Kingdom, [13]Univ. of Oxford, Oxford, United Kingdom, [14]Univ. of Helsinki, Helsinki, Finland, [15]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [16]Inst. of Translational Genomics, Helmholtz Zentrum München, German Res. Ctr. for Environment, Munich, Germany, [17]Inst. of Translational Genomics, Neuherberg, Germany, [18]UCLA, Encino, CA, [19]deCODE genetics, Reykjavik, Is, Iceland, [20]Univ Alabama at Birmingham, Birmingham, AL, [21]Erasmus Med Ctr, Rotterdam, Netherlands, [22]Helmholtz Munich /Helmholtz Zentrum Muenchen GmbH, Neuherberg / Munich, Germany

Abstract:

Background: Approximately 50% of those with a significant knee injury develop osteoarthritis, so called post-traumatic osteoarthritis (PTOA) of the knee. The health consequences of PTOA are well-described, but the genetic risk of developing OA after knee injury is not understood. Knowledge of genetic variants associated with PTOA would improve our understanding of this disease. To address this gap, we performed a genome-wide association study (GWAS) meta-analysis in knee PTOA cases and controls from 6 international studies participating in the Genetics of Osteoarthritis consortium: UK Biobank, DeCODE, FinnGen, Estonian Biobank, Heredigene and Million Veteran Program. **Methods:** Knee PTOA cases were defined as either an eligible acute knee injury and subsequent knee OA/PTOA code, or a knee PTOA 'case code' only. Comparators for this analysis were those with an eligible acute knee injury, with no subsequent knee OA/PTOA code on follow up. Eligible knee injuries occurred between age 16-50, due to acute trauma of an internal knee joint structure. A window of 6 months - 20 years between knee injury and knee OA code, or censoring/death was applied. Exclusions were defined, e.g., ever having inflammatory arthritis. Diagnostic codes were adjudicated and compiled from electronic health record systems, primarily ICD-9/10. Study stratified GWAS were performed in each cohort adjusting for age, sex and principal components controlling for ancestry and then

combined by fixed effects meta-analysis using METAL. Fine mapping was performed using SuSiE. **Results:** A total of 23,334 knee PTOA cases and 24,062 knee injured comparators were identified across the 6 cohorts, with a median follow up of 16 years after injury. For 8,930 knee PTOA cases identified by injury followed by OA codes, the weighted mean age at injury was 41. Knee PTOA was more common in males (74%). In the combined meta-analysis, a locus on chromosome 2 encompassing the *TEX41* gene was significantly associated with knee PTOA (rs28414958: OR =1.2 , SE= 0.034, P= 1.4E-8, MAF_{range} = 0.06 - 0.10, p_{het}= 0.97). The effect direction was the same in all six participating cohorts. The fine-mapping credible set included only variants within the *TEX41* gene which encodes a long noncoding RNA previously associated with calcific aortic valve disease, with no previous described genetic associations with OA. **Conclusion:** We report a novel locus encompassing the gene *TEX41* associated with knee PTOA. Ongoing research is focused on downstream follow up. Better understanding of molecular determinants of PTOA will enable its prediction, stratification, and prevention.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5132F: Novel genetic loci associated with knee osteoarthritis in 284,184 cases and controls in the Million Veteran Program and the All of Us Research Network

Authors:

C. Nguyen[1,2], A. Rocco[1,2], V. Srivasasainagendra[1,2], D. Dochtermann[3], H. Tiwari[1], J. Richman[1,2], F. Watt[4,5], J. Singh[1,2], M-L. McDonald[6]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Birmingham VA Hlth.Care System, Birmingham, AL, [3]Ctr. for Data and Computational Sci., VA Boston Hlth.care System, Boston, MA, [4]Ctr. for Inflammatory Disease, Dept. of Immunology & Inflammation, Imperial Coll. London, London, United Kingdom, [5]Ctr. for Osteoarthritis Pathogenesis Versus Arthritis, Kennedy Inst. of Rheumatology, Univ. of Oxford, Oxford, United Kingdom, [6]Univ. of Alabama, Birmingham, Hoover, AL

Abstract:

Background: In the U.S., there are currently 32.5 million adults with osteoarthritis (OA), a debilitating disease characterized by deterioration of the articular cartilage and underlying subchondral bone. The etiology of OA is poorly understood with only limited treatment options available including total joint replacement surgery. We aim to investigate genetic variants influencing knee OA within the Million Veteran Program (MVP) for discovery and All of Us (AoU) for replication, performing large-scale genome-wide association studies (GWAS) that may offer insights in disease mechanisms and risk prediction associated with susceptibility, heterogeneity,

and progression. **Methods:** Knee OA cases and controls were identified using ICD 9 and 10 Electronic Health Record (EHR) codes. Knee OA cases were required to have at least two OA codes 30 days apart and OA controls were excluded if they had codes for concomitant findings typically more prevalent in patients with osteoarthritis. MVP genetic variants were genotyped on the Affymetrix Axiom Biobank Array (MVP release 3*)* with additional variants imputed using the TOPMed reference panel. AoU genotypes were generated using the Illumina Global Diversity Array with whole-genome sequencing in the v7 release. In each study, both cosmopolitan and ancestry stratified analyses were performed in each cohort using SAIGE adjusting for age, sex, body mass index and principal components adjusting for population structure. **Results:** In the MVP cosmopolitan analysis, a total of seven genetic loci were identified as significantly associated with knee OA. Of these, two are novel with one encompassing the *IKBKB gene* (8:42281902 G/A, OR = 1.1, SE = 0.014, P = 4.3E-11) and the second encompassing the *DKK4* gene (8:42381579 G/A, OR = 1.1, SE = 0.014, P = 1.1E-8). Both the *IKBKB* (OR = 1.1, SE = 0.014, P = 5.9E-6) and the *DKK4* (OR = 1,1, SE = 0.025, P = 8.5E-5) loci replicated in AoU. Additional significant loci associated with knee OA were also identified in ancestry stratified analyses. **Discussion:** We report two novel gene loci encompassing the genes *IKBKB* and *DKK4* associated with knee OA. These findings may contribute to our understanding of the genetic basis of OA and provide potential targets for therapeutic interventions and more personalized approaches to OA management.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5133F: Novel rare variant associations with late-life cognitive performance

Authors:

A. Regelson[1,2], D. Archer[1,2], A. Durant[1,2], S. Mukherjee[3], M. L. Lee[3], S-E. Choi[3], P. Scollard[3], E. H. Trittschuh[4,5], J. B. Mez[6], W. S. Bush[7], M. L. Cuccaro[8], The Alzheimer's Disease Sequencing Project (ADSP), C. Cruchaga[9,10], L. A. Farrer[11,12], L-S. Wang[13], G. D. Schellenberg[13], R. P. Mayeux[14,15,16], W. A. Kukull[17], C. D. Keene[17], A. J. Saykin[18], S. C. Johnson[19], C. D. Engelman[20], D. A. Bennett[21], L. L. Barnes[21], A. M. Goate[22], A. Renton[22], E. Marcora[22], B. Fulton-Howard[22], T. Patel[22], S. L. Risacher[18], A. Destefano[11], J. A. Schneider[21], S. Seshadri[23], A. W. Toga[24], D. Tosun[25], J. Vance[26], E. C. Mormino[27], T. J. Montine[27], S. Biber[17], G. W. Beecham[26], P. L. De Jager[28], A. J. Lee[28], B. Vardarajan[14], J. J. Manly[14], A. M. Brickman[14], M. A. Renteria[14], Q. Lu[29], P. K. Crane[17], T. Hohman[30], L. Dumitrescu[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Genetics Inst., Nashville, TN, [3]Univ. Of Washington, Seattle, WA, [4]Univ. of Washington Sch. of Med., Seattle, WA, [5]GRECC, Seattle, WA, [6]Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [7]Case

Western Reserve Univ, Cleveland, OH, [8]John P. Hussman Inst. for Human Genomics, Miami, FL, [9]Washington Univ. Sch. of Med., Saint Louis, MO, [10]NeuroGenomics and Informatics Ctr., St. Louis, MO, [11]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [12]Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, Boston, MA, [13]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [14]Columbia Univ., New York, NY, [15]The Taub Inst. for Res. on Alzheimer's Disease and The Aging Brain, New York, NY, [16]Columbia Univ. Med. Ctr. and The New York Presbyterian Hosp., New York, NY, [17]Univ. of Washington, Seattle, WA, [18]Indiana Univ. Sch. of Med., Indianapolis, IN, [19]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [20]Univ. of Wisconsin Madison Sch. of Med. and Publ. Hlth., Madison, WI, [21]Rush Univ. Med. Ctr., Chicago, IL, [22]Icahn Sch. of Med. at Mount Sinai, New York, NY, [23]Univ. of Texas Hlth.Sci. Ctr. at San Antonio, San Antonio, TX, [24]Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [25]Univ. of California, San Francisco, Stanford, CA, [26]Univ. of Miami Miller Sch. of Med., Miami, FL, [27]Stanford Univ. Sch. of Med., Stanford, CA, [28]Columbia Univ. Med. Ctr., New York, NY, [29]Univ. of Wisconsin-Madison, Madison, WI, [30]Vanderbilt Univ. Med Ctr., Nashville, TN

Abstract:

Background: Despite evidence that Alzheimer's Disease (AD) is a highly heritable disease, there remains substantial "missing" heritability, likely due to the clinical and neuropathologic heterogeneity inherent in the disease. Here, we leverage sensitive longitudinal cognitive measures of (i.e., endphenotypes of AD) in a rare variant analysis to identify novel genetic drivers of cognitive decline in aging and AD. **Methods:** We leveraged 8 cohorts of cognitive aging with whole genome sequencing data from the AD Sequencing Project to conduct rare variant analyses of multiple domains of cognition (N=9,317; mean age=73; 56% female; 52% cognitively unimpaired). Harmonized scores for memory, executive function, and language were derived using confirmatory factor analysis models. Participants of European ancestry inferred using SNPweights and 1000G reference panel were included in analysis. Variants included in the analysis had a minor allele frequency < 0.01, a minor allele count of ≥ 10, and were annotated as a high or moderate impact SNP using VEP. Associations baseline scores in each cognitive domain were performed using SKAT-O, including 92,905 rare variants among 16,243 genes. All tests were adjusted for sex, baseline age, sequencing center and platform, and the first 5 principal components of genetic ancestry. Correction for multiple comparisons was completed using the false discovery rate (FDR) procedure. **Results:** *APOE* was negatively associated with baseline memory, language, and executive function, though only memory survived correction (p_{FDR}=0.001). Outside of *APOE*, *ITPKB* was negatively associated with baseline executive function (p_{FDR}=0.048). *SHCBP1L*, and *CCNF* showed nominal negative associations and *AKTIP* showed nominal positive associations with multiple domains of cognition that did not survive correction for multiple comparisons (all p_{FDR}<0.07). **Conclusions:** These results highlight novel rare variants associated with cognition. *ITPKB* is an AGORA-nominated gene target for potential AD

treatment. It is important in the regulation of immune cells and displays increased expression in the cortex of AD patients compared to controls. *CCNF* and *AKTIP* are brain eQTLs and have differential RNA expression in AD brains. Previously, variants in *AKTIP* have been associated with educational attainment, intelligence, and memory, while variants in *CCNF* have been associated with neuritic plaques and neurofibrillary tangles. Future work will incorporate longitudinal data into our analyses.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5134F: Objective Evaluation of Clustering Methods for Resolving the Heterogeneity of PCOS

Authors:

K. Brewer[1], R. Sisk[2], M. Dapas[2], C. Li[3], A. Dunaif[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Northwestern Univ., Chicago, IL, [3]Univ. of Southern California, Los Angeles, CA

Abstract:

Clustering methods have been used to resolve heterogeneity of complex diseases to elucidate underlying biologic mechanisms. We (Dapas et al, 2020) and others have applied these methods to PCOS and identified discrete clusters associated with distinct reproductive or metabolic pathways. We compared two widely used methods, hierarchical clustering (HC), which recursively merges subjects based on their similarity, and k-means (Km), which iteratively groups individuals to minimize their distance to cluster centroids. European ancestry PCOS cases, n=874, diagnosed by NIH criteria were studied. Eight traits (BMI, testosterone, SHBG, DHEAS, LH, FSH, fasting insulin, fasting glucose) and 9 traits (adding anti-Müllerian hormone [AMH]) were used for clustering, with cluster number (k) 2, 3, or 4. Genomewide association study (GWAS) meta-analysis was performed in a discovery (620 cases, 2951 controls) and a replication cohort (371 cases, 926 controls); genotypes were imputed to the TOPMED (r2) panel. Clustering using 8 traits and k=3 resulted in the best cluster stability for both HC and Km (pairwise consensus proportion: HC 0.98, Km 0.93) compared to k=2 with 8 traits (HC 0.88, Km 0.88) or 9 traits (HC 0.95, Km 0.92). No stability was observed for k=4 with 8 or 9 traits for either method. Km showed slightly better cluster separation than HC (average silhouette width: Km 0.12 vs HC 0.09) with k=3 and 8 traits. Clusters were designated as reproductive (higher LH, FSH, SHBG), metabolic (higher BMI, insulin, glucose), and background subtypes. GWAS was performed as an orthogonal (i.e., uncorrelated) approach to assess whether the clusters were associated with distinct genetic loci suggesting that they captured biologically relevant subtypes. Clusters identified by HC had genomewide significant associations (metabolic

subtype, c9orf3/FANCC, rs10761370, minor allele frequency [MAF] 0.47, p=1.21 x 10-8; background subtype, FSHB/ARL14EP, rs10835649, MAF 0.17, p=8.49 x 10-10). There were no genomewide significant signals in the clusters identified by Km, despite better cluster separation. In summary, HC clusters appeared to identify etiologically distinct subtypes since two of the three clusters were associated with genomewide significant loci. Neither the addition of AMH nor increasing the number of groups improved cluster metrics. Our results emphasize the importance of independent validation of clustering approaches using an orthogonal confirmation strategy such as GWAS. We conclude that HC clustering using 8 traits is superior to Km for resolving the genetic heterogeneity of PCOS.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5135F: OPRD1 encoding delta opioid receptor: a new culprit in type 2 diabetes and insulin secretion

Authors:

A. Bonnefond[1], S. Meulebrouck[1], J. Merrheim[1], B. Balkau[2], G. Charpentier[3], S. Franc[3], M. Marre[4], M. Derhourhi[5], P. Froguel[6]; [1]Inserm, LILLE, France, [2]Inserm, Villejuif, France, [3]CERITD, Paris, France, [4]Inserm, Paris, France, [5]Inserm, Lille, France, [6]CNRS UMR8199-1283, Lille, France

Abstract:

Functional genetics has been proven to be instrumental to identify drug targets for metabolic disorders. Opioid use impacts metabolic homeostasis, although mechanisms remain elusive. Here, we used functional genetics focused on the OPRD1 gene (encoding the delta opioid receptor, DOP) to decipher its impact on type 2 diabetes.
OPRD1 was initially sequenced in 6,971 participants from the French RaDiO study and functional investigation of each variant was done in vitro. Association analyses between OPRD1 variants and metabolic traits were performed in RaDiO and confirmed in up to 222,054 participants (UK Biobank, T2D Knowledge Portal). We found that loss-of-function variants were associated with higher adiposity and lower hyperglycemia risk, while gain-of-function variants were associated with lower adiposity and higher type 2 diabetes risk. These last results on gain-of-function variants were in line with observational studies of opium addicts. Furthermore, based on RNA-seq analyses, we found that OPRD1 was expressed in human pancreatic islets and beta cells (but not in rodents) in addition to various brain regions. Of note, the other opioid receptor genes (OPRM1 and OPRK1) were not expressed in islets and beta cells. We also showed that DOP inhibition by a specific antagonist (naltrindole) strongly enhanced insulin secretion from human pancreatic beta cells (EndoCbetaH5) and islets (InSphero).

Finally, RNA-seq in human beta cells identified several pathways regulated by DOP antagonism, including nerve growth factor, circadian clock and nuclear receptor pathways. In conclusion, our study highlights DOP as a major link between opioids and metabolic homeostasis in humans, playing a direct role in the physiology of insulin secretion, with a potential impact for novel class of treatment against type 2 diabetes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5136F: Paternal age, *de novo* mutation, and age at onset among co-affected schizophrenia sib-pairs

Authors:

S-H. Wang[1], Y-C. Feng[2], W. Chen[1], M-C. Lin[1], J. Hsu[3], C-F. Cheng[1], C-M. Liu[4], H-G. Hwu[4], Y-T. Huang[5], T-P. Lu[2]; [1]Natl. Hlth.Res. Inst.s, Miaoli County, Taiwan, [2]Natl. Taiwan Univ., Taipei, Taiwan, [3]Natl. Taiwan Univ., Taipei City, Taiwan, [4]Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [5]Academia Sinica, Taipei, Taiwan

Abstract:

Whether delaying fatherhood leads to more mutations, thereby increasing psychiatric risk in offspring, remains under debate. No study has directly examined the role of de novo mutations (DNMs) in the paternal age-offspring disease risk/onset age relationships. This study aimed to explore the association between paternal age, the number of DNMs, and age at onset of schizophrenia by sequencing the whole genome of multiplex schizophrenia families. Whole-genome sequencing (30x) was performed in 5 Taiwanese families, each comprising 3 co-affected siblings and healthy parents. Causal mediation analyses were used to explore the mediating role of DNMs in the paternal age effect. Paternal age predicted increased DNMs (+1.50 DNMs/year, 95% CI: 0.81, 2.19, $p<0.0001$) over maternal age (+0.09 DNMs/year, 95% CI: -1.01, 1.19, $p=0.87$). The effect of paternal age on the number of DNMs varied across families. Each additional DNM resulted in a 0.16-year earlier onset age of schizophrenia (95% CI: 0.04, 0.27, $p=0.009$). The mediation analyses showed that 30% of the observed association of paternal age with onset age of schizophrenia might be mediated through paternal age-related DNMs. Our study, the first to directly quantify the mediating effect of DNMs, provides support for a causal role of paternal age-related mutations in the increased psychiatric risk in offspring.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5137F: Pediatric Case of Hereditary Polyneuropathy: Charcot-Marie-Tooth Disease Type 2Q and CMT1A gene mutation.

Authors:

A. Poudyal[1], R. Mishra[2]; [1]Civil Service Hopstital, Kathmandu, Nepal, [2]Civil Service Hosp., Kathmandu, Nepal

Abstract:

Background: This case report presents a rare case of a pediatric patient with hereditary polyneuropathy, specifically Charcot-Marie-Tooth (CMT) disease type 2Q and CMT1A gene mutation. The report emphasizes the importance of whole exome sequencing (WES) testing in confirming the diagnosis of CMT disease with pathogenic variant and guiding treatment decisions, especially in complex cases with atypical presentations.
Case Summary: A 10-year-old male born of a non-consanguineous marriage presented with clinical indications of developmental delay, seizures, atonia, shrinking of eyes, twitching of lips, left-sided hemiparesis, and recurrent pneumonia. Diagnostic evaluation revealed a pathogenic variant in the DHTKD1 gene associated with CMT disease type 2Q, as well as a likely pathogenic duplication on chromosome 17 overlapping genes associated with CMT1A. The patient was managed with physical and occupational therapy, with no significant improvement in motor function noted on follow-up.
Conclusion: This case emphasizes the importance of genetic testing in confirming the diagnosis of complex hereditary polyneuropathies and guiding treatment decisions, highlighting the need for a comprehensive and personalized approach to patient care in such cases.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5138F: Peroxisome Biogenesis Disorder in a Patient with heyerozygous PEX6 gene mutation and 3’UTR TAAA deletion representing allelic expression imbalance

Authors:

A. Akalin[1], I. Akalin[2], E. Kaya[3], N. Elcioglu[4]; [1]Bahcesehir Univ. Sch. of Med., Istanbul, Turkey, [2]Metagentech Ctr. for Genetic Diseases, Istanbul, Turkey, [3]Riga Univ. Sch. of Med., Riga, Latvia, [4]Marmara Univ. Pediatric Genetics, Istanbul, Turkey

Abstract:

Peroxisome Biogenesis Disorder 4B (PBD4B) is genetically heterozygous disorder which is inherted autosomal recessively or dominantly. The recessive forms are caused by homozygous mutation in *PEX6* gene. Peroxisome assembly is diminished leading to metabolic activity impairment that encourages progressive disabilities and complex neurodevelopmental sequelae.An 12-year-old female patient was born 1260g, at 27 weeks of pregnancy. After birth, her heart stopped twice and was kept in neonatal intensive care unit for two months and considered as cerebral palsy. Macrocephaly, frontal bossing, hypertelorism, severe congenital hypotonia, seizures, developmental delay, triangular mouth, wide frontal fontanel and wide big toe were detected on physical examination. Whole exome sequencing (WES) analysis was run for mutation analysis.WES revealed likely pathogenic c.2254C>G: p.L752V heterozygous mutation, variant of unknown significance of c.2814_2816delinsACA homozygous mutation and heterozygous 3'UTR c.*445_*442delTAAA deletion in *PEX6* gene identified. We report another PBD4B prient having heterozygous *PEX6* mutation suggesting allelic expression imbalance.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5139F: Phenotypic consequences of variation in elastic fiber genes

Authors:

A. Justice[1], N. Josyula[2], T. Luperchio[3], M. A. Kelly[1], B. A. Kozel[4], M. S. Williams[1]; [1]Geisinger, Danville, PA, [2]Geisinger Hlth.System, Danville, PA, [3]NIH - (C) Guidehouse, Ellicott City, MD, [4]Nationwide Childrens' Hosp., Columbus, OH

Abstract:

Elastic fibers provide recoil to tissues and are found throughout the body in the skin, lungs, arteries, veins, connective tissue, etc. Thus, variation in these genes may lead to a myriad of health consequences. Changes in *ELN, LOX, FBLN5, EFEMP2*, or *MFAP4* have been linked to defects in aortic diameter and tortuosity, emphysema, or skin laxity. We hypothesize that variation in elastic fiber genes may contribute more widely to connective tissue diseases than is now appreciated. We took a gene-first approach to identify disease phenotypes associated with elastic fiber gene variation using exome sequencing data in Geisinger's MyCode Study (N = 173,585). Using Ensembl's VEP, we extracted rare variants (MAF<0.01) with a REVEL score > 0.183, predicted high or moderate impact, and a PolyPhen score > 0.5. We applied two variant masks based on the consequence in the canonical transcript: Strict) missense variants only; Broad) missense, frameshift, stop/gain, splice sites and inframe indels. We performed gene-based Phenome-wide association analyses (PheWAS) applying both SKAT and SKAT-O methods.

We conducted a Kolmogorov-Smirnov equality-of-distribution test to identify significant enrichment in phenotypic categories after correcting for number of masks, methods and categories ($P < 7.35 \times 10^{-4}$). Of the 1,542 PheCodes analyzed, we identified one PheWAS significant association ($P < 3.14 \times 10^{-5}$, SKAT-O, Broad) for *ELN* with PheCode 727.7, "Contracture of tendon (sheath)". For the remainder, the results did not reach Bonferroni significance but have the following top associations: *EFEMP2* with 401 "Hypertension" ($P=4.65 \times 10^{-5}$, SKAT-O, Strict); *MFAP4* with 370.2 "Superficial keratitis" ($P=8.60 \times 10^{-5}$, SKAT, Broad); *FBLN5* with 264.3 "Delayed milestones" ($P=2.33 \times 10^{-4}$, SKAT-O, Strict); and *LOX* with 758.1 "Chromosomal anomalies" ($P=7.61 \times 10^{-4}$, SKAT-O, Strict). We identified significant enrichment of associations for hematopoietic traits (SKAT, Broad) and endocrine/metabolic traits (SKAT, Strict) with *MFAP4*; as well as circulatory system disorders with *ELN* (SKAT-O, Strict). Also, when considering all five genes, we observed significant enrichment for circulatory system (Strict, both methods), respiratory (SKAT, both masks), and mental disorder (SKAT, both masks) traits. Overall, we identified one robust association between a new musculoskeletal trait and potentially damaging variation in *ELN*. Further, we detected enrichment for nominal associations within circulatory system and respiratory categories known to relate to elastic fibers, but also mental disorders, which may indicate more broad and downstream impacts of damage to elastic fibers.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5141F: Polygenic Risk for Substance Abuse in the ABCD Study

Authors:

R. Loughnan[1], J. Ahern[2], W. Thompson[3], N. Wade[1], C. Fan[3]; [1]Univ. of California, San Diego, La Jolla, CA, [2]UC San Diego, La Jolla, CA, [3]Laureate Inst. for Brain Res., Tulsa, OK

Abstract:

Studying substance abuse and its genetic contributions is crucial for identifying at-risk individuals and developing targeted prevention strategies. Investigating these factors during adolescence is vital, as this period is critical for brain development and represents an age of experimentation and initiation of substances. Here we generate polygenic risk scores (PRS), using pgscatalog, across a range of substance abuse and use related traits in over 11,000 individuals aged 9-11 years old, at baseline, from the Adolescent Brain and Cognitive Development (ABCD) Study®. Due to the large contribution of European-like individuals in the discovery samples, we performed ancestry stratified analysis in EUR-like (n=6,695), AFR-like (n=811) and MIX-like (n=3,664) groups reflecting genetic similarity to continental ancestry groups. We evaluate the performance of PRS to predict i) measures of impulsivity taken from

the UPPS questionnaire and ii) self reported use of tobacco, marijuana, alcohol and caffeine. In the EUR-like group we find PRS related to tobacco and alcohol use were associated with greater impulsivity across all subscales of the UPPS: Smoking Status PRS with Lack of Perseverance ($t(6,695)=11.28$, $p=2.03x10^{-29}$) and Alcohol Consumption with Sensation Seeking ($t(6,695=5.37$, $p=7.73x10^{-8}$). Longitudinal analysis across ages 9-15 years old revealed PRS vs impulsivity associations that: a) grew over as the sample aged (e.g. Age at Smoking PRS with Lack of Perseverance: 9-11 years: $r^2=0.004$, 11-13 years: $r^2=0.007$, 13-15 years: $r^2=0.014$) and b) others that diminished as the sample aged (e.g. Alcohol Consumption PRS with Sensation Seeking: 9-11 years: $r^2=0.005$, 11-13 years: $r^2=0.004$, 13-15 years: $r^2=0.001$). Evaluating the performance PRS against self reported substance use and exposure we found that PRS of smoking traits were associated with weekly consumption of caffeine at 11-13 years old. At 13-15 years of age we found PRS of smoking traits was associated with marijuana and tobacco exposure, in addition to weekly caffeine consumption. Replication of our EUR-like results in AFR-like and MIX-like sub-samples revealed an attenuation of effects underscoring the importance of collecting genetic studies in more ancestrally diverse cohorts. Our results highlight the dynamic relationship between genetic risk factors of substance abuse and different factors of impulsivity through adolescence. Significant associations between regular caffeine consumption and PRS, at a relatively illicit drug-naive age, may represent the detection of early addictive behaviors related to substance abuse risk in later life.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5142F: Polygenic Risk Scores can predict gestational diabetes and weight gain among pregnant women in Taiwan

Authors:

Y-W. Chang[1], H-F. Lu[2], H-I. Hsieh[3], Y-C. Huang[1], W-L. Liao[1], F-J. Tsai[4]; [1]China Med. Univ., Taichung City, Taiwan, [2]China Med. Univ. Hosp., Taichung City, Taiwan, [3]Natl. Taiwan Univ., Taipei City, Taiwan, [4]China Med. Univ. Children's Hosp., Taichung City, Taiwan

Abstract:

Background: Gestational diabetes mellitus (GDM) is a prevalent pregnancy complication. Abnormal glucose metabolism in mothers may increase diabetes susceptibility due to placental hormones and weight changes. Effective risk-reduction strategies are essential for high-risk individuals. This study investigated the impact of BMI-related genetic variants on weight gain and GDM risk during pregnancy among Taiwanese women. **Method:** We analyzed electronic medical records of 3,875 pregnant women from a medical center in Taiwan, including 3,162

diagnosed with GDM based on oral glucose tolerance tests and 2,483 with complete body weight data. The polygenic risk score for BMI (PRS_BMI) was constructed using genetic variants from a genome-wide association study. **Results:** PRS_BMI significantly influenced weight gain during pregnancy and GDM risk. The odds ratio (OR) for GDM per 1 SD increase in PRS_BMI was 1.34 (1.24-1.45, $P < 0.001$) after adjusting for maternal age. Compared to the first quartile (Q1) of PRS_BMI, the risk of GDM increased in the second (Q2), third (Q3), and fourth quartiles (Q4), with ORs of 1.22 (0.96-1.54), 1.31 (1.03-1.65), and 1.93 (1.54-2.43), respectively (P for trend < 0.001). PRS_BMI was positively associated with baseline weight and weight gain during pregnancy. Each unit increase in PRS_BMI increased baseline weight by 6.386 kg/m² ($R^2 = 0.391$) and weight gain by 0.428 kg/m² ($R^2 = 0.010$), with P values < 0.001. Stratified analyses showed significant correlations between PRS_BMI and weight gain in women with BMI <25 kg/m² ($R^2 = 0.054$, $P < 0.001$) and BMI ≥25 kg/m² ($R^2 = 0.016$, $P = 0.010$) at the start of pregnancy. **Discussion:** These findings indicate that women with a higher genetic predisposition for higher BMI are more likely to gain weight during pregnancy. PRS may help monitor pre-pregnant women and offer lifestyle interventions for those at high risk of GDM.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5143F: Polygenic risk scores for incident dementia in the Multi-Ethnic Study of Atherosclerosis

Authors:

D. Xue[1], E. Blue[1], T. Sofer[2], T. Hughes[3], J. Rotter[4], A. Fohner[1]; [1]Univ. of Washington, Seattle, WA, [2]Beth Israel Deaconess Med. Ctr. / Harvard Med. Sch., Boston, MA, [3]Wake Forest Univ. Sch. of Med., Winston-Salem, NC, [4]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA

Abstract:

Background: Over 75 Alzheimer's disease (AD) and dementia-associated variants have been identified. Most were identified by studies in populations of European descent, and few have replicated in studies among groups differing in genetic ancestry. The utility of polygenic risk scores (PRS) for predicting AD and dementia in diverse and admixed populations is still unclear. We compared how PRS approaches that differ in p-value thresholds, variant weights, and source ancestry perform in predicting late-onset dementia in a multi-ethnic cohort.
Methods: We used Cox models and C-statistics to compare clumping and thresholding (C+T) methods against Bayesian approaches (PRS-CS, PRS-CSx) in the Multi-Ethnic Study of Atherosclerosis (MESA), a longitudinal cohort of 6,814 men and women aged 45-84 at baseline who identified as Black/African American, Chinese, Hispanic, or White. For the C+T and PRS-CS

methods, we used summary statistics from the International Genomics of Alzheimer's Project GWAS of non-Hispanic White individuals, excluding the *APOE* region. PRS-CSx was designed to improve cross-population performance and allowed inclusion of summary statistics from the African Genome Resources Panel GWAS. For C+T, we varied p-value thresholds: 0.01 (nominal), 1e-05 (suggestive), and 5e-08 (conservative). All PRS were regressed over the first 3 principal components and standardized. We compared the ability of each method to predict incident dementia for MESA participants overall and stratified by self-reported race/ethnicity. For admixed groups, we additionally analyzed performance based on the estimated proportion of non-Finnish European (NFE) ancestry.
Results: The median follow-up time was 16.8 years, and 569 participants developed dementia according to hospital and death certificate ICD codes. Only the C+T PRS with conservative (15 SNPs, HR = 1.21 [95% CI: 1.11-1.31]) or suggestive (53 SNPs, HR=1.12 [95% CI: 1.03, 1.22]) p-value thresholds were significantly associated with dementia. The C+T method with p<5e-08 had the strongest predictive performance overall (C= 0.55, SE=0.01) and among the Black (C=0.57, SE = 0.02) and White (C= 0.56, SE=0.02) participants. PRS performance did not vary significantly with NFE ancestry among the admixed participants.
Conclusion: The PRS based on C+T method with only 15 SNPs with strong evidence of association with AD is more strongly associated with dementia than PRS derived from Bayesian models that include >20,000 SNPs, even in target populations genetically dissimilar from that of the source data. When it comes to PRS for dementia, it pays to be picky: including more variants does not improve PRS performance.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5144F: Polygenic risk scores of rheumatoid arthritis associated with seropositivity & bone erosions in a Taiwanese population

Authors:

Y-M. Chen[1], T-S. Wu[2]; [1]Taichung Veterans Gen. Hosp., Taiwan, Taichung, Taiwan, [2]Chung Shan Med. Univ., Taiwan, Taichung, Taiwan

Abstract:

Background: Polygenic risk scores (PRS) are widely used to estimate disease risks and predict clinical outcomes. However, ethnical differences exist across the globe. We aimed to use PRS for predicting the development of rheumatoid arthritis (RA) in a Taiwanese population and investigate whether PRS of RA may be associated with structural damage. Methods: We calculated PRS using GWAS data from a hospital-based cohort of 2,044 RA cases and 7,950 non-

RA controls participating in the Taiwan Precision Medicine Initiative. LDpred2, PLINK and PRSice-2 models were used to evaluate the area under curve (AUC) of RA susceptibility. Demographic data, seropositivity of rheumatoid factor (RF), anti-citrullinated protein antibody (ACPA), treatment with biologic and targeted synthetic disease modifying anti-rheumatic drugs (bDMARDs and tsDMARDs), and bone erosions identified by either plain X-ray or ultrasonography were compared among quartiles of PRS. Results: PRS of 97,396 single nucleotide polymorphisms (SNPs) derived from LDpred2 exhibited the highest AUC for predicting the development of RA as compared to the remaining algorithms. Participants in the top RA-PRS quartile have the highest proportion of RF and ACPA positivity (74.8% & 65.0%, respectively), bone erosion (86.4%), and a higher chance to receive bDMARDs or tsDMARDs (42.3%) as compared to their counterparts. In addition, RA-PRS was a significant risk for bone erosion in the multivariate regression model with the adjustment of RF, ACPA positivity and therapeutic medication, specifically in patients aged < 60 years. Conclusions: PRS is associated with seropositivity, erosive bone disease, and need for advanced therapy among Taiwanese patients with RA.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5146F: Prediction of clinical diabetes subclasses using polygenic risk scores considering enrichment of diabetes pathways

Authors:

A. Miyake[1], H. Tanabe[2], A. Narita[3], M. Shimabukuro[2], G. Tamiya[1,3]; [1]Tohoku Univ., Sendai, Japan, [2]Fukushima Med. Univ., Fukushima, Japan, [3]Tohoku Med. Megabank Organization, Sendai, Japan

Abstract:

Background: Diabetes exhibits significant heterogeneity in clinical features, treatment responses, and complication risks. However, current treatments often fail to address the diverse pathophysiological aspects adequately, relying solely on indicators such as blood glucose levels and HbA1c. Ahlqvist et al. (2018) identified five subtypes of diabetes: SAID, SIDD, SIRD, MOD, and MARD, each associated with distinct complication risks, emphasizing the need for tailored treatments. Despite the clinically significant potential of subtype-specific polygenic scores (PRS) for risk prediction, their predictive accuracy may be limited by smaller sample sizes due to subtype-specific genome-wide association studies (GWAS).
Objective: To overcome these challenges, this study aims to develop predictive models tailored to the clinical subtypes, accounting for the pathophysiological differences, and enhancing the

accuracy of polygenic risk scores (PRS) compared to those based on subtype-specific GWAS.
Method: Utilizing data from the Tohoku Medical Megabank Project (TMM) and Fukushima Medical University in Japan, we classified 3,950 diabetes cases into subtypes and partitioned them into training and validation datasets alongside 26,470 controls. Next, we computed partitioned polygenic scores (pPS), which are derived by deconstructing the overall polygenic score into components representing different etiological pathways contributing to diabetes. These pPS help capture the elements of genetic influences that contribute to clinical heterogeneity. Specifically, in this study, we used pPS to estimate individual-level pathway enrichment for each of the 12 diabetes-related pathways identified by Smith et al. (2024) through Bayesian non-negative matrix factorization. These scores are then utilized as predictive factors for each subtype. Penalized predictive models were subsequently developed, taking into consideration the uncertainty in subtype boundaries.
Result: Our predictive models demonstrated consistent performance across all clusters in the validation dataset, particularly excelling in the SIRD subtype (AUC=0.72). These models outperformed PRS based on the publicly available subtype-specific GWAS summary statistics.
Conclusion: Despite the challenges posed by small sample sizes, particularly in subtypes like SIRD, our pathophysiology-based risk prediction models exhibited high accuracy, underscoring their clinical utility.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5148F: Prevalence and Impact of CYP1A1 and GST Gene Polymorphisms on Aplastic Anemia in Northern India

Authors:

S. Shukla, A. Dwivedi; King George's Med. Univ., Lucknow, India

Abstract:

Title: Prevalence and Impact of CYP1A1 and GST Gene Polymorphisms on Aplastic Anemia in Northern IndiaAuthor Block: S. Shukla, A. Dwivedi, Tripathi A.K.; King George's Med. Univ., Lucknow, India**Introduction**: The cause of acquired aplastic anemia (AA) in most individuals is still unknown. Patients with impaired detoxification capacity for environmental contaminants are likely to have a higher susceptibility to developing AA. Cytochrome P450 (*CYP450*) and glutathione S transferase (*GST*) are the primary enzymes responsible for metabolizing foreign substances in phases I and II of the body's detoxification process.**Material and methods:** We investigated the effect of polymorphisms in the *CYP4501A1*, *GSTM1*, and *GSTT1* genes on susceptibility and disease severity in 200 patients with AA and compared it with the frequency

to 200 healthy age-matched controls using RFLP.**Results:** The frequency of the *CYP1A1m4* allele was considerably higher in AA patients compared to normal controls, with an odds ratio of 2.10 (95% confidence interval 1.39-3.17; p = 0.01). There were no significant differences observed in the *CYP1A1* genotypes or *GST* genotypes between AA patients and controls.**Conclusion**: The patients may have a higher risk of developing AA because of a genetic variation that affects the body to metabolize benzo(a)pyrene. In India, this is an important *CYP1A1m4* allele that can be used to assess AA risk. Keywords: Aplastic anemia, Cytochrome P450, Glutathione S transferase, Polymorphism

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5150F: Quantifying extent of overlap in gene expression regulation between common and rare variants in psychiatric traits

Authors:

H. Abe[1], C. Liao[2], L. Han[3], T. Morley[3], D. Ruderfer[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Introduction: Genetic studies have identified thousands of gene associations to complex diseases from both common and rare variations. While prior work has demonstrated overlap between the genes identified from rare and common variation, the extent to which these genes share the same risk pathways remains unclear. Here, we use gene-gene coexpression data to infer shared downstream regulatory effects of disease genes identified from either rare or common genetic studies and assess the biological overlap to gain a deeper picture of disease biology. **Methods:** For rare variation, we collected large-scale exome gene burden summary data for six brain diseases (Autism, Schizophrenia, Alzheimer's, Epilepsy, Parkinson's, and Bipolar-disorder) and identified exome-wide significant genes. For common variation, we mapped the nearest gene for each index genome-wide significant SNP using the latest GWAS for the same traits. For both variation types, we meta-analyzed coexpression data from post-mortem brain tissue data from Common Mind Consortium (n=733), to derive a "Convergence Z-score". This score correlates with differential expression results from CRISPR perturbations proxying downstream regulated genes. The inferred downstream genes were used to quantify overlap between common and rare variants and identify enriched pathways. **Results:** Supporting previous findings, there was partial overlap of rare and common implicated genes at genome wide level for Schizophrenia and Alzheimer's but no significant overlap for Autism, Epilepsy, Bipolar, and Parkinson's. Among genes in the top decile of

convergence Z-score, we observed significant overlap ($p < 2e-16$) between common and rare implicated genes for all diseases (Autism OR= 0.17, Epilepsy OR = 14, Alzheimer's OR = 3.6, Parkinson's OR = 0.5, Bipolar OR = 9, Schizophrenia OR = 19). The shared genes were enriched in disease relevant pathways, such as synaptic vesicle cycle and exocytosis for Schizophrenia. Top convergent genes showed enrichment in both common and rare disease genes for other phenotypes and excess overlap in haploinsufficient and FDA approved Drug targets. Moreover, variants near top convergent genes are enriched in heritability. **Discussion:** Our study establishes a pipeline that uses co-expression, GWAS, and exome data to expand understanding of shared transcriptional regulation between rare and common variations. When incorporating downstream regulated genes, we find greater significance of overlap between common and rare variants than comparing at genome wide significant level, highlighting downstream mechanisms that can be further investigated for their contribution to disease risk.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5151F: Rare and common variants associated with alcohol consumption identify a conserved molecular network

Authors:

B. Leger[1], J. J. Meredith[1], T. Ideker[2], S. Sanchez-Roige[1,3], A. Palmer[1]; [1]Univ. of California San Diego, La Jolla, CA, [2]Univ California, San Diego, San Diego, CA, [3]Vanderbilt Univ., Nashville, TN

Abstract:

Genome-wide association studies (GWAS) have identified hundreds of common variants associated with alcohol consumption. In contrast, genetic studies of alcohol consumption that use rare variants are still in their early stages. No prior studies of alcohol consumption have examined whether common and rare variants implicate the same genes and molecular networks, leaving open the possibility that the two approaches might identify distinct biology. To address this knowledge gap, we used publicly available alcohol consumption GWAS summary statistics (GSCAN, N=666,978) and whole exome sequencing data (Genebass, N=393,099) to identify a set of common and rare variants for alcohol consumption. Gene-based analysis of each dataset have implicated 294 (common variants) and 35 (rare variants) genes, including ethanol metabolizing genes *ADH1B* and *ADH1C,* which were identified by both analyses, and *ANKRD12*, *GIGYF1*, *KIF21B*, and *STK31*, which were identified only by the rare variant analysis, but have been associated with other neuropsychiatric traits. We then used a network colocalization procedure to propagate the common and rare gene sets onto a shared molecular network, revealing significant overlap. The shared network identified gene families that function

in alcohol metabolism, including *ADH*, *ALDH*, *CYP*, and *UGT*. 71 of the genes in the shared network were previously implicated in neuropsychiatric or substance use disorders but had not been previously identified for alcohol-related behaviors (e.g. *EXOC2*, *EPM2A*, and *CACNG4)*. Differential gene expression analysis showed enrichment in the liver and several brain regions. Thus, genes implicated by common and rare variants identify shared functions relevant to alcohol consumption, which also underlie psychiatric traits and substance use disorders that are comorbid with alcohol use, providing a more holistic understanding of two disparate sources of genetic information.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5152F: Rare variant analysis for post- traumatic stress disorder using whole-exome-sequencing data.

Authors:

T-H. Nguyen[1], J. R. I. Coleman[2], A. Gentry[1], B. Webb[3], R. Peterson[4], K. Kendler[1], B. Riley[1], A. Amstadter[1], C. M. Sheerin[1]; [1]Dept. of Psychiatry, Virginia Inst. for Psychiatric and Behavioral Genetics, Virginia Commonwealth Univ., Richmond, VA, [2]Social, Genetic and Dev.al Psychiatry Ctr., Inst. of Psychiatry, Psychology, and NeuroSci., King's College London, United Kingdom, [3]RTI Intl., Research Triangle Park, NC, [4]SUNY Downstate Hlth.Sci. Univ.State Univ. of New York Downstate Hlth.Sci. Univ, Brooklyn, NY

Abstract:

Posttraumatic stress disorder (PTSD) is a common, clinically and economically impactful psychiatric disorder. Genetic discovery for PTSD has accelerated at a rapid pace with numerous common variants identified from genome-wide association studies (GWAS). However, a large-scale study of rare variants associated with PTSD using next-generation sequencing data has not been conducted. Here, we analyzed the latest whole-exome-sequencing (WES) dataset from the UK Biobank (UKB) to uncover additional genetic signals for PTSD. We conducted WES analyses of PTSD threshold status for 133,738 samples of European ancestry using mental health questionnaire data. First, we used established quality-control steps and pipelines from UKB. Next, synonymous, missense and loss-of-function variants were identified for analyses. We then ran SAIGE-GENE+ to generate variant and gene-level p-values for variants with a minor allele frequency (MAF) < 0.01. The GAUSS package was then used to conduct gene-set enrichment analysis for gene-level p-values. We used gene-sets from a recent GWAS of PTSD and curated gene-sets from studies of neuropsychiatric disorders. For variant-level analysis, there were no genome-wide significant variants detected. The top associated variant had a p-value = 4.98x10^-

7 (1:152356894:A:G in gene *FLG2*, MAF < 0.01). For gene-level analysis, from 18,190 genes tested, one gene (*MAST4*) had a nominal p-value = $8.2x10^{-6}$ for loss-of-function and missense variants. For gene-set enrichment analysis, 10 out of 189 tested gene-sets were nominally significant. However, these gene-sets were not significant after multiple testing correction. Eight of these gene-sets were from mouse mutants with central nervous system phenotypes with the smallest p-value observed for genes relating to abnormal innervation (p-value = 0.003). Two of these ten were significant gene-sets in studies of other neuropsychiatric disorders. Interestingly, the gene-set from the latest PTSD GWAS had a modest p-value of 0.089. In summary, for rare variant analysis, we did not see statistically significant signals from the analysis of approximately 130K samples using a proxy phenotype of PTSD from the UKB. Future studies which focus on understanding the overlapping information of rare and common variants can shed additional light on the genetic architecture of PTSD. Our future extensions of these analyses will include integrative analytic approaches and importantly expand to multiple ancestry samples across several biobanks in order to advance the understanding of PTSD.*The UK Biobank Resource application number 30782.*

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5153F: Rare variant analysis of developmental stuttering identifies novel genes

Authors:

A. Scartozzi[1], Y. Yu[2], H. G. Polikowsky[1], D. M. Shaw[1], L. E. Petty[3], D. G. Pruett[3], J. M. Beilby[4], K. Z. Viljoen[4], S. Kraft[5], C. D. Huff[2], J. E. Below[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Curtin Univ., Perth, Australia, [5]Wayne State Univ., Detroit, MI

Abstract:

Developmental stuttering is a common speech disorder characterized by repetitions, sound prolongations, and blocks of speech sounds. Stuttering is strongly enriched in families and is influenced by both familial and population-level variation. Although six candidate causal stuttering genes have been identified, this variation does not explain stuttering risk in other families or general populations.
Here, we use whole-exome sequenced data from The International Stuttering Project to map gene action impacting stuttering risk. The current study uses the Variant Annotation Analysis and Search Tool ("VAAST") to identify stuttering risk genes.
Using our whole-exome sequenced data in those of European genetic ancestry ("EA", ncase=

862, ncontrol= 4856) and African genetic ancestry ("AA", ncase= 82, ncontrol= 1012), we joint called unrelated samples and controls using XPAT with default QC parameters and performed VAAST to identify enrichment of rare functional variation in our datasets. For our preliminary EA analyses, we report the top genes with enrichment of rare functional variation. Several top genes are known to be phenotypically and genetically correlated with stuttering, including insomnia (*GMPS*, $p = 3.2 \times 10^{-4}$), testosterone (*RIBC1*, $p = 5.5 \times 10^{-4}$), asthma and lung function (LRRC3C, $p = 7.00 \times 10^{-5}$; *PAMR1*, $p = 3.4 \times 10^{-4}$), and brain measurements (*MAST4*, $p = 4.2 \times 10^{-4}$, *FRMD3*, $p = 5.7 \times 10^{-4}$).
We also tested for evidence of rare variant effects from previously reported genes in the literature, including six candidate causal stuttering genes and 60 genes identified in population-based GWAS of stuttering. For our AA sample, one gene identified in familial studies reached nominal significance (*GNPTAB*, p = 0.015) and one approached significance (*NAGPA*, p = 0.087). No candidate causal genes surpassed significance in the EA sample. Next, for our EA sample, three population-based stuttering genes reached nominal significance (*SGCD*, p = 0.0097, *ADCY5*, p = 0.029, and *SLC39A8*, p = 0.045). In the AA sample, one population-based gene reached nominal significance (*SORCS1*, p = 0.038)
Future work will leverage family-based rare variant analysis tools (pVAAST) and complete pedigree information to increase power and identify novel stuttering risk genes enriched in families.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5154F: Rare variation in *LMNA* underlies polycystic ovary syndrome (PCOS) pathogenesis in two independent cohorts

Authors:

R. Bauer[1], C. Parker[1], L. Gorsic[2], M. G. Hayes[1], D. Ehrmann[3], C. Welt[4], R. Legro[5], **M. Urbanek**[1]; [1]Northwestern Univ., Chicago, IL, [2]Mayo Clinic, Naperville, IL, [3]Univ of Chicago, Chicago, IL, [4]Univ. of Utah, Salt Lake City, UT, [5]MS Hershey Med. Ctr, Hershey, PA

Abstract:

Background: PCOS is a complex, heritable endocrinopathy that is a common cause of anovulatory infertility in reproductive age women. Common non-reproductive features of PCOS include obesity, insulin resistance (IR), and type 2 diabetes. The directionality of this relationship, whether IR is primary and underlies PCOS or develops due to PCOS-driven elevated testosterone or obesity, is continuously debated. Familial partial lipodystrophy type 2 (FPLD2) is a disorder of lipid storage and IR caused by dominant missense variants in the *LMNA* gene.

Women with this Mendelian disorder of IR also experience symptoms of PCOS such as amenorrhea and hyperandrogenism. We hypothesize that genetic variation in *LMNA* also underlies PCOS, and that rare missense variation in this gene represents a monogenic form of the complex trait. **Objective:** We aim to identify and evaluate variation in *LMNA* that underlies PCOS pathogenesis in two independent cohorts. **Methods:** We sequenced the *LMNA* gene in two cohorts of European descent, totaling 1530 PCOS patients and 472 reproductively healthy controls. We comprehensively screened *LMNA* for protein-altering variants with minor allele frequency < 0.05. **Results:** In the discovery cohort, we identified 8 missense variants in 15/811 cases, and 1 variant in 1/172 reproductively healthy controls. Six variants are pathogenic and 5 are likely pathogenic by ACMG classification. While we are not powered to detect evidence for association between *LMNA* variants and PCOS with our phenotyped controls (p=0.3335, OR=3.2), there is strong evidence for association between the variants and PCOS with gnomAD non-Finnish European population controls (X^2=17, p=3.7x10^{-5}, OR=2.9). In the replication cohort, we identified 8 unique variants in 9/719 cases, and no variants in 281 reproductively healthy controls. Of the 8 variants, 2 are pathogenic, 1 is likely pathogenic, and the remaining 5 are variants of unknown significance. Again, our power to detect association with PCOS against our phenotyped controls is limited (p=0.069, OR = 7.48), but there is strong evidence for association with population controls (X^2=373.1, p<1X10^{-15}, OR= 51.84). In both the discovery and replication cohorts, variants in *LMNA* identify a specific phenotype of PCOS characterized by elevated circulating triglycerides. **Conclusion:** Rare missense variation in *LMNA* is reproducibly associated with PCOS and identifies a sub-phenotype of the complex disorder characterized by high triglycerides. The overlap between this PCOS phenotype and genetic partial lipodystrophy syndromes warrants further investigation into additional lipodystrophy genes and their potential in PCOS etiology.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5155F: Recruitment biases substantially impact GWAS results and their out-of-sample portability

Authors:

O. Smith, A. Harpak; The Univ. of Texas at Austin, Austin, TX

Abstract:

Individual data in genome-wide association studies (GWAS) is often sourced from non-representative sample populations. This limits the extent to which conclusions from GWAS of one cohort are applicable to another i.e. a lack of “portability.” Factors such as sex, age,

ancestry, and socioeconomic status often contribute to this lack of portability in modern biobank data sets. In our work, we investigate whether recruitment bias - bias in the way in which participants select or are selected to participate in biobank research - may also be impacting GWAS results. Using combined electronic health records, surveys, and genotype data from both the UK Biobank (UKB, n=500,000) and the US-based All of Us biobank (AoU, n=312,000 genotyped individuals) we test whether recruitment biases within and across these databases substantially impact GWAS associations and their downstream applications. The UKB, which used a mailer to invite all individuals in the UK to participate, is composed of individuals who are older, live in areas of higher socioeconomic status, and report fewer health conditions than the general population in the UK. AoU aims to expand sample diversity by using multiple recruitment methods and specifically reaching out to individuals who are underrepresented in genetic studies. To compare within-biobank recruitment, we performed GWAS on White-British individuals in the UKB for various diseases and physiological measurements and looked at the effect of interactions between genotype and recruitment center on these phenotypes. Bias in GWAS results may be amplified further when computing polygenic scores (PGS) which are composed of hundreds or thousands of effects across the genome. Thus we compared GWAS and PGS results using ancestry-matched participants in the UKB and All of Us databases to understand whether intra-biobank participant differences affect portability. Taken together, our analyses show that recruitment biases can substantially impact genotype-phenotype associations and their applications in research and healthcare.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5156F: Risk of Parkinson's and Alzheimer's disease co-localises with expression of *SNCA*, *LRRK2* and *BIN1* in peripheral innate immune cells in an ancestry-dependent manner.

Authors:

A. Wagen[1], R. H. Reynolds[2], S. Gandhi[1], M. Ryten[3]; [1]The Francis Crick Inst., London, United Kingdom, [2]Univ. Coll. London, London, United Kingdom, [3]The Univ. of Cambridge, London, United Kingdom

Abstract:

Immune activation is increasingly implicated in neurodegeneration. Despite this, little is known about how the genes that cause neurodegeneration are expressed in peripheral immune cells, and how this expression mediates the genetic risk of disease across diverse ancestries.
Here we use peripheral blood phenotype and expression quantitative trait loci (QTL) studies

from Asian, European and African ancestries to show that genes central to Alzheimer's disease (AD), Parkinson's disease (PD) and Frontotemporal dementia (FTD) are prominently expressed in peripheral innate immune cells, including *APP* and *PSEN2* in dendritic cells; *APOE*, *LRRK2* and *C9orf72* in monocytes; and *SNCA*, *PINK1* and *FUS* in natural killer cells. Moreover, multi-ancestry phenotype QTL studies show that this expression is not by chance, with genetic loci at many neurodegenerative genes contributing to changes in peripheral white blood cell counts.
To explore how the expression of genes in peripheral immune cells might directly mediate neurodegenerative disease, we utilised ancestry-specific genome wide association studies (GWAS) from African, Asian and European populations in a colocalisation analysis. We show that an increased risk of PD is genetically associated with increased expression of *LRRK2* in stimulated monocytes in those with European ancestry (posterior probability of hypothesis 4, PPH4 = 0.92), and decreased expression of *SNCA* in monocytes and dendritic cells from those with East Asian ancestry (PPH4 = 0.90). Similarly, we show that *BIN1* expression in peripheral monocytes can mediate the risk of AD in Europeans.
These results suggest that peripheral immune responses associate with the risk of Alzheimer's disease and Parkinson's disease, contributing to the hypothesis that neurodegeneration interacts with, and may be driven by, peripheral responses to the environment. The implications of this include the possibility of treating neurodegeneration via peripheral targets, as well as providing a mechanistic understanding regarding how environmentally-based selection pressures influence these diseases in an ancestry-specific manner. Finally, this work demonstrates that while increasing efforts to generate ancestry-specific GWAS are important, equal effort must be placed on diversity of functional genomic annotation to enable multi-ancestry GWAS analyses to be used effectively and for global benefit.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5157F: Sex-specific transcriptional differences in Gulf War Illness patients under stressful conditions.

Authors:

J. Frank[1], L. Tehrani[2], D. Van Booven[3], S. Ballarin[2], R. Rossman[2], A. Reuthebuck[4], S. Uppalati[4], A. Edelstein[2], R. Pangeni[1], N. Klimas[1], L. Nathanson[1]; [1]Nova Southeastern Univ., Dr. Kiran C. Patel Coll. of Osteopathic Med., Inst. for Neuro Immune Med., Fort Lauderdale, FL, [2]Nova Southeastern Univ., Dr. Kiran C. Patel Coll. of Osteopathic Med., Fort Lauderdale, FL, [3]Univ Miami Miller Sch Med, Miami, FL, [4]Nova Southeastern Univ., Fort Lauderdale, FL

Abstract:

Background. Gulf War Illness (GWI) is a severely debilitating condition with chronic symptoms affecting up to 30 percent of the veterans deployed to the 1990-91 Persian Gulf War. The symptoms associated with GWI include fatigue, musculoskeletal/joint pain and disturbances of multiple organ systems. The mechanisms of GWI are unknown, although the symptoms suggest that GWI is most likely caused by a combination of genetic predisposition and it's interaction with environmental factors such as exposure to toxic chemicals. **Objectives.** This research project aims to assess how sex differences in individuals with GWI impact their response to stress, specifically through an exercise challenge. Additionally, it highlights the importance of considering sex differences in the diagnosis and treatment of GWI. **Methods.** We utilized RNA-seq to examine the differential gene expression (GE) in peripheral blood mononuclear cells of 24 female GWI patients, 20 male GWI patients, and 40 healthy controls (HC) who were matched in for sex, age, and BMI. We evaluated the gene expression changes at three time points: T0 (baseline), T1 (at maximal exertion), and T2 (4 hours into recovery after T1). To ensure consistency, we excluded genes located on the sex chromosomes and performed separate analyses for men and women. Our focus was to compare the response to exercise within each group (GWI patients and HC) between different time points. **Results.** In the male GWI cohort, pathways associated with immune system process, particularly IL-12, and natural killer cell cytotoxicity were activated during exercise. The female GWI patients exhibited less significant changes in gene expression than men. During the recovery period, male GWI patients demonstrated distinct alterations in the regulation of innate immune responses. On the other hand, female GWI patients showed significant changes in genes related to regulation of intracellular signal transduction. These functional pathways and differentially expressed genes identified in this project provide valuable insights into the sex-specific pathophysiology of GWI. **Conclusions.** The identification of sex-specific biomarkers and therapeutic targets in GWI holds great potential for understanding the distinct onset and progression of this disease in different sexes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5158F: Shared genetic basis of obesity, type 2 diabetes, and Alzheimer's disease: Novel insights into sex and age-specificity

Authors:

C. Cunha[1,2], R. Martinez[1], R. Loos[1,2], T. Kilpeläinen[1,2]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., University of Copenhagen, Denmark, [2]Novo Nordisk Fndn. Ctr. for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, MA

Abstract:

Obesity and type 2 diabetes (T2D) are major modifiable risk factors for Alzheimer's Disease (AD). Whether obesity and T2D may share genetic risk factors with AD remains unclear. We aimed to identify shared genetic risk loci and elucidate their functional roles by leveraging the largest available GWAS summary statistics for BMI, T2D, and AD from the GIANT Consortium, T2D Global Genomics Initiative, and European Alzheimer's Disease DNA BioBank, respectively. Recognizing the complex relationship between BMI and AD during aging, and the elevated AD risk in women, we incorporated sex and age-stratified BMI and sex-stratified T2D data in our analysis. We observed a global negative genetic correlation between BMI and AD (r_g= -0.13), as well as T2D and AD (r_g= -0.09). The genetic correlation between BMI was more pronounced in women (r_g= -0.16) than men (r_g= -0.11), particularly in women under 50 years of age (r_g=[-0.22; -0.18]). Through colocalization analysis, we identified four loci associated with BMI and AD, and two distinct loci associated with T2D and AD. In addition, we identified one locus concurrently associated with BMI, T2D, and AD. Notably, the four BMI-AD loci showed a pronounced effect on BMI in women below 50, while the BMI-T2D-AD locus had a more pronounced effect on BMI in women over 50. This locus was also associated with earlier menarche and lower sex-hormone binding globulin levels. We prioritized the most likely effector genes in each locus based on colocalization with coding variants, fine-mapped cis-eQTLs, and enhancer-gene pairs. The effector genes within the BMI-AD and T2D-AD loci are predominantly expressed in peripheral tissues, including adipose tissue, vasculature, endocrine glands, and female gonads. In contrast, the prioritized effector gene in the BMI-T2D-AD locus, *DOC2A*, is primarily expressed in the brain, specifically in prefrontal cortical excitatory neurons and hypothalamic ependymal cells. A perturbation screen evaluating cellular responses to pharmacological agents showed changes in protein expression encoded by our prioritized effector genes in response to 100 unique compounds, seven of which relate directly to cardiometabolic conditions. Overall, our findings indicate that while the genes identified for BMI-AD and T2D-AD act peripherally, the BMI-T2D-AD locus points to central, brain-related processes. Our sex-stratified analyses highlighted the role of women-specific genetic regulators of obesity, T2D, and AD, potentially elucidating the heightened AD risk in women. These insights into the shared genetic basis of cardiometabolic diseases and AD may open avenues for new prevention strategies and drug repurposing.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5159F: Single cell multi-omic profiling in hiPSC-derived neurons reveals regulatory mechanisms of neuron activity dependent response and genetics of neuropsychiatric disorders ⭐

Authors:

L. Liang[1], Z. Wang[2], S. Zhang[3], H. Zhang[3], C. Li[4], Z. Pang[3], A. Sanders[3], J. Duan[5], X. He[2]; [1]Univ. of Chicago, Chicago, IL, [2]The Univ. of Chicago, Chicago, IL, [3]North Shore Univ Hlth.System, Evanston, IL, [4]Northshore Res. Inst., Evanston, IL, [5]Northshore Univ Hlth.system/Univ of Chicago, Evanston, IL

Abstract:

How neurons respond to stimulation is fundamental to neurodevelopment and neuronal function. Dysregulation of this process, including that from genetic variations, underlies many neuropsychiatric disorders (NPD). However, neuronal activity-dependent gene regulatory networks, and to what extent genetic variation can influence this process and its relevance to NPD are largely unknown.

To fill in these gaps, we profiled single-cell RNA and ATAC-seq in 100 iPSC derived neurons (>700K) co-cultured with glia, in untreated and stimulated conditions. Focusing on GABAergic and two subtypes of Glutamatergic neurons, we found complex patterns of expression dynamics, summarized into 15 regulatory modules. Interestingly, while early responses to stimulation are shared across cell types, some late response programs are cell type-specific. Using ATAC-seq data, we linked genes with regulatory elements (REs). Interestingly, some REs of early response genes show prolonged chromatin openness, suggesting “epigenomic memory”. Integrative analysis discovered a core set of Transcription Factors (TFs) involved in regulation of multiple modules, including the well-known AP-1 family, and other TFs, such as SMARCC1 and BACH2. We also found TFs specifically regulating early responses, including SREBF2, a master regulator of lipid biosynthesis; and several TFs regulating GABA-specific late response.

We mapped expression and chromatin QTLs. Focusing on eQTLs, we found ~8,000 eGenes. ~1,600 of these eQTLs showed variable effects across cell types and conditions. Dynamic eQTLs (stimulation-specific) tend to be more cell type specific compared to other eQTLs, and are less likely to be shared with adult brain eQTLs. Using a recently published method, cTWAS, we integrated eQTL data, across multiple cell types and conditions, with GWAS of NPDs. Our results highlighted nearly 100 likely causal genes of NPDs. In the majority of cases, cTWAS suggested that causal variants of NPDs are dynamic eQTLs. Interestingly, the candidate genes were enriched for lipid metabolism.

Altogether, our study uncovered complex expression dynamics and regulatory mechanisms of neuron stimulation, highlighting the importance of this process in the genetics of NPDs.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5160F: Sleep inertia drives the causal association of evening chronotype with psychiatric disorders: genetic and epidemiological evidence

Authors:

A. Burns[1], D. Windred[2], I. Daghlas[3], M. Rutter[4], H. Dashti[5], A. Phillipps[6], S. Cain[2], H. Ollila[7], R. Saxena[8], J. Lane[9]; [1]Harvard Med. Sch.; Broad Inst., Boston, MA, [2]Monash Univ., Melbourne, Australia, [3]Stanford Univ., Stanford, CA, [4]Univ. of Manchester, Manchester, United Kingdom, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Flinders Univ., Adelaide, Australia, [7]Univ. of Helsinki, Helsinki, Finland, [8]Mass Gen. Hosp., Harvard Med. Sch., Boston, MA, [9]BWH, Boston, MA

Abstract:

Late chronotypes ("night owls") have a consistently higher risk of developing psychiatric disorders. The proposed mechanism for this is greater risk for circadian misalignment in this group. Sleep inertia, or difficulty awakening, acts as a proxy for the phase-angle of awakening and circadian misalignment. Here, we complete the first genome-wide association study of sleep inertia in the UK Biobank cohort (457,737 European individuals) and identified 98 independent genomic risk loci ($p < 5x10^{-8}$; $\lambda_{GC} = 1.52$; SBayesS SNP $h^2 = 0.08$) that were enriched for circadian rhythm-related genes such as *PER2, PER3, CRY1* & *CIART*. Bivariate Gaussian mixture modeling identified significant polygenic overlap between chronotype and sleep inertia (MiXeR $r_g = 0.82$), however significant unshared polygenicity was observed for both traits and 16% of the shared variants were discordant. Marginal Mendelian randomization analyses (MR; GSMR2) identified significant causal associations of late chronotype with increased risk for MDD, suicide attempt, anxiety, schizophrenia and bipolar disorder (all $p < 0.001$). However, in MR analyses conditioning on sleep inertia, the causal association of late chronotype with these psychiatric disorders was ablated and in some cases reversed. Epidemiological analyses in the UK Biobank cohort and Finnish Older Twin Cohort corroborated these findings, such that the effect of late chronotype on psychiatric disorder risk was in all cases entirely mediated by sleep inertia. Our findings suggests that circadian misalignment in people with late chronotype increases the risk of psychiatric disorders, not simply their delayed timing. These results challenge the common understanding that late chronotypes are at increased risk for psychiatric disorders, suggesting it is dependent on phase alignment.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5161F: Spectrum of genetic variants in a cohort of patients with long QT interval

Authors:

B. Burnyte, R. Meskiene, A. Kemezyte; Vilnius Univ., Vilnius, Lithuania

Abstract:

Objective: Long QT syndrome (LQTS) is caused by pathogenic variants in the genes encoding ion channels, which can lead to disturbances in ventricular repolarization. This increases the risk of cardiac arrest or death in young patients. We aimed to define the clinical and genetic findings in a cohort of pediatric patients with suspected LQTS. Methods: In this retrospective study, we included patients <18 years of age with suspected LQTS who presented to genetic testing (2021-2022). Results: Clinical and genetic data from 92 unrelated patients (55%, 51/92 of females) were analysed. The mean age at time of testing was 10±4.7SD years. Pathogenic or likely pathogenic variants were identified in 17% (16/92) of those patients. The most implicated gene was *KCNQ1*, with 69% (11/16) of the patients affected. Variants in *KCNH2* gene were found in 12% (2/16) of patients, while 19% (3/16) were distributed in other genes. Many variants of unknown significance were identified. A quarter of patients (25%, 4/16) had a first degree relative with clinical LQTS. Among the patients, mean Schwarz score and longest measured QTc interval were 3.4±1SD and 497ms±32SD, respectively. More than half (62%, 10/16) were asymptomatic, diagnosed accidentally based on ECG findings and/or family history. Conclusion: This study confirms the importance of genetic testing for LQTS patients showing 17% of them to benefit from genetic result that will guide their clinical management.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5162F: *Staphylococcus aureus*bacteremia is associated with *HLA DRB1* amino acid changes 26Y and 71K in a multi-ethnic sample

Authors:

W. Scott[1], D. A. Dorfsman[1], D. M. Dykxhoorn[1], F. R. Ruffin[2], Y. Tao[2], J. L. Kair[2], A. M. Jackson[2], V. G. Fowler, Jr[2]; [1]Univ. of Miami, Miami, FL, [2]Duke Univ. Med. Ctr., Durham, NC

Abstract:

Bloodstream infection by *Staphylococcus aureus* is a frequent complication of inpatient healthcare. *S. aureus* infection or bacteremia (SAB) is a component of several genetic syndromes. Risk of SAB is increased in individuals with complex diseases, such as rheumatoid arthritis and diabetes, that have strong genetic risk factors. Prior studies have associated risk of SAB with single nucleotide variants (SNV) between *DRA1* and *DRB1* in the human leukocyte antigen (HLA) class II region, and this association may be partially attributed to the HLA class II DR4 haplotype clade. To further investigate the association of HLA class II variation with SAB, we conducted a regional association study in 3,194 inpatients (1,616 with SAB, 1,578 without) from Duke University Medical Center. Illumina GSA arrays were used to genotype participants, the TopMed Imputation Server was used to impute 38,077 HLA SNV (chr6: 29,000,111-33,999,677) and the Michigan Imputation Server was used to impute canonical haplotypes in the HLA class I and II regions using SNP2HLA and a multi-ancestry reference panel (Sakaue et al., 2023). We considered 3,236 SNV in a 25 kilobase region identified as harboring elevated European ancestry in an admixture mapping study of SAB (Cyr et al., 2017). Logistic regression models were constructed in the R package GENESIS to evaluate association of each SNV or haplotype with SAB, including 10 principal components generated by PCAir to control for population structure. Age, sex, and ethnicity were not significantly associated with SAB and were not retained as covariates. Nominal association ($p<0.05$) was detected at 72 SNV, but no SNV was significant after Bonferroni adjustment for multiple comparisons. Analysis of haplotypes in the HLA *DRB1* locus did not identify any one- or two-field haplotypes significantly associated with SAB, although *DRB1**03:01 was 1.17 times more likely in people with SAB than controls (95% CI: 0.99-1.37); $p=0.06$). Considering coding sequence variation at the amino acid level, 4 of 167 amino acid changes in *DRB1* were associated ($p<0.05$) with SAB: 26Y and 71K were positively associated with SAB and 30H and 67I were negatively associated. Notably, 71K is present on both the *DRB1**03:01 and *DRB1**04:01 haplotypes, which are positively associated with SAB in this and prior studies, respectively. *DRB1**03:01 (which carries 26Y and 71K) is associated with increased risk of type 1 diabetes, and risk of SAB is increased in individuals with diabetes. These results suggest that the development of SAB is in part influenced by HLA variation, and that variants that increase risk of autoimmune disease may also predispose to development of bacteremia.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5163F: Structural Haplotypes in the 17q21.31 *MAPT* Region and Cognitive Resilience to Alzheimer's Disease Pathology

Authors:

X. Han[1], T. Stein[1,2,3], L. Farrer[1,4], J. Mez[1], G. Jun[1,4]; [1]Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [2]VA Boston Hlth.care System, Boston, MA, [3]VA Bedford Hlth.care System, Bedford, MA, [4]Boston Univ. Sch. of Publ. Hlth., Boston, MA

Abstract:

IntroductionAlzheimer's disease (AD) pathology, including neuritic amyloid-β plaques and neurofibrillary tangles, is often found in postmortem brains of resilient individuals who were cognitively normal prior to death, [herein termed Asymptomatic AD (AsymAD)]. The 17q21.31 region, including *MAPT*, contains a megabase-long inversion polymorphism (H1/H2) and many copy number variations (CNVs), including α, β and γ segments, which determine various structural haplotypes. Here we imputed 17q21.31 structural haplotypes and tested their associations with AsymAD compared to symptomatic AD (SymAD).**Methods**We obtained genome-wide single nucleotide polymorphism (SNP) array and brain RNA sequencing data from 439 brain donors with SymAD, 349 with AsymAD from the Rush Religious Orders Study and Memory and Aging Project (ROSMAP), Boston University Alzheimer's Disease Research Center (BUADRC), and Framingham Heart Study (FHS). We imputed Inversion status (H1/H2), CNVs in α/β/γ segments, and 5 common subhaplotypes (H1β1γ1, H1β1γ2, H1β1γ3, H1β2γ1, and H2α2γ2) using the Boettger et al. reference haplotype panel, SHAPEIT and IMPUTE2. We tested their associations between AsymAD and SymAD, as well as with several cognitive tests. We further evaluated associations of these SVs with gene expression levels in the 17q.21.31 region and examined their differential expression patterns between SymAD and AsymAD.**Results**We identified 2 subhaplotypes were protective against SymAD in brains with AD pathology: H1β1γ2 (Freq=0.10, OR=0.61, FDR=0.03) and H1β1γ3 (Freq=0.07, OR=0.58, FDR=0.04). Conversely, the H1β2γ1 subhaplotype was a risk for SymAD (Freq=0.24, OR=1.44, FDR=0.03). Among AD brains carrying H1 homozygotes, CNVs in the γ segment were protective against SymAD (β=-0.51, FDR=2.3×10^{-4}), whereas CNVs in the β segment increased the risk for SymAD (β=0.14, FDR=0.03). Additionally, β segment CNVs were significantly associated with increased levels of *KANSL1* (β=52.7, P=1.4×10^{-15}), which was also upregulated in SymAD (Log2FC=0.08, P=5.9×10^{-6}) compared with AsymAD. These associations among AD brains were probably driven by performance in memory and language. For example, we observed nominal associations (P<0.05) of γ segment CNVs with improved performance in several cognitive tests, including Logical Memory (immediate and delayed recall), the Boston Naming Test, and verbal fluency tests (phonemic FAS and categorical Animal Naming) for BUADRC and FHS participants.**Conclusion**Our study identified structural variants in the 17q21.31 region associated with cognitive resilience among brain donors with AD pathology.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5164F: T-cell receptor repertoire sequencing at single-cell resolution in endemic pemphigus foliaceus

Authors:

V. Calonga Solis[1], T. Della Justina Farias[1], I. Wolf[1], G. A. Cipolla[2], D. Malheiros[2], M. Petzl-Erler[2], D. G. Augusto[1]; [1]The Univ. of North Carolina at Charlotte, Charlotte, NC, [2]Univ.e Federal do Parana, Curitiba, Brazil

Abstract:

Pemphigus foliaceus (PF) is a rare autoimmune skin disease with a global incidence of 1.1 per million per year. However, PF presents an exceptionally high prevalence in certain regions of South America and Northern Africa, probably due to environmental factors, making it the only autoimmune disease to be named endemic. The highest prevalence is seen in some areas of Brazil, where it can reach astonishing rates of more than 3%. As an antibody-mediated disease with a strong genetic component, we have previously explored the B cell receptor (BCR) repertoire in PF. Now, we performed a challenging and pioneering evaluation of the characteristics of T cells and the T-cell receptor (TCR) repertoire in this disease, through RNA sequencing of single cells. The variable region of TCRs are encoded by genes that are non-functional in the germline DNA and can only be transcribed after somatic rearrangement of three types of gene segments (V, D and J). The somatic recombination generates a high diversity of functional sequences, and each T cell will express a different TCR rearranged gene. The total set of these transcripts in an individual is the TCR repertoire. To better understand the role of activation and clonal expansion of these cells in PF, we conducted a single-cell RNA sequencing of peripheral blood mononuclear cells from four PF patients with active lesions and without any treatment and four healthy controls residing in an endemic area of Brazil. We obtained the whole transcriptome of 5,154 T cells, encompassing CD8+, CD4+, and regulatory T cells (Tregs). Next, we sequenced the TCR alpha and beta transcripts, revealing the TCR repertoire of 2,692 cells. Our findings indicate that patients with PF exhibit a higher proportion of high-frequency clonotypes (hyperexpanded clonotypes) compared to controls ($p < 8x10^{-4}$). Further analysis of cell types expressing these hyperexpanded clonotypes showed significant differences between patients and controls. Specifically, the proportion of effector memory CD8+ T cells was significantly elevated in patients (71.5%) compared to controls (32%), while the proportion of naïve CD4+ T cells was reduced in patients (8.4%) relative to controls (32.9%; $p < 2x10^{-16}$). Additionally, we identified clonotype sequences shared among the four patients, suggesting a common antigenic exposure or shared immune response pattern within this group. These results provide novel insights into T cell dynamics and clonal expansion in PF, including identifying disease-specific clonotypes. Our groundbreaking findings offer potential pathways

for understanding disease mechanisms and developing therapies to target specific pathogenic T and B cells.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5165F: The Association of a Polygenic Risk Score for Idiopathic Pulmonary Fibrosis with Diagnosis and Mortality in Biobank Cohorts

Authors:

S. Kalra[1], Q. Chen[2], C. Tern[2], J. Lasky-Su[2], G. Hunninghake[1], E. Silverman[2], D. Schwartz[3], M. Moll[2], M. Cho[2]; [1]Brigham and Women's Hosp., Boston, MA, [2]Channing Div. of Res. Med., Boston, MA, [3]Natl. Jewish Hlth., Denver, CO

Abstract:

BACKGROUND AND AIMS: Idiopathic Pulmonary Fibrosis (IPF) is characterized by chronic progressive pulmonary fibrosis and high mortality. A polygenic risk score (PRS) for IPF associates with IPF in case-control studies, however the PRS association with diagnosis and trajectory of IPF in hospital administrative data from biobanks is unclear. **METHODS**: We conducted an observational study including unrelated, self-reported white participants with genetic data in the Mass General Brigham (MGB) and UKBiobank. We tested the association of the PRS with an International Classification of Disease (ICD) IPF diagnosis using multivariable logistic regression, and with transplant-free survival among those with IPF using multivariable Cox proportional hazards, adjusting for age, gender, *MUC5B* genotype dose and principal components of ancestry. **RESULTS**: Of 18,583 and 458,573 participants from MGB and UKBiobank respectively, 552 (3%) and 2760 (0.6%) participants had IPF. The PRS was associated with IPF diagnosis in MGB (OR 1.34 95%CI 1.21-1.47) and UKBiobank (OR 1.35 95%CI 1.30-1.41). Among those with IPF, the PRS was associated with transplant-free survival in MGB (HR 1.17 95%CI 1.00-1.36, p = 0.053) and UKBobank (HR 1.15 95%CI 1.06-1.24, p = 0.00064). **CONCLUSIONS**: The PRS for IPF is significantly associated with a diagnosis of IPF in two large biobank cohorts. Among those with IPF, the PRS is associated with transplant-free survival.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5166F: The Genetic Etiology of Alzheimer’s Disease Mediated by Regional Brain Gray Matter Densities

Authors:

J. Bao[1], S. Mu[1], H. Xu[1], E. Mamourian[1], Q. Zhan[1], G. Erus[1], H. Shou[1], C. Davatzikos[1], Q. Zhang[1], J. WEN[2], Q. Long[1], L. Shen[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Southern California, LA, CA

Abstract:

Background: Alzheimer's disease (AD) is a prevalent neurodegenerative disorder characterized by high heritability and accelerated brain atrophy. While numerous studies have identified genetic loci associated with AD onset and highlighted specific brain regions of interest (ROIs) influencing disease outcomes, few have integrated these approaches to explore the pathway from genetics to regional brain volumes and, ultimately, AD diagnosis. To address this gap, we perform a mediation analysis to uncover the genetic etiology of AD mediated by regional brain gray matter densities.
Methods: Data for this study were sourced from the Alzheimer's Disease Sequencing Project (ADSP) study hosted by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689). We downloaded and quality-checked whole genome sequencing data, T1-weighted MRI-derived endophenotypes preprocessed using the MUlti-atlas region Segmentation (MUSE) pipeline, harmonized AD case-control phenotypes, and population characteristics. A genome-wide association study (GWAS) was performed separately for white and other populations, adjusting for age, sex, and the first 10 principal components from genotyping data. In the white population, mediation analyses were conducted for SNP variants meeting the 5e-8 significance threshold. Significant mediation signals were then analyzed in a replication study in non-white populations.
Results: Our study included 6,888 AD cases and 6,475 controls in the white population, and 3,537 cases and 10,395 controls in non-white populations. Out of 5,607,693 SNP variants analyzed, 340 reached genome-wide significance. These significant SNPs were used as exposures in mediation analyses, with 119 MUSE-preprocessed regional gray matter densities as mediators and the AD case-control phenotype as the outcome. After FDR correction, 2,716 significant mediation signals were identified in the white population. Replication in non-white populations yielded 126 mediation signals at a p-value threshold of 0.05.
Conclusion: Our brain-wide genome-wide mediation analyses identified 126 mediation signals replicated across different ancestries, addressing the gap in imaging-genomics etiologies of AD. Future work will involve ancestry-specific analyses across various studies, including ADSP, UK Biobank, and Penn Medicine Biobank, to discover more robust and reliable signals.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5167F: The role of genetics in predicting progression to severe obesity in midlife

Authors:

S. Harrison[1], V. Plagnol[1], M. Weale[2], D. Wells[1], P. Donnelly[3]; [1]Genomics plc, Cambridge, United Kingdom, [2]King Charles House, Oxford, United Kingdom, [3]Genomics plc, Oxford, United Kingdom

Abstract:

BackgroundTargeted prevention of obesity and related complications is a major challenge for healthcare systems and would benefit from identification of individuals with the highest risk of weight gain. We aimed to evaluate if polygenic risk scores (PRS) would be useful in predicting clinically relevant weight trajectories, in adults in the UK. This is particularly important given the availability of novel weight loss drugs.
MethodsLeveraging data from 1,365,965 individuals, we developed and optimised a multi-ancestry polygenic risk score (PRS) for body mass index (BMI). In UK Biobank, we identified a separate testing subset of 3,953 individuals with BMI measured both in early adulthood (26 - 34 years, before recruitment to UK Biobank) and midlife (46 - 54 years, after recruitment to UK Biobank), and evaluated if PRS could predict transitions across BMI categories, relevant to clinical practice.
ResultsThe optimised PRS explained 14.6% phenotypic variance in individuals with European ancestry, and 8.5% in non-European ancestries. For individuals in the top 5% of the PRS distribution, the risk of transitioning from healthy weight (BMI<25) to obesity (BMI≥30) was 16.8% (95% CIs 10.71%-24.53%), compared to 0.8% (95% CIs 0.02%-4.38%) for those in the bottom 5%. Possession of a high PRS (top 5%) was associated with 10.0% of all healthy -> obesity transitions (2.0-fold enrichment, 95% CI 1.26 - 2.98), and for 23.5% of all individuals presenting as severely obese (BMI≥40) at recruitment (4.7-fold enrichment, 95% CI 4.41-4.97). In contrast, possession of a low PRS (bottom 5%) was associated with 13.4% of all reversions from overweight status (BMI≥25) in early adulthood to healthy weight in midlife (2.7-fold enrichment, 95% CI 1.56 - 4.15).
ConclusionA PRS developed for BMI can identify individuals in early adulthood who are most likely to transition to obesity in midlife, thereby providing a window for intensive education and earlier intervention, to avoid complications of obesity. Policies aimed at obesity prevention should consider inclusion of genetic information.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5168F: The role of *Tumor necrosis factor-alpha* and *Interleukin-6* gene polymorphisms in genetic susceptibility of juvenile idiopathic arthritis in Kuwaiti Arab children

Authors:

M. Haider[1], M. Alsaeid[1,2], K. Alsaeid[1,3]; [1]Kuwait Univ., Coll. of Med., Jabriya, Kuwait, [2]Amiri Hosp., Dasman, Kuwait, [3]Mubarak Alkabeer Hosp., Jabriya, Kuwait

Abstract:

Juvenile idiopathic arthritis (JIA) is a common autoimmune disease in which a combination of risk alleles from different susceptibility genes predisposes the patients to develop clinical symptoms following exposure to environmental factors. Several immune related risk factors have been associated with JIA; these include single nucleotide polymorphisms (SNPs) in some cytokine genes. These cytokine gene polymorphisms are thought to influence the levels of these cytokines in serum and thus can impact the disease-onset and/or progression. The objective of this study was to investigate the role of Tumor necrosis factor-alpha (*TNF-alpha*) and Interleukin-6 (*IL-6*) gene polymorphisms in genetic susceptibility of JIA in Kuwaiti children. We have determined the genotypes of *TNF-alpha* gene (-308A/G; rs1800629) and *IL-6* gene (-174G/C; rs1800795) polymorphisms in 78 Kuwaiti children with JIA and compared it to that in approximately 200 healthy controls of the same ethnicity. The diagnosis of JIA was based on the American College of Rheumatology (ACR) criteria. The genotypes for *TNF-alpha* and *IL-6* gene polymorphisms were determined by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) methods and confirmed by DNA sequencing. The frequency of GG genotype of *TNF-alpha* gene polymorphism showed a strong association with JIA in both co-dominant and dominant models of genetic analysis ($p = 0.0004$ and 0.003 respectively). In contrast, the GG genotype of *IL-6* gene polymorphism showed a significant association with JIA only in the dominant model ($p = 0.01$). Amongst the two clinical subgroups of JIA patients, 98% of oligoarticular and 100% of the polyarticular JIA patients had at least one 'G' allele of the *TNF-alpha* gene polymorphism. The distribution of *IL-6* gene polymorphism genotypes in JIA subgroups showed that 35% of the oligoarticular and 13% of the polyarticular JIA patients carried at least one 'G' allele. Our data showed significant association of two cytokine gene polymorphisms (i.e. *TNF-alpha* gene -308A/G, rs1800629 and *IL6* gene -174G/C; rs1800795 polymorphism) with JIA in Kuwaiti children highlighting their significant contribution in genetic susceptibility of this chronic disease possibly along with other factors.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5169F: Trajectory analysis reveals islet cell depletion exacerbated in individuals of African ancestry with Type 2 diabetes ⭐

Authors:

E. Heuston, A. Oteng, A. Doumatey, C. Rotimi, A. Adeyemo; Ctr. for Res. on Genomics and Global Hlth., NHGRI, NIH, Bethesda, MD

Abstract:

Type 2 diabetes (T2D) is a cardiometabolic disease disproportionately affecting individuals of African ancestry in the United States. A unique pathophysiology in this group could be due to ancestry-specific risk variants impacting islet function and which could be identified using single cell trajectory analysis to model cell development. We hypothesize that comparing pancreatic cell scRNA-Seq profiles from donors of European (EA) or African ancestry (AA) with and without T2D (T2D+, T2D-) will reveal unique trajectories between these cell populations.
We analyzed scRNA-Seq data from the Human Pancreas Analysis Program for 38 T2D- (84000 EA and 30000 AA cells) and 14 T2D+ (46000 EA and 7400 AA cells) individuals. After quality control 167188 cells were pooled and analyzed in *Seurat5* to generate 24 clusters. We used over representation analysis (ORA) and *Garnett* to assign cluster identities. EA T2D+ and T2D- cell categories were downsampled to match cell numbers in the corresponding AA categories before *Monocle3* trajectory analysis to model relationships between cell clusters.
T2D- AA had 2 of 5 alpha cell clusters depleted (36%, 20%) compared to EA (64%, 80%). Alpha_1 cluster depletion (high *CRH* expression, linked to reduced apoptosis) produced a poorly defined trajectory compared to EA models. Alpha_2 cluster depletion (high BIRC5 and *PTTG1* expression, mitosis-related) resulted in complete loss of a trajectory branch. 1 of 4 beta cell clusters was also significantly depleted (15% AA, 20% EA) resulting in the loss of a terminal trajectory branch. Interestingly, a population of acinar cells was significantly depleted in AA (16% AA, 32% EA).
T2D+ AA alpha and beta cell clusters were too depleted to generate trajectory models, while the downsampled EA clusters maintained similar trajectories compared to their T2D- counterparts. Alpha cells dropped from 53% to 12% of all cells in AA, compared to an unchanged 32% in EA. Beta cells decreased from 28% to 6% of AA, versus a 37% to 25% decrease in EA. Acinar cell clusters maintained similar trajectories between AA and EA T2D- and T2D+ donors, implying that these observations are specific to AA alpha and beta cells. Finally, AA and EA donors shared a unique T2D+ cluster enriched for Zinc Homeostasis and Mineral Absorption activities (FDR ≤ 0.001), echoing emerging evidence of metallothionein activity in beta cell function.
Our findings show that pancreatic cells show different trajectories between AA and EA that is observable in donors without T2D and is exacerbated in T2D. These findings as well as

differences in cellular cluster composition may contribute to the basal and stimulated measures of beta cell dysfunction observed in AA.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5170F: Transcriptomic characterization of induced pluripotent stem cell-derived neural cells from patients with Williams Syndrome

Authors:

C. da Silva Camargo, S. Detera-Wadleigh, N. Akula, A. Schulmann, M. Gregory, D. Eisenberg, K. Berman, F. McMahon; NIH/NIMH, Bethesda, MD

Abstract:

Williams Syndrome is a rare genetic neurodevelopmental disorder caused by a hemizygous deletion on 7q11.23. The syndrome is characterized by mild intellectual disability, specific cognitive traits, distinctive facies, overly social comportment, and other system abnormalities. While this syndrome has a well-defined cognitive/behavioral profile, the neurocellular and molecular underpinnings are not well-defined. To model the neurobiological mechanisms involved in neural cell development, we analyze neural progenitor cells (NPCs) and cortical neurons differentiated from iPSCs derived from a group of 7q11.2 deletion carriers and normal siblings. Single cell RNA sequencing (scRNAseq) on NPCs from an initial sample derived from three WS and one control revealed clusters of differentially expressed genes (DEGs) representing three stages of temporal differentiation, the majority of which comprise neural cells at earlier developmental stages that are positive for the NPC markers, nestin and PAX6. The remaining neural cells showed progression toward maturation as evidenced from genes that belong to the GABAergic and glutamatergic categories. Genes localized in the 7q11.23 deleted region such as GTF2I and LIMK1 were downregulated, confirming the genotype expected for the cellular provenance. DEGs in earlier stage were enriched for the cell cycle and mitotic processes. Those DEGs in the later stage of NPC maturation were enriched for various biological processes including nervous system process and synaptic signaling. Future directions include multi-omics analysis of NPCs and cortical neurons from additional individuals. Emergent genetic and molecular profiles will be correlated with available extensive neuroimaging and clinical phenotypes for these study samples. The neurobiological and neuromolecular characterizations may reveal signature pathways that could be targeted in the development of new, efficacious therapeutic agents.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5171F: Transferability of tobacco and alcohol use polygenic scores across populations

Authors:

G. Saunders[1], S. Vrieze[1], D. Liu[2]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Penn State Univ., Hershey, PA

Abstract:

Polygenic risk scores (PRS) based on a recent genome-wide association study (GWAS) of nearly 2.7 million individuals have been found to explain up to 9.6% of the variance in alcohol and tobacco use. This level of predictive power, however, is only true for individuals of European ancestries, the same ancestries upon which the discovery GWAS was based. The relative accuracy of scores derived from this same GWAS were between 0.5% and 62% in individuals of non-European ancestry, with average relative accuracies of 12% for African (AFR) ancestries, 26% for East Asian (EAS) ancestries, and 31% for American (AMR) ancestries. This is consistent with a growing body of literature showing that the predictive power of polygenic scores declines with increasing genetic distance from the discovery sample ancestry. With the potential of PRS to provide clinical, research, and personal utility, understanding of the causes of reduced accuracy, and relative contributions to the problem, has important implications for equitable potential of PRS. In the current study we use summary statistics from a large-scale GWAS of tobacco and alcohol use to describe the accuracy of European-based PRS across diverse ancestries using approximately 250,000 individuals from the All of Us cohort. We find that the relative accuracy of PRS for smoking behaviors is 32% for AMR, 26% for EAS, and 10% for AFR ancestries. Further, we find that the predictive accuracy of scores decline as a function of genetic distance (GD) from the discovery GWAS even within ancestry groups. For example, within EUR individuals in All of Us, those in the GD bin furthest from the discovery sample show a predictive accuracy 50% lower than those in the closest GD bin. Comparisons of predictive accuracy by allele frequency (AF) and recombination rates suggest that differences in AF and linkage disequilibrium explain much, but not all of the reduced prediction (accounting for approximately 85% of the loss in AFR ancestries, 58% in EAS ancestries, and 25% in AMR ancestries). We compare these results with those of medical and anthropometric traits to evaluate how patterns differ as a function of phenotype. The results of this work inform efforts to understand the genetic architecture of complex traits across genetic ancestries and to create more accurate polygenic risk scores for individuals from ancestries that have largely been left out of genomic research.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5172F: Type 2 Diabetes protective effects of rare loss of function variants in African Americans: an All of Us cohort study

Authors:

S. Sharma[1], C. Astore[1], L. Marino-Ramirez[2], I. Jordan[1]; [1]Georgia Inst. of Technology, Atlanta, GA, [2]NIH, Rockville, MD

Abstract:

The NIH All of Us Research Program (All of Us) is one of the world's most diverse population biobanks in support of equitable precision medicine. Type 2 diabetes is a complex disease that disproportionately affects African Americans in the United States. Although multiple genome- and exome-wide association studies have been performed for type 2 diabetes, these studies have relied on predominantly European ancestry cohorts. We hypothesized that studying the effects of type 2 diabetes rare loss of function variants in African and European ancestry cohorts could yield potentially novel therapeutic targets. In pursuit of this aim, we conducted exome-wide and gene-based association studies using the All of Us Research Program cohort ($n_{African}$ = 27,982; $n_{European}$ = 88,932). We identified rare predicted loss-of-function (pLOF) variants (variant frequency < 0.01) in the *SLC30A8*, *STX19*, & *UCP2* genes for the African ancestry cohort and the *SNX18* & *KDELR3* genes for the European ancestry cohort. We further examined the effects of these rare pLOF variants on blood hemoglobin A1C (HbA1C) levels and body mass index (BMI). We identified pLOF rare variants in *SLC30A8* associated with lower prevalence of type 2 diabetes, reduced average HbA1C levels, and lower BMI in the African ancestry cohort. Accordingly, we propose *SLC30A8* as a potential type 2 diabetes therapeutic target, which would not have been discovered via analysis of a European ancestry cohort alone. Our findings demonstrate that genetic association studies on ancestrally diverse cohorts can power the discovery of novel therapeutic targets for complex diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5173F: Understanding Lecithin-cholesterol acyltransferase Deficiency

Authors:

A. Wokhlu[1], V. Prakash[2]; [1]Univ. of Texas at Dallas, Richardson, TX, [2]Dallas Renal Group, McKinney, TX

Abstract:

Lecithin-cholesterol acyltransferase (LCAT) deficiency is a rare autosomal recessive genetic disorder with a prevalence of 1 in 1,000,000 individuals. LCAT deficiency results from a missense mutation that replaces methionine with arginine in position 293 in the amino acid sequence of the LCAT protein. LCAT is vital for converting free cholesterol into cholesteryl ester, which is necessary for high-density lipoprotein (HDL) function. Patients who showcase a complete loss of activity in the α-LCAT and β-LCAT are diagnosed with Familial LCAT deficiency (FLD). A 32-year-old female patient was referred to Nephrology for proteinuria with a history of obesity, PCOS, and FLD. In addition to treating her Hyperlipidemia, her risk of progressive kidney disease is high with 2+ proteinuria on UA despite a current EGFR > 100. The median age for ESKD is the mid-40s to early 50s and progression does not entirely correlate with proteinuria. That said, with the advent of multi-modal therapies for proteinuria we hope to reduce her risk of ESKD. She was started on an ARB and will be followed closely for the addition of SGLT2i and/or an MRA. Understanding the genetic basis and clinical implications of LCAT deficiency underscores its importance in clarifying the interplay between genetics, lipid metabolism, and renal function, offering insights into potential therapeutic strategies and improving patient outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5174F: Understanding sleep and metabolic traits in ME/CFS

Authors:

L. Urpa[1], A. Tervi[1], H. Haapaniemi[1], H. Ollila[1,2,3,4]; [1]Inst. for Molecular Med. Finland (FIMM), HiLIFE, Univ. of Helsinki, Helsinki, Finland, [2]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst., Cambridge, MA, [4]Dept. of Anasthesia, Critical Care and Pain Med., Massachusetts Gen. Hosp., Boston, MA

Abstract:

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a poorly-understood disease that affects an estimated 17-24 million individuals worldwide. Often preceded by an infection, the symptoms manifest as debilitating fatigue that in severe cases requires near-constant isolation in a dim, quiet room. Earlier reports show that circulating metabolites and particularly glucose are affected in ME/CFS. In chronic viral infection, hypothesized to be a cause of ME/CFS in at least a subset of patients, high levels of the cytokine interferon gamma leads to isolated insulin resistance in skeletal muscle. This local insulin resistance leads to a compensatory increase in insulin production in the liver, reducing glucose production via gluconeogenesis and resulting in transient peripheral hypoglycemia. To investigate the role of sleep and altered

glucose metabolism in ME/CFS, we obtained genome-wide association study (GWAS) summary statistics for sleep traits from the UK Biobank and fasting glucose, fasting insulin and other glucose metabolism traits from the MAGIC consortium. We additionally obtained body mass index (BMI) GWAS summary statistics from the GIANT consortium. We compared these summary statistics to a genome-wide association study meta-analysis of ME/CFS performed on individuals from the national FinnGen study, the UK Biobank, the Estonian Biobank, and the Mass General Brigham Biobank (N = 3908 cases with ME/CFS and over one million disease free controls). We performed genetic correlation with LD Score Regression (LDSC) between ME/CFS and metabolic traits, and two-sample mendelian randomization with significant single-nucleotide polymorphisms as exposures predicting ME/CFS as an outcome. We found no significant genetic correlation between ME/CFS and glucose metabolism traits or BMI, suggesting that genetic variants associated with typical glucose metabolism were not related to those underlying liability to chronic fatigue syndrome. In contrast, ME/CFS correlated with long sleep duration and chronotype ($P < 0.05$). While glucose metabolism and ME/CFS are likely controlled by separate genetic risk factors, the findings indicate a connection between sleep and ME/CFS. We also found no significant causal relationship between genetic instruments associated with glucose metabolism and ME/CFS, indicating that altered glucose metabolism is likely not causal in the development of chronic fatigue.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5175F: Unlocking Knee Osteoarthritis: Whole-Exome Sequencing Identifies Key Susceptibility Genes

Authors:

A. Biswas, B. Garg, R. Malhotra; All India Inst. of Med. Sci., New Delhi, India

Abstract:

IIntroduction: Knee osteoarthritis (kOA) is a major cause of disability in the elderly with the global prevalence of 16% and 22-39% in India. Major risk factors for the kOA is aging, diet, lifestyle and most importantly genetic modality. Till date, genetic predisposition is not fully understood and also there is lack of genetic studies from India.
Materials and methods: In this study, we have sequenced kOA patients through exome sequencing to decipher the associated genetic variation. 70 kOA patients and 20 healthy subjects were screened clinically and confirmed the diagnosis through radio-graphical intervention. Clinical history, demography & 5mL blood sample were collected. DNA was isolated and subjected to exome sequencing (100X). Through expression studies, these

identified's genes level were compared with healthy subjects to correlate with genetic profile.
Results: We uncovered germline variants in established knee osteoarthritis (kOA) candidate genes like *BTNL2 and HLA-DPB1*, alongside several other genes harboring potentially harmful variants for OA. Employing independent filtering for causal variants associated with OA or related bone disorders using the ClinVar database, we identified previously reported variants in genes *NACA2, TLR10, GLE1, and MATN3,* exhibiting low minor allele frequency compared to 1000 Genomes and South Asian population data. Furthermore, we detected numerous rare and novel mutations, subject to thorough bioinformatic scrutiny for potential kOA association. Intriguingly, whole-exome analysis revealed multiple modifier variants across most samples, with rare variants being individual-specific. Notably, the mRNA expression of these candidate genes showed significant upregulation or downregulation, manifesting distinct gene expression profiles.
Conclusion: In conclusion, our study sheds light on the genetic landscape of knee osteoarthritis (kOA), a prevalent cause of elderly disability globally and particularly in India. Our research suggests that individuals with knee osteoarthritis (kOA) possess harmful rare and novel germline variants in genes involved in various pathways, such as immune responses, inflammation, and cartilage degradation. Expanding genetic testing to include a broader range of OA-predisposing genes may enhance screening accuracy for identifying high-risk osteoarthritis populations.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5176F: Unveiling Mechanistic Overlaps in Opioid and Nicotine Addiction through Integrative Systems Biology: Incorporating Single-Cell RNA-Seq, Multiplex Networks, and Explainable-AI

Authors:

D. Jacobson[1], M. Lane[2], K. Sullivan[1], A. Townsend[2], J. Merlet[2], C. M. Alvarez[2], A. Wells[3], B. Quach[4], C. Willis[4], G. Keele[4], E. Johnson[5], D. Hancock[4], D. Hyatt[1]; [1]Oak Ridge Natl. Lab., Oak Ridge, TN, [2]Univ. of Tennessee, Knoxville, TN, [3]Rhodes Coll., Memphis, TN, [4]RTI Intl., Durham, NC, [5]RTI Intl., Duluth, MN

Abstract:

Opioid use results in approximately 480,000 deaths worldwide with an estimated economic impact of $1.5 trillion in the United States alone. Smoking causes 7 million deaths globally and incurs a $1.4 trillion economic impact. Given the global impact of these substance use disorders, our research aims to uncover significant mechanistic overlaps between these highly comorbid

conditions. To incorporate cell-type specificity into our analysis, we generated 15 single-cell type dorsolateral prefrontal cortex (dlPFC) RNA-seq-based networks using our explainable-AI algorithm iRF-LOOP. These networks included various types of inhibitory interneurons, excitatory principal neurons, astrocytes, microglia, vascular cells, and oligodendrocyte precursor cells, and mature oligodendrocytes. We used these cell-type-specific predictive expression networks (scPENs) to create a multiplex network, integrating them with curated layers of protein-protein interactions, literature mining-based associations, gene co-essentiality, molecular pathway databases, gene phylogenetic associations, and transcription factor-target networks, comprising 51,183 genes and millions of connections. Our MENTOR algorithm, based on random walk with restart (RWR), used network connectivity (topology) to mechanistically cluster 357 genes: 149 genes associated with smoking cessation (SmkCess) GWAS and 245 genes linked to opioid use disorder (OUD) via GWAS and dlPFC differential epigenome, transcriptome, and proteome data. The MENTOR algorithm used either a 25-layer scPEN multiplex network or a 10-layer multiplex lacking single-cell data. The 25-layer multiplex network identified 22 clusters with much greater granularity than the non-scPEN 10-layer multiplex. One cluster encompassed genes related to BDNF, its receptor NTRK2, the entire MAPK signaling pathway, and synaptic plasticity transcriptional regulation circuits. Another cluster included genes involved in neuronal development, activation, cytoskeletal remodeling, and beta-catenin signaling, while a different cluster included genes related to neural and synaptic function, including axon guidance and nicotinic acetylcholine receptors associated with addiction susceptibility. We are systematically exploring the mechanisms represented in each of the 22 clusters. Overall, 86% of MENTOR-derived clusters contained both OUD and SmkCess genes, indicating significant mechanistic overlap. Our findings demonstrate the utility of scPENs in revealing intricate biological narratives, aiding the identification of potential therapeutic targets for multiple substance use disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5177F: Using polygenic risk score to assess the role of common genetic variants in phenotypic variability of inflammatory bowel disease in a rare immunodeficiency cohort

Authors:

S. Srinivasan[1], M. N. Similuk[1], R. Ghosh[1], B. A. Seifert[1], M. Tokita[1], J. Lack[2], W. Cao[2], B. E. Marciano[1], C. Zerbe[1], P. A. Frischmeyer-Guerrerio[1], G. Uzel[1], I. Fuss[1], A. Freeman[1], M. Lionakis[1], L. Notarangelo[1], H. Malech[1], J. R. E. Bergerson[1], S. M. Holland[1], M. A. Walkiewicz[1], J. Yan[1]; [1]Div. of Intramural Res., Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]NIAID

Collaborative Bioinformatics Resource, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD

Abstract:

Background
Inflammatory bowel disease (IBD) is a disorder of the gastrointestinal tract marked by diarrhea, abdominal pain, and rectal bleeding. The etiology of IBD is multifactorial, with contributions from environmental, common genetic, and rare genetic factors such as monogenic inborn errors of immunity (IEI). Here, we aimed to determine common variant contributions to IBD in a cohort of individuals with known or suspected IEI and to explore the relationship between rare and common variants in IBD manifestation.

Methods
Two thousand sixty-seven unrelated probands with immune conditions underwent genome sequencing. The cohort was primarily female (n=1097), adult (median 38; 0-94 years), and of European ancestry (n=1551). Variants with frequency>0.01 were used to calculate IBD polygenic risk scores (PRS) based on prior publications. The PRS was tested for association with IBD, defined as having one or more of the Human Phenotype Ontology terms: *ulcerative colitis*, C*rohn's disease*, *colitis*, *chronic colitis, pancolitis, enterocolitis, necrotizing enterocolitis,* and *inflammation of the large intestine* (n=102). The association was tested using logistic regression controlling for sex, age, and 10 ancestry principal components. To test the specificity of the IBD PRS in the cohort, IBD PRS was assessed for association with a phenotype correlated to IBD (fatigue, n=508) and one not correlated to IBD (atopic dermatitis [AD], n=277). Lastly, to see whether the polygenic threshold for disease differed based on the presence of a rare monogenic variant for IBD, monogenic IBD status was assigned to those with a pathogenic or likely pathogenic variant in a monogenic IBD gene. Welch *t*-test was run to test the difference in mean IBD PRS between those with and without monogenic IBD among IBD cases.

Results
The IBD PRS score was associated with IBD (OR=1.34[1.09,1.64], p=0.004). The IBD PRS score was not associated with fatigue (OR=1.07[0.97,1.19], p=0.19) nor AD (OR=0.96[0.84,1.10], p=0.55). Monogenic IBD diagnoses among cases (n=24) consisted of *CYBB* (n=11); *STAT1* (n=3); *NCF1* (n=2); *BTK*, *CYBA*, *G6PC3*, *IL2RG, LRBA*, *PTEN*, *TGFBR2*, and *XIAP* (n=1 each). There was no significant difference between mean IBD PRS in those with monogenic IBD versus those without in the IBD cases.

Discussion
Within a cohort with rare immune conditions, IBD PRS was associated with IBD. The IBD PRS was specific to the IBD phenotype in the cohort - it was not associated with fatigue or AD. The polygenic threshold for disease did not appear to differ based on presence of a monogenic IBD

variant. These results support a role of common genetic variation in IBD etiology among individuals with known or suspected IEI.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5178F: Utilizing prescription data from electronic health records to better characterize chronic pain patients and uncover novel genetic associations

Authors:

J. Eicher[1], C. Harlow[2], E. UZOCHUKWU[3], H. Fernando[2], N. Bowker[2], J. Liu[4], C. Webber[3], P. Wilson[2], L. Addis[2], J. Davitte[4], Z. Cader[5]; [1]GSK, Cambridge, MA, [2]GSK, Stevenage, United Kingdom, [3]Cardiff Univ., Cardiff, United Kingdom, [4]GSK, Collegeville, PA, [5]Univ. of Oxford, Oxford, United Kingdom

Abstract:

Chronic pain, defined as pain that persists or recurs for 3 months or longer, encompasses a wide-range of disorders with complex etiologies, heterogenous clinical presentations, and a lack of effective and safe therapies. Presently, there is no single biomedical test nor biomarker used in clinical practice to diagnose chronic pain. Rather, diagnosis is made following clinical assessment of the potential underlying conditions combined with a review of patient history. Consequently, genetics-studies of pain are often limited to investigations of individual pain-related diseases, which may miss common biology shared across multiple pain-related diseases. Our study tests a novel method for studying the genetics of chronic pain by leveraging medication data from two large biobanks (UK Biobank and FinnGen) to identify continuous prescription analgesic use (>90 days duration) as a proxy for the commencement of any chronic pain condition. We construct novel phenotypes to describe the chronic pain population including: chronic pain vs. healthy individuals; time until analgesic cessation; strong opioid usage; and, total time on analgesics. After deriving these chronic pain phenotypes, we performed genome-wide association studies (GWASs) in each cohort followed by meta-analysis alongside rare variant gene-based association tests using UK Biobank whole-genome sequencing data. Across the 11 GWAS meta-analyses, we identified 140 associations with chronic pain with 30 loci (representing 76 of the associations) showing association with multiple chronic pain definitions. Of the associations we identified, many had previously described links to pain phenotypes (e.g. *SLC39A8, DCC, TRPM8*), opioid use disorder and drug metabolism—indicative of our use of medication data in phenotype derivation—
(e.g. *ARHGAP15*, *OPRM1*, *CYP2D6*), and newly detected links to chronic pain

(e.g. *ARPP21*, *CNTNAP2*). Using LD-score regression, we found stronger genetic correlations with osteoarthritis and psychiatric traits and weaker but significant correlations with immune traits. Our study demonstrates how clinical and medication data available in large biobank studies can be used to better define and characterize conditions which are difficult to measure with disease-focused diagnosis codes, are heterogenous, or lack specific diagnosis. Taking advantage of these data can lead to the identification of novel underlying biological mechanisms enabling clinically informed therapeutic discovery.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5179F: Ventricular arrhythmia risk is predicted by a polygenic risk score for atrial fibrillation

Authors:

M. Puckelwartz[1], D. H. Acciani[1], T. Monroe[1], L. Pesce[1], S. Kearns[1], L. Castillo[1], G. Webster[2], E. McNally[1]; [1]Northwestern Univ., Chicago, IL, [2]Lurie Children's Hosp., Chicago, IL

Abstract:

Background: Cardiomyopathy is a prominent cause of heart failure (HF), and both cardiomyopathy and HF independently associate with atrial fibrillation (AFib). Approximately half of all patients with HF eventually develop AFib, accelerating stroke risk and all-cause mortality, and Afib carries an increased risk for HF. Ventricular arrhythmias also associate with cardiomyopathy and HF. Because ventricular arrhythmias can be lethal, current guidelines recommend implantable cardioverter-defibrillator (ICD) placement using left ventricular ejection fraction (LVEF) <35% as a risk stratification tool. However, the odds ratio for ventricular fibrillation for patients with an LVEF <35% was only 2.9. Years of clinical experience demonstrate that EF as a standalone risk factor is neither fully sensitive, nor specific. Additional testable variables, like polygenic risk scores (PRS), could aid in arrhythmia risk stratification. There is growing evidence that atrial dysfunction is both a cause and consequence of ventricular disease. AFib is highly heritable with rare monogenic contribution and mostly polygenic effects. Genome-wide association studies (GWAS) have identified >100 loci that contribute to AFib which have been used to develop an AFib polygenic risk score (PRS). AFib PRS have been validated, but it is not known whether the genomic architecture predicting AFib is also associated with risk of ventricular fibrillation. **Approach:** As part of an AHA Arrhythmia and Sudden Death Strategically Funded Research Network, we performed genome sequencing and analysis on a cohort of 1043 individuals across the arrhythmia spectrum including ~500 controls >40 years and without known cardiac disease. Arrhythmia risk in cases ranged from low-risk

individuals with non-ischemic cardiomyopathy without arrhythmia to higher-risk individuals who underwent ICD placement for aborted sudden death and sudden death. **Results:** We examined an AFib PRS with 161 SNPs to define ventricular outcomes in the cohort (PMID: 38374346). Of the 440 participants with disease, 81% had cardiomyopathy, 8% channelopathy and 7% idiopathic arrest. We found the AFib PRS stratified subjects without significant ventricular arrhythmia risk from subjects with known ventricular arrhythmia (p=0.03), indicating that the AFib PRS includes components that are important markers of ventricular dysfunction and a low PRS for AFib may be protective in the setting of ventricular disease. **Conclusions:** The AFib PRS is a useful tool to stratify ventricular arrhythmia risk identifying shared genetics for atrial and ventricular arrhythmias.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5180F: Vitamin D, cognition, and Alzheimer's disease:Observational and two-sample Mendelian randomization studies.

Authors:

Y. Ilboudo[1], S. Yoshiji[2], T. Lu[3], G. Butler-Laporte[4], S. Zhou[4], B. Richards[5]; [1]Jewish Gen. Hosp., Montreal, QC, Canada, [2]McGill Univ./Broad Inst., Montreal, QC, Canada, [3]Univ. of Toronto, Toronto, ON, Canada, [4]Mcgill Univ., Montreal, QC, Canada, [5]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada

Abstract:

Objective: Observational studies have found that vitamin D supplementation is associated with improved cognition. Further, recent Mendelian randomization (MR) studies have shown that higher vitamin D levels, 25(OH)D, may protect against Alzheimer's disease. Thus, it is possible that 25(OH)D may protect against Alzheimer's disease by improving cognition.Methods: We assessed this hypothesis by studying 26,787 older individuals, examining the relationship between 25(OH)D levels and seven cognitive measurements. To mitigate bias from confounding, we performed two-sample Mendelian randomization analyses. We used instruments from three publications: Manousaki et al (2020), Sutherland et al (2022), and the Emerging Risk Factors Collaboration/EPIC-CVD/Vitamin D Studies Collaboration et al (2021).Results: Our observational studies suggested a protective association between 25(OH)D levels and cognitive measures. An increase in the natural log of 25(OH)D by 1 SD was associated with a higher PACC score (BetaPACC score= 0.06, 95%CI = (0.04 - 0.08); P-value=1.8 x 10-10). However, in the MR analyses, the estimated effect of 25(OH)D on cognitive measures was null. Specifically, per 1 SD

increase in genetically estimated natural log of 25(OH)D, the PACC scores remained unchanged in the overall population, (BetaPACC score= -0.01, 95% CI (-0.06 to 0.03); P-value= 0.53), and amongst individuals aged over 60 (BetaPACC score =0.03, 95% CI (-0.028 to 0.08); P-value=0.35).Conclusion: In conclusion, our MR study found no clear evidence to support a protective role of increased 25(OH)D concentrations on cognitive performance in European ancestry individuals. However, our study cannot entirely dismiss the potential beneficial effect on PACC for individuals over the age of 60.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5181F: Waist-to-hip ratio as a sex-specific genetic proxy for visceral adipose tissue: Apples and (pine)apples?

Authors:

R. Smit[1,2], Z. Wang[2], R. J. F. Loos[1,2]; [1]Univ. of Copenhagen, Copenhagen, Denmark, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Excess abdominal visceral fat (VAT) is a well-established risk factor for cardiometabolic disease, but direct measurement using MRI remains prohibitively expensive for widespread implementation. Waist-to-hip ratio (WHR), particularly when adjusted for body mass index (adjBMI), has been suggested as a proxy for understanding the biology of VAT. However, to date, genome-wide association studies (GWAS) of WHR have mostly identified loci in women, despite similar sample sizes for men. As such, the impact of sex on the effectiveness of WHR as a proxy of VAT remains unclear, and direct comparison of GWAS results challenging. Under the assumption that GWAS of WHR capture diluted effects on VAT, we applied PheMED to quantify sex-specific phenotypic misclassification between WHR and VAT, with and without adjustment for BMI. For comparison, we calculated genetic correlation (rg) across traits and sexes using LD score regression. For the analyses we used non-overlapping, publicly available GWAS summary statistics from populations of primarily European ancestry with MRI-derived VAT (UK Biobank, N: men 19k, women 20k) and WHR (GIANT consortium, N: men 93k, women 117k). For PheMED, we used a set of 268k approximately independent common variants. Finally, we explored sex-specific, dilution-aware GWAS meta-analyses combining the two traits. While WHR and VAT each show comparable genetic correlation across sexes (both 0.74), BMI adjustment widens the genetic gap for WHR (0.66) but narrows it substantially for VATadjBMI+height (0.94). Relative to VAT, WHR GWAS effect sizes for men are effectively diluted by 2.6-fold [95%CI: 2.4,3.0] (with rg 0.66) and by 2.8-fold [2.4, 3.2] for women (with rg 0.53). This suggests that

these WHR GWAS effect sizes would need to be multiplied by 2.6 and 2.8 to achieve comparable effect size distributions to VAT. In contrast, between WHRadjBMI and VATadjBMI+height, dilution is larger for men than women (3.1-fold [2.7, 3.5] vs. 2.6-fold [2.3, 3.0], respectively), and genetic correlation lower (rg 0.52 and 0.60, respectively). This greater dilution for men suggests that their effective GWAS sample size of WHRadjBMI will be lower for joint analysis of summary statistics which account for dilution. Dilution-aware meta-analyses adding WHRadjBMI increased the number of independent loci found in VATadjBMI+height GWAS alone from 6 to 16 in men and from 5 to 43 in women. Ongoing analyses include comparison with MTAG and investigating the impact of the additional adjustment of VAT for height. Our results question whether WHRadjBMI captures the same trait and genetic architecture in men and women and its suitability as a (genetic) proxy of VAT in either sex.

Session Title: Complex Traits and Polygenic Disorders Poster Friday Session

Board 5182F: Whole genome sequencing analysis of mitochondrial genetic variation associated with cachexia in COPD

Authors:

B. Lukusa-Sawalena[1], M. Wang[2], A. Rocco[1], K. Lavin[1], J. Chiles[1], H. Tiwari[1], D. L. Demeo[3], E. K. Silverman[4], D. Arking[5], C. Liu[2], M-L. McDonald[1]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Boston Univ., Boston, MA, [3]Harvard Med. Sch., Cambridge, MA, [4]Brigham & Women's Hosp., Boston, MA, [5]Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Abstract:

Chronic Obstructive Pulmonary Disease (COPD) is a progressive lung disease and leading cause of death and disability worldwide. Characterized by chronic airway obstruction, COPD has a complex etiology influenced by genetic risk factors and environmental exposure to smoke. Increased COPD severity is associated with extra-pulmonary manifestations such as cachexia. The mitochondrion, a cellular organelle with its own genome (mtDNA), is responsible for cellular metabolism and redox homeostasis. MtDNA variants differ from those in the nucleus, and are categorized as either homoplasmic or heteroplasmic, based on whether the variant is or is not present in all mtDNA copies in the same sample. The aim of this study was to test the association of homoplasmic mtDNA variants with cachexia in individuals with COPD. We analyzed whole genome sequencing (WGS) data generated from blood samples taken from 1,936 participants with COPD from the Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE) cohort. mtDNA variants were called with Mitochondrial High-Performance Caller (MitoHPC) using WGS data. Homoplasmic variants were classified using a

variant allele frequency threshold of 3%. Specifically for a given homoplasmic site participants with less than 3% alternate allele reads were coded as 0 and those greater than 97% were coded as 1. Whereas mtDNA SNVs in participant samples with alternate allele frequencies between 3% and 97% were classified as heteroplasmic and excluded from analyses. Homoplasmic variants present in 5% or more of participants were included in analyses. Cachexia was defined as weight loss of 5% or more in 12 months or less. Logistic regression models assessing the association of each homoplasmic variant with cachexia as the outcome were adjusted for sex, age, and pack years of smoking. 62 homoplasmic variants were considered common, 5 of the variants were nominally significantly associated with the development of cachexia (p-value < 0.05) and found in regions of the *ND5, ND1,* and *COX3* genes, or the *D-loop*. In particular, the chrM13708G>A variant in *ND5* (OR=1.46, SE=0.17, p= 0.022) is a missense variant. *ND5* encodes NADH-ubiquinone oxidoreductase chain 5, a protein involved in the electron transport chain. Variants in this gene have been linked to Leigh syndrome, mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes, and mitochondrial complex I deficiency. Our planned future analyses will expand our research to include other TOPMed cohorts with participants with COPD. This study provides suggestive evidence for the relationship between mtDNA variation and COPD cachexia.

Session Title: Epigenetics Poster Friday Session

Board 1001F: A Cell-Type-Specific DNA Sequence Model for Understanding the Role of Noncoding Variants in Chromatin Accessibility

Authors:

S. Tiwari[1,2], A. Karbalayghareh[1], K. K. Dey[1], C. S. Leslie[1]; [1]Mem. Sloan Kettering Cancer Ctr., New york, NY, [2]Weill Cornell, New york, NY

Abstract:

Recent efforts have used deep learning models to understand the effects of genetic variation in cis-regulatory elements (CREs) on function across cell types. Current deep learning models either employ a multitasking approach (Enformer: Avsec et al., Nat. Methods, 2021) across cell types, which tends to suffer in accuracy and feature attribution due to feature leakage, or fit individual models for each cell type (ChromBPNet: Pampari et al., GitHub, 2023) that cannot leverage information about the differences between cell types.
Here, we address these limitations through a Large Language Model framework based on HyenaDNA (Nguyen et al., arXiv, 2023) which has demonstrated greater efficiency and

performance over transformer models in many genomics tasks. We created a cell-type-specific model that is trained jointly across cell types. For each 1kb sequence, we append a unique cell type token at the start of the sequence and the model only predicts the CRE accessibility for that cell type. We compared this model to a multitasking HyenaDNA model without cell type tokens, which predicts the accessibility of all cell types simultaneously. We first pretrained using a next token prediction task on 1 million candidate-CRE regions from ENCODE, then finetuned it to predict the accessibility of those regions in 161 known cell lines and tissues. For 152 cell types, our cell type token model showed an improvement in Pearson correlation > 0.03 (max = 0.08) compared to the multitasking model. Our model also achieved competitive results with 161 HyenaDNA models trained independently on each cell type (mean correlation difference of 0.002).
We validated our approach using in-silico mutagenesis to assign attribution scores at a base pair level to 10,000 distal enhancers. Utilizing TF-MoDISco, we found that our model can identify several important transcription factor (TF) motifs such as PU.1 in B-cell related cell lines, which was missed by the multitasking model. When using the model attributions as TF binding scores tested against TF ChIP-seq profiles for PU.1 in 4 ENCODE immune cell lines (K562, GM12878, GM12891, HL-60), our approach showed a higher average precision in every cell type compared to the multitasking model (average 0.18 versus 0.12).
These results demonstrate that new language-based modeling methods hold promise for creating a unified model to learn underlying mechanisms of the genome while avoiding feature leakage and competing with individually trained models. Ongoing efforts revolve around systematic benchmarking of our model against models like ChromBPNet and Enformer using chromatin accessibility QTL datasets and further analyzing attributions of TF motifs.

Session Title: Epigenetics Poster Friday Session

Board 1002F: A universal spike-in normalization strategy for CUT&RUN, CUT&Tag, and ATAC-Seq

Authors:

S. Traynor, B. Townsley, S. Wang, S. Jeon, B. Egan, B. Delatte; Active Motif, Carlsbad, CA

Abstract:

The development of simple and efficient genome-wide approaches has recently democratized the study of epigenetics. For instance, CUT&RUN and CUT&Tag were developed as fast and affordable alternatives to ChIP-seq, whereas ATAC-seq readily identifies regions of open chromatin. These methods are ideal to look for genome-wide alterations in chromatin

structures but are only semi-quantitative and can miss global changes when performed without normalization spike-ins. In fact, many standard computational pipelines fail to detect a reportedly substantial global H3K27me3 decrease in the presence of GSK126, a therapeutic target for lymphoma and potent inhibitor of the EZH2 methyltransferase. To overcome this challenge, we developed a spike-in approach using Drosophila melanogaster nuclei for normalization of ATAC, CUT&RUN and CUT&Tag experiments using human cells. The addition of the spiked-in nuclei allowed for the detection of the expected global changes in open regions of chromatin (ATAC-seq), H3K27me3 (CUT&Tag), R-loops (CUT&Tag), EED and p53 binding (CUT&RUN) upon treatment of K562 cells with GSK126. By contrast, these same changes were all missed when omitting the spike-in nuclei. This streamlined spike-in normalization strategy can be easily applied to any number of cells and on a large variety of conditions.

Session Title: Epigenetics Poster Friday Session

Board 1003F: Allelic imbalance in chromatin accessibility in the *KCNQ1* SNP region that associates with type 2 diabetes

Authors:

Y. Muller[1], A. Nair[2], K. Wiedrich[1], C. Bogardus[1], L. Baier[3]; [1]NIDDK, NIH, Phoenix, AZ, [2]NIDDK/NIH, Phoenix, AZ, [3]NIH, Phoenix, AZ

Abstract:

A prior genome-wide association study in 7701 southwestern Indigenous Americans utilized a population-specific custom Axiom array (Affymetrix) which contained 474,196 SNPs. These SNPs tagged 92% of common variants ($R^2 \geq 0.85$, minor allele frequency ≥0.05) identified in the genome of this population. SNP rs79788804, which tags 38 additional SNPs in intron 15 of *KCNQ1*, achieved genome-wide significance in its association with type 2 diabetes (T2D; OR=1.31, $p=2 \times 10^{-8}$). *KCNQ1* resides in an imprinted region of the genome and a three SNP haplotype (rs79788804, rs2299620 and rs2237892) with maternally derived alleles strongly associated with T2D ($OR_M=2.05$, $p=2 \times 10^{-12}$). The ENCODE and Islet regulome databases mapped a putative strong enhancer to this region with multiple epigenetic marks and transcription factor binding sites. To increase evidence that the region encompassing rs2299620, rs2237896, rs2237897 functions as a regulatory region, we assessed chromatin accessibility during various stages of pancreatic development using an isogenic induced pluripotent stem cell (IPSC) derived pancreatic islet-based model system. Chromatin at 3 stages of development, baseline (Day 0), pancreatic progenitor stage (PP, Day10) and endocrine progenitor stage (EP, Day 13, Day16) was assessed using an Assay for Transposase Accessible Chromatin qPCR (ATAC-qPCR). This assay

showed that chromatin accessibility is increased by 9.7-fold during the endocrine progenitor stage (EP, Day 13) as compared with baseline (iPSC), and cells with the T2D risk allele at rs2299620 SNP region had a 1.6-fold higher ATAC-qPCR signal than the non-risk cells. This allelic imbalance in chromatin accessibility in the *KCNQ1* SNP region, specifically at the endocrine progenitor stage, was replicated in an independent isogenic iPSC-derived pancreatic islet differentiation. Sequence analysis suggests that this region of the chromatin could potentially bind transcription factors including CTCF, NKX2.2, MAFB, NKX6.1, FOXA2, and PDX1. Our finding of a specific developmental stage of chromatin accessibility during islet development extends the findings of others who reported an alteration in an ATAC peak mapping to this *KCNQ1* region in pancreatic islets from individuals with type 2 diabetes. In summary our data supports that a 341 bp region containing SNPs rs2299620, rs2237896 and rs2237897 may confer risk for type 2 diabetes via acting as a regulatory region during specific stages of islet development.

Session Title: Epigenetics Poster Friday Session

Board 1004F: Brain epigenetic regulatory atlas reveals context-dependent trans-regulatory programs in Alzheimer's disease ⭐

Authors:

T-T. Fu[1], J. Mahendran[2], B. Kapalli[2], N. Rao[2], L. Farrer[1], L. Hou[1]; [1]Boston Univ Sch Med, Boston, MA, [2]Boston Univ., Boston, MA

Abstract:

An increasing number of single-nucleus RNA-seq/ATAC-seq studies on Alzheimer's disease (AD) have revealed molecular alterations in AD at the cell-type level. However, how the interplay among genetic and non-genetic factors leads to AD alterations remains elusive. There are three major gaps: (1) the impacts of disease-associated gene variants on cis-regulatory-elements (CREs) activity and gene expression could be context-dependent, and how these impacts are modulated by transcription factors (TFs) in the context of AD is largely uncharted; (2) the CRE-gene linking sheds light on context-dependent cis-regulation and how TFs mediate alternation of such linking in AD is missing; (3) the relationships among dysregulation of various molecular machineries and the genetic drivers in AD are still unknown. In this study, we bridge these gaps by characterizing TF-mediated trans-regulatory programs and their relationship in AD-associated contexts, based on which we delineate cis-dysregulation. We have curated ~ 100 sorted-cell and single-nucleus ATAC-seq datasets covering major and rare brain cell types under various perturbations and disease conditions. We then integrate TF footprint scores based on TOBIAS with protein-protein interaction networks to identify TF-CRE regulatory modules shared

or specific across different contexts, i.e., combinations of cell types and conditions. For instance, our regulatory annotation reveals that FOSL1-represented TF cluster (e.g., FOSL1/2, JUND, JUNB, BACH2) shows consistently up-regulated (P < 1e-114) footprint scores across contexts, such as IFN-beta stimulated microglia and AD astrocytes and oligodendrocytes. Meanwhile, CTCF shows overall down-regulated (P < 1e-111) footprint scores in AD but not in IFN-beta-stimulated microglia, suggesting loss of CTCF binding relates to chronic neuroinflammation in AD rather than acute response. We finally dissect AD-associated variants based on genome-wide association study (GWAS) with trans-regulatory programs in AD. We found TFs from FOSL1 cluster show upregulated footprint scores at introns of AD GWAS genes, such as *SORL1*, *BIN1* and *GRN*. Intriguingly, the same FOSL1 TF cluster also shows upregulated footprint scores in sorted microglia cells with AD risk variant $SORL1^{A528T}$, further connecting genetic drives to the trans-regulatory programs. Overall, our work systematically constructs context-dependent TF-CRE-gene networks across brain cell types under various conditions, providing a comprehensive resource to dissect key regulators for functional AD risk variants, and investigate gene-by-environment interplay during the pathogenesis of brain disorders.

Session Title: Epigenetics Poster Friday Session

Board 1005F: Characterizing the effect of flowcell and basecaller upgrades on 5mC and 5hmC methylation calling in ONT long read sequencing technologies

Authors:

R. Genner[1], S. Akeson[2], M. Meredith[3], K. Billingsley[4], K. Daida[4], P. Alvarez[5], L. Malik[6], C. Blauwendraat[4], M. Jain[2]; [1]Johns Hopkins Univ. / NIH, Bethesda, MD, [2]Northeastern Univ., Boston, MA, [3]Univ. of California, Santa Cruz, Santa Cruz, CA, [4]NIH, Bethesda, MD, [5]NIH, Washington, DC, [6]NIH, Bethesda, MD

Abstract:

5-methylcytosine (5mC) is the most common and most studied DNA methylation mechanism. Variations in 5mC levels have been associated with human disease. Traditionally, 5mC is analyzed using bisulfite conversion followed by short read sequencing techniques. In the past few years, Oxford Nanopore Technologies (ONT) and Pacific Biosciences have released long-read platforms that yield increased sequencing accuracy and enable direct detection of 5mC and other base modifications. They also also allow for additional downstream analyses like haplotype phasing and the detection of structural variants. ONT recently upgraded the flow cell chemistry on their long read sequencing platform from version R9 to R10, which yields

increased throughput and accuracy; however, the effects of this upgrade on methylation calling have yet to be documented.
We previously performed a series of computational benchmarking experiments to compare methylation calling differences between the flow cell kits using a lymphoblastoid cell line (HG002), a frontal cortex brain tissue sample, and blood sample that had been sequenced with both R9 and R10. We found that R10 made more frequent methylation calls at the extrema (0% and 100% methylated) compared to R9, regardless of sample type. Additionally, a pairwise comparison of HG002 cell line methylation calls made by ONT R9 and R10 with other sequencing platforms revealed that ONT R10 and bisulfite sequencing had the highest concordance.
The HG002 methylation analysis has been expanded to include a methylation calling comparison between the Guppy basecaller and its upgraded version, Dorado. We have also conducted a preliminary analysis of 5-hydroxymethylcytosine (5hmC) levels in the Dorado-basecalled cell line since 5hmC calling accuracy was greatly improved by the Dorado upgrade. This work will help facilitate combined methylation analyses between samples that have been sequenced and/or basecalled with different versions of ONT long read sequencing technologies. This work also tests feasibility for measuring 5hmC levels in the HG002 cell line. We anticipate these analyses will enable comparison of multiple DNA modifications (5mC, 5hmC) across large cohorts of sequencing data. This, in turn, could help identify clinically-relevant methylation profiles and hotspots.

Session Title: Epigenetics Poster Friday Session

Board 1006F: Cross species epigenomic analysis to assess long-term molecular responses to environmental chemicals: from conserved mechanisms to the One Health framework ⭐

Authors:

G. del Giudice, A. Serra, L. A. Saarimäki, M. Fratello, J. Morikka, L. Ylä Outinen, S. Peltola, S. Buonocore, N. D' Alessandro, A. Federico, D. Greco; Tampere Univ., Tampere, Finland

Abstract:

In real-world scenarios, organisms are exposed to myriad of chemicals, from drugs to environmental pollutants over prolonged periods of time. Current toxicology and pharmacology approaches focus on acute responses and phenotypic effects, which vary across species and limit safety assessments. However, epidemiological studies highlight the long term impacts of chemical exposures on disease susceptibility and biodiversity. This emphasizes the need for

cross species and long term monitoring, based on the One Health paradigm.
The epigenome captures long term molecular alterations and reliably indicates exposure induced effects. Furthermore, epigenetic mechanisms are usually well conserved in the evolution, representing a powerful dimension to assess cross species effects.
Our study analyses a large collection of molecular profiles of multiple species to environmental stimuli, from chemicals to nanoparticles. We analyse more than 600 exposures to nanoparticulate matter, as well as altered genes associates with over 1600 chemicals across 6 species. We found that a limited number of gene regulatory mechanisms are responsible for most responses to environmental stimuli in all species (from *H. sapiens* to *C. elegans*).
Specifically, by examining nanoparticle responses, we identified binding sites for zinc finger transcription factors C_2H_2, involved in stress responses, chromatin remodeling and immunity. Studying this regulatory mechanism, we highlight the important role of immunomodulation. The immune system mediates interactions between biological systems and environmental cues, underpinning many human diseases. Using epidemiological epigenetic data (sample size 710,302), we aim to link chemical responses to known phenotypes, improving long-term response assessments and promoting sustainable innovation in the chemical and pharmaceutical industries.
Adopting a One Health framework highlights the need to consider the interconnectedness of animals and the environment. Our discovery of a ZNF C_2H_2 model for nanoparticles underscores the importance of understanding conserved epigenetic regulation mechanisms. We aim to develop similar models for chemicals, providing knowledge that integrates social and environmental protection, ensuring planetary sustainability and long term impact.

Session Title: Epigenetics Poster Friday Session

Board 1007F: Differentially Methylated Regions in Neurexin 3 are Associated with Chronic Heavy Ethanol Use in Rhesus Macaques

Authors:

C. Lowe[1], K. Zimmerman[2], L. Wilhelm[1], F. Miller[1], D. Zhu[1], B. Ferguson[3], K. Grant[3], R. Cervera Juanes[4]; [1]Wake Forest Sch. of Med., Winston Salem, NC, [2]Ctr. for Precision Med., Winston-salem, NC, [3]Oregon Hlth.& Sci. Univ., Beaverton, OR, [4]Atrium Hlth.Wake Forest Baptist, Winston-Salem, NC

Abstract:

Chronic ethanol use is associated with differential DNA methylation (DNAm), influencing gene expression and neural function. However, cross-sectional studies do not enable distinguishing

pre-existing epigenetic risk from adaptive changes caused by alcohol use. We conducted genome-wide DNAm analyses of the dorsolateral prefrontal cortex (dlPFC) of male rhesus macaques collected prior to alcohol use (alcohol-naïve state) and following 14 months of chronic ethanol access to identify these two sets of signals. In alcohol-naïve individuals, DNAm profiling revealed pre-existing neural network signatures predicting future ethanol consumption. We identified 3,539 differentially methylated regions (DMRs), mapping to genes involved in synaptic transmission, neuron development, axonogenesis and glutamatergic signaling. One of the differentially methylated gene was neurexin 3 (*NRXN3*), a pre-synaptic gene with a critical role in the establishment, maintenance and plasticity of synapses. *NRXN3* encodes for two main transcripts, alpha and beta, known to have different functions. In the chronic samples, after 14 months of ethanol exposure, we identified 906 DMRs between heavy drinking macaques and controls. These DMRs mapped to genes involved in DNA repair, cell adhesion, synaptic transmission, and G protein-coupled receptor signaling. NRXN3 was also identified as differentially methylated, although the DMR was in a different location to that identified in the alcohol-naïve individuals. Combining studies, there were three DMRs mapping to *NRXN3*. Two intragenic DMRs near the first exon of the beta transcripts, DMR1 and DMR2, were hypomethylated in an alcohol-naïve state and after chronic use in high drinkers; respectively. DMR1 methylation levels correlated with increased intake. A hypermethylated DMR (DMR4) was identified upstream of the alpha transcripts following chronic use. Functional assays revealed DMR1 and DMR2 showed promoter activity ($p < 0.003$ and $p < 0.05$; respectively), with DMR2 also showing significant enhancer activity ($p < 0.001$). Gene expression analyses conducted in rhesus macaques, showed that NRXN3 beta is downregulated in the dlPFC of heavy drinking macaques. Similar findings were observed in mice exposed to intermittent access ethanol intake, where lower levels of Nrxn3 beta transcripts and protein isoform were observed in the prelimbic cortex. These results indicate an effect of alcohol in altering the expression of a specific transcript/isoform of Nrxn3. These findings highlight ethanol's impact on neural networks, particularly in synaptic plasticity and neurotransmitter signaling, contributing to neural adaptations.

Session Title: Epigenetics Poster Friday Session

Board 1008F: Dissecting chromatin regulatory landscape during B cell activation

Authors:

B. Soskic, P. Marchesan, P. Demela, L. Esposito; Human Technopole, Milan, Italy

Abstract:

Impaired B cell activation results in high susceptibility to infection, immune deficiency and autoimmune diseases. Genetic variants linked to immune-mediated diseases show enrichment in enhancers and promoters specifically active in B cell activation. Considering the dynamic nature of B cell activation and the diverse population of B cells, we mapped gene expression regulation and chromatin accessibility using single-cell transcriptomics (scRNA-seq) and single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq) across three time points of B cell activation. We stimulated human naive and memory B cells with anti-IgM, CD40L and a cocktail of cytokines. The stimulation cocktail triggered B cell activation, proliferation and antibody class switching in vitro. As expected, the majority of activated memory B cells proceeded to later cell divisions than naive B cells. Furthermore, memory B cells produced a higher level of antibodies, suggesting a higher effector capacity. The chromatin accessibility and transcriptomic analysis of 64.513 high-quality cells demonstrated that a large number of transcriptional changes are detected in the early time point of activation. Differential genes were highly time point specific, with only 32% of them overlapping between the time points. This suggests dynamic and large transcriptional differences in two similar B cell states. Interestingly, transcription of the secreting isoform of IgM was higher in the day six activated memory cells compared to naïve, suggesting an intrinsic mechanism of fast differentiation of memory cells to antibody-producing cells. Next, we performed GWAS enrichment analysis and showed that variants associated with lupus, type 1 diabetes and asthma were enriched in differentiated plasma cells, while variants associated with ulcerative colitis and Crohn's disease were enriched in marginal zone-like B cells. Finally, to investigate transcription factor activity and gene-regulatory networks driving the differences in the activation responses between naïve and memory B cells we linked peaks of open chromatin to the activation-induced genes. Our results demonstrated cell type-specific chromatin elements affecting differences in gene expression in naive and memory B cells. This functional map of active chromatin regions, coupled with gene expression dynamics, demonstrates heterogeneity of B cell responses furthering our understanding of B cell differentiation and antibody class switching.

Session Title: Epigenetics Poster Friday Session

Board 1009F: DNA methylation and genetic regulation in natural populations of East Asian and mixed Eurasian ancestry

Authors:

S. Cheng[1,2], S. X[2,3,4,5,6], Z. Ning[3], Y. Lu[2], Y. Du[2], X. Yang[7], M. Li[7], D. Maimaitiyiming[8]; [1]Fudan Univ., Shanghai, China, [2]State Key Lab. of Genetic Engineering, Human Phenome Inst., Zhangjiang Fudan Intl. Innovation Ctr., Ctr. for Evolutionary Biology, Sch. of Life Sci., Fudan Univ., Shanghai,

China, [3]Key Lab. of Computational Biology, Shanghai Inst. of Nutrition and Hlth., Univ. of Chinese Academy of Sci., Chinese Academy of Sci., Shanghai, China, [4]Ministry of Ed. Key Lab. of Contemporary Anthropology, Collaborative Innovation Ctr. for Genetics and Dev., Fudan Univ., Shanghai, China, [5]Dept. of Liver Surgery and Transplantation Liver Cancer Inst., Zhongshan Hosp., Fudan Univ., Shanghai, China, [6]Sch. of Life Sci. and Technology, Shanghai Tech Univ., Shanghai, China, [7]Sinotech Genomics Co., Ltd., Shanghai, China, [8]The First Affiliated Hosp. of Xinjiang Med. Univ., No 137 Liyushan Road, Urumqi, China

Abstract:

It has been shown that DNA methylation (DNAm) plays an essential role in complex traits and diseases, and patterns of DNAm can be differentiated among diverse populations. Nevertheless, natural populations of East Asian and mixed Eurasian ancestry are under-represented in DNAm studies. In this study, we systematically analyzed the DNAm of the Han Chinese and Uyghurs with whole-genome variants related to methylation quantitative traits (meQTL). Interestingly, DNAm differentiation is highly correlated with genetic differentiation between populations, and genetic variation explained 20.19% of the variance in the observed DNA methylation differences. The Uyghurs showed a distinct DNAm pattern compared with the Han Chinese, and the two populations were well distinguished with 660 differentially methylated regions enriched in transcriptional regulatory areas compared with genome-wide CpG sites. Accordingly, meQTLs were also identified to explain the genetic basis of DNAm diversity within and between populations. We further identified a number of genes, including *SEPT8*, *WT1*, *RUNX1*, *DOCK1*, and *COLEC11*, that were significantly differentiated between the Han Chinese and Uyghurs due to DNAm variations. Notably, *SEPT8*, a gene that plays an important role in neurological disorders, cell structure, and kidney damage, showed the highest differentiation of DNAm levels and meQTL between the two populations, possibly due to admixture-induced effects. This study enhanced our understanding of population differentiation in DNAm level and the associated genetic basis in ethnic groups of East Asian ancestry and mixed Eurasian ancestry.

Session Title: Epigenetics Poster Friday Session

Board 1010F: Epigenetic signatures of age-related changes in total gene expression and splicing in human skin

Authors:

C. Shore[1], D. Wang[1], D. Gunn[2], K. Small[3], J. Bell[1]; [1]King's Coll. London, London, United Kingdom, [2]Unilever, London, United Kingdom, [3]King s Coll. London, London, United Kingdom

Abstract:

Exploring age-related changes in gene expression and its regulation give insight into biological mechanisms of ageing and age-related traits and diseases. This is particularly true in skin, the largest human organ, which is subject to a plethora of age-related phenotypic changes which can impact health and well-being.
We explored the impact of ageing on gene expression levels in whole-skin using RNA-sequencing data from up to 664 healthy older female twins (mean:59, range:38-84) from TwinsUK. We observed that 9,745 genes had significant age-related changes in total gene expression in skin at a threshold of q-value<0.05. We also investigated changes in seven different types of splicing variation in the same dataset: Alternative 3' splice site, alternative 5' splice site, alternative first exon, alternative last exon, mutually exclusive exons, intron retention, and skipped exons. Overall, 761 individual splicing events were associated with age after multiple testing correction (q-value<0.05). Of these, most events (N=457) were classified as intron retention, a form of mis-splicing, which increased with age in 342 genes, and decreased with age in only 38 genes, suggesting age-related reduction in splicing fidelity in skin. In total, we observed 327 genes with both age-related changes in total expression and splicing variation in skin.
Age-associated changes in skin DNA methylation were also explored in a subset of 414 female twins. We identified 7,877 age-related methylation signals (q-value<0.05), including in 24 genes which also exhibit altered expression and splicing with age. For example, COL1A1, which encodes pro-alpha1 chains of type I collagen, showed age-related changes across all analyses (decreased gene body methylation, decreased expression, and increased intron retention). We are currently characterising expression Quantitative Trait Methylation (eQTM) and spliceQTM loci to identify putative epigenetic regulators underlying age-related changes in gene expression in skin.
Our work explores the regulatory networks underpinning changes in the ageing human skin transcriptome, including epigenetic regulation of these changes, to provide insight into the molecular mechanisms underlying ageing and age-related skin phenotypes.

Session Title: Epigenetics Poster Friday Session

Board 1011F: Epigenetics for genome-wide target discovery of cardiometabolic diseases

Authors:

M. Gonzalez-Ramirez, G. Atla, L. Marsh, C. Vangjeli, J. Howson; Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom

Abstract:

GWAS have been useful for drug target discovery for complex diseases. However, more than 90% of the disease-associated variants are non-coding, complicating the identification of the causal genes. Quantitative Trait Loci (QTL) are useful to nominate causal genes from non-coding loci but do not resolve all disease-associated loci. We also lack QTLs from diverse populations across tissues which makes it challenging to identify causal genes from GWAS of diverse ancestries. Even for loci where the causal genes have been identified, the knowledge of the molecular mechanism by which the variants lead to disease, as well as the cellular context, are often lacking.

Here, we present our approach to enhance genetics-led drug target discovery with epigenetics data. We first use stratified LD score regression (s-LDSC) and chromatin accessibility data from publicly available single nucleus ATAC-seq (snATAC-seq) from human tissues to determine the cell types enriched with disease heritability. We then obtain a set of high-confidence cell type-specific regulatory element-gene pairs by combining snATAC-seq with predicted enhancer-gene links, which we benchmarked against expression QTLs. We then use statistical finemapping to identify candidate causal variants of cardiometabolic traits, and overlap finemapped variants with high-confidence regulatory element-gene pairs of relevant effector cell types to nominate candidate causal genes.

As a proof of concept, we applied this strategy to Type 2 Diabetes (T2D) and glycaemic traits. Beta cells, hepatocytes, adipocytes and myocytes are among the most enriched cell types with heritability of T2D and glycaemic traits. Our regulatory element-gene pairs for these relevant cell types have a precision of around 60% and perform better than assigning closest gene. Our approach to identify candidate causal genes is validated by the fact that we can confirm 40 out of 132 curated T2D effector gene predictions from the Common Metabolic Diseases Knowledge Portal. Among these known genes, we highlight *PPARG* in adipocytes, which plays a key role in adipocyte biology and glucose metabolism. Compared to genes nominated through QTLs, rare and common coding variants, we increased the number of candidate causal genes by 12%. Among these novel T2D causal genes we highlight *SRRM3*, a splicing factor previously implicated in T2D risk that regulates a beta-cell splicing program.

We envisage that our approach will aid in the identification of causal genes from non-coding loci that lack QTLs, will assist in resolving causal variants and genes in complex regions and will inform on effector cell type when genes are nominated by QTLs from bulk tissues.

Session Title: Epigenetics Poster Friday Session

Board 1012F: ERG impacts the shear stress response in human aortic endothelial cells

Authors:

K. Vazquez, B. Tran, M. Whalen, L. Stolze, C. Romanoski; Univ. of Arizona, Tucson, AZ

Abstract:

The ETS-family transcription factor ERG is required for endothelial lineage specification by driving the expression of lineage-specific genes. ERG's control of gene expression occurs through context-dependent hierarchical and collaborative interactions with other transcription factors to activate specific enhancers. Endothelial cells are constantly exposed to shear stress, a mechanical force produced by the friction blood flow exerts on the walls of blood vessels. Two primary patterns of shear stress exist in the vasculature known as laminar and oscillating shear stress. However, the mechanism with which ERG affects the shear stress response in endothelial cells through the collaborative binding with differentially regulated transcription factors is unknown. To address this, we utilized a flow device to mimic laminar and oscillating shear stress in wild-type and ERG knockout telomerized human aortic endothelial cells *in vitro* followed by RNA-seq. We performed differential gene expression between wild-type and ERG knockout cells subjected to laminar or oscillating shear stress and identified 7,568 differentially expressed genes with an FDR ≤ 0.05 across all conditions. Following this, we compared the sequences from 1 kb upstream to 1 kb downstream, to include regulatory elements, of each differentially regulated gene with cardiovascular disease GWAS and found these genes overlap with 25,585 GWAS SNPs. We then utilized WGCNA to create a co-expression network by grouping all transcripts with similar expression patterns across perturbations into modules. We found seven modules correlated with only shear stress, nine modules correlated with only ERG expression, and nine correlated with both. In one module, gene expression was increased under laminar shear stress relative to oscillating shear stress but upon ERG knockout the effect of shear stress on the expression of those genes was suppressed. Gene ontology analysis of this module revealed endothelial cell differentiation as an enriched pathway which included *NR2F2*, a transcription factor that interacts with ERG to promote endothelial identity. These results indicate ERG is involved in regulating the shear stress response by increasing endothelial-lineage gene expression under laminar shear stress. To complement these results, we are evaluating changes in ERG binding using CUT&Tag to identify novel transcription factors that interact with ERG under different shear stress patterns. The integration of these results will provide a better understanding of how collaborative binding of ERG with various transcription factors directs the shear stress response in endothelial cells.

Session Title: Epigenetics Poster Friday Session

Board 1013F: Functional annotation of risk loci for genetically complex disease using epigenomic liquid biopsy

Authors:

Z. Zhang, B. Sylvan, S. Sotudian, K. Semaan, Z. Zhang; Dana-Farber Cancer Inst., Boston, MA

Abstract:

Background: Epigenomic profiling datasets have been utilized in molecular quantitative trait loci (QTL) mapping and allelic imbalance (AI) analysis to functionally annotate genomic risk loci for traits and diseases identified by genome-wide association studies (GWAS). However, using these methods to understand disease heritability is limited by the lack of large epigenomic reference panels from relevant tissues. Performing epigenomic profiling of circulating cell-free nucleosomes using liquid biopsy is highly scalable and measures epigenetic information from a diverse composition of tissues and cells of origin.
Methods: We used epigenomic profiling datasets generated by an immunoprecipitation-based approach targeting histone modifications and DNA methylation on circulating cell-free chromatin from a total of 1,033 liquid biopsy samples across 15 cancer types. By incorporating publicly available GWAS summary statistics of 14 cancer types and 10 complex traits, we performed cistrome-wide association studies (CWAS) to identify germline genetic variants that contribute to disease by affecting the activity of nearby regulatory elements.
Results: We identified 4,757, 7,357, and 206,447 candidate regulatory elements (CREs) with significant AI or chromatin QTLs (cQTLs) for H3K27ac, H3K4me3, and DNA methylation, respectively. CREs with AI or cQTLs showed significant enrichment for risk heritability of multiple cancer types, including prostate cancer, colorectal cancer, and renal cell carcinoma (RCC). In prostate cancer, CWAS identified significant CRE-trait associations at 79 of 98 GWAS risk loci using predictive models of DNA methylation and 35 using models of H3K4me3. The integration of CWAS and liquid biopsy epigenomic profiling also identified CREs associated with risk of non-oncologic diseases and complex traits. For instance, the CWAS of rheumatoid arthritis implicated variation in DNA methylation at CREs within 11 of 21 GWAS risk loci for this disease.
Conclusions: Integrating liquid biopsy epigenomic profiling assays with cQTLs mapping, AI analysis, and CWAS has the potential to identify genetic variants that confer disease risks through their effects on transcriptional regulation. This novel paradigm may allow rapid scaling of sample sizes for epigenomic reference panels, increasing the power of CWAS to reveal the genetic architecture of complex diseases.

Session Title: Epigenetics Poster Friday Session

Board 1014F: Functional interactions between LINE-1 retrotransposon signaling and the innate immune system in lung cancer

Authors:

B. Hoden[1], A. Chamorro[1], M. Tavera-Garcia[1], A. Hassanin[2], D. Zhang[1], K. Ramos[2]; [1]Texas A&M Hlth.Sci. Ctr. Inst. of BioSci.s and Technology, Houston, TX, [2]Texas A&M Inst. of BioSci.s and Technology, Houston, TX

Abstract:

Lung cancer remains the leading cause of cancer-related deaths worldwide. Recent studies in our laboratory and others have implicated Long Interspersed Element-1 (LINE-1) retrotransposons as mediators of genomic instability in lung cancer. LINE-1 retrotransposons are mobile genetic elements that copy and paste themselves into new genomic locations to alter chromatin architecture and activate oncogenic signaling. Negative correlations between LINE-1 and immune activity have been reported for many cancer types, including lung cancer. However, specific mechanisms have not been established and our current understanding remains incomplete. The heterogeneous nature of lung cancer poses challenges for effective immunotherapy, necessitating a deeper understanding of tumor subpopulations and their interactions with the immune system. As key players in innate immunity, Toll-like receptors (TLRs) recognize various pathogen-associated molecular patterns (PAMPs) and danger-associated molecular patterns (DAMPs). We posit that understanding the intricate interplay between TLRs and LINE-1 retrotransposons holds promise for unraveling novel biomarkers and therapeutic targets in lung cancer. In studies to examine functional interactions between LINE-1 and the innate immune system we found that various TLRs induced LINE-1 as well as genes within its regulatory network including, ICAM-1, AGL, PKIA, RBM39, RGS3, RGS13, VCAM1, VAMP3, and TGFB1. Gene inductions were seen in whole cell and in exosomes, with TLR5 activation by bacterial flagellin showing the greatest induction of LINE-1 followed by TLR3 by Poly(I:C) LMW, TLR2 by HKLM, and TLR4 by LPS. Given our findings and the established role of exosomes in regulating cell-cell communication through the transfer of proteins, lipids, and RNA across the tumor microenvironment, TLRs like TLR5 may regulate the cargo of exosomes to modulate LINE-1 signaling in lung cancer. The study of functional interactions between LINE-1 retrotransposons and TLR signaling may open the door for the development of novel strategies to combat inflammation, genomic instability, and immune evasion in lung cancer.

Session Title: Epigenetics Poster Friday Session

Board 1015F: Genetic and epigenomic cross talk between subcutaneous adipose tissue cell-types contributes to abdominal obesity risk ⭐

Authors:

Z. Chen[1,2], S. Das[3], A. Kar[3,4], S. T. Lee[3], K. Abuhanna[3], M. Alvarez[3], M. G. Sukhatme[3], K. Gelev[3], M. G. Heffel[3], Y. Zhang[3], O. Avram[1,2,5], E. Rahmani[2], S. Sankararaman[1,2,3], S. Heinonen[6], H. Peltoniemi[7], E. Halperin[2], K. H. Pietiläinen[6,8], C. Luo[3,4], P. Pajukanta[3,4,9]; [1]Dept. of Computer Sci., Univ. of California, Los Angeles, CA, [2]Dept. of Computational Med., Univ. of California, Los Angeles, CA, [3]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [4]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [5]Dept. of Anesthesiology and Perioperative Med., Univ. of California, Los Angeles, Los Angeles, CA, [6]Obesity Res. Unit, Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [7]Eira Hosp., Helsinki, Finland, [8]Hlth.yWeightHub, Endocrinology, Abdominal Ctr., Helsinki Univ. Central Hosp. and Univ. of Helsinki, Helsinki, Finland, [9]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract:

Abdominal obesity (AOb) is a well-established risk factor for common cardiometabolic diseases. Previous GWAS studies have implicated genes highly expressed in subcutaneous adipose tissue (SAT) for AOb. However, cell-type level epigenomic landscape of SAT and its interplay with AOb risk variants have remained elusive. To address this knowledge gap, we obtained cell-type level DNA methylation and chromatin conformation profiles by conducting single nucleus methyl-3C sequencing (sn-m3C-seq) on nuclei isolated from SAT biopsies from five females with obesity undergoing liposuction. In addition, we generated matching single cell level expression profiles by applying single nucleus RNA-seq on the SAT nuclei from the same cohort. After careful quality control, we obtained seven main cell-types: adipocytes, adipose stem and progenitor cells (ASPCs), endothelial, perivascular, myeloid, mast, and lymphoid cells when utilizing both the methylation and chromatin conformation profiles, also seen using each modality separately. We further leveraged the gene-body methylation patterns to co-embed all sn-m3C-seq nuclei with those generated by snRNA-seq, which showed high concordance between de-novo and RNA-derived annotation (ARI=0.975). We next investigated the global SAT regulatory dynamics by identifying 2,233,540 differentially methylated mCG regions (DMRs) with methylpy. Strikingly, 57.5% and 58.2% of the DMRs are hypo-methylated in adipocytes and ASPCs, contrasting with only 14.2% in myeloid cells, while conversely, more hyper-methylated DMRs were observed in myeloid cells (77.6%) over adipocytes (16.7%) and ASPCs (22.6%) (-log10p>300). At the chromatin level, we found that 44% of the genome is differentially conformed between the SAT cell types, with an average of 54% A (active) and 46% B (inactive) compartments across cell-

types. To investigate whether the DMRs or cell-type level A compartments contain AOb risk, we constructed partitioned AOb PRSs for variants in DMRs and cell-type level A compartments in the UK Biobank using plink. We discovered 7 of 10 constructed PRSs for variants in DMRs (hypomethylated regions in adipocytes, ASPCs, endothelial, and perivascular cells, and hypermethylated regions in endothelial, myeloid, and perivascular cells) to be significantly enriched AOb predictors ($p_{R2}<0.05$; $p_{perm}10000<0.05$), as were all 5 tested cell-type level A compartment PRSs ($p_{R2}<0.05$; $p_{perm}10000<0.05$), which performed similarly to the genome-wide AOb PRS. Taken together, we discover epigenomic cross talk between SAT cell-types and show that variants in SAT cell-type level epigenomic sites significantly contribute to AOb risk.

Session Title: Epigenetics Poster Friday Session

Board 1016F: Genotype-epigenotype-phenotype correlations in Cornelia de Lange Syndrome: Insights from a case with a novel intragenic deletion in the *NIPBL* gene

Authors:

Y. Murofushi[1], Y. Nishio[1,2,3], T. Kawai[4], M. Inaba[5], S. Mizuno[5], M. Nakajima[6], A. Ishikawa[7], Y. Muramatsu[3], H. Muramatsu[3], Y. Nakazawa[2], Y. Oka[2], T. Yamada[8], Y. Takahashi[3], T. Ogi[2], S. Saitoh[1]; [1]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Dept. of Genetics, Res. Inst. of Environmental Med., Nagoya Univ., Nagoya, Japan, [3]Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [4]Dept. of Maternal-Fetal Biology, Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [5]Dept. of Clinical Genetics, Central Hosp., Aichi Dev.al Disability Ctr., Kasugai, Japan, [6]Dept. of Pediatrics, Hakodate Central Gen. Hosp., Hakodate, Japan, [7]Dept. of Med. Genetics, Sapporo Med. Univ., Sapporo, Japan, [8]Div. of Clinical Genetics, Hokkaido Univ. Hosp., Sapporo, Japan

Abstract:

Background: *NIPBL*, essential for loading the cohesin complex onto sister chromatids, is a major causative gene in Cornelia de Lange Syndrome (CdLS). Patients with *NIPBL* truncating mutations (CdLS-trunc*NIPBL*) typically exhibit more severe phenotypes than those with *NIPBL* missense mutations, as well as mutations in other causative genes (CdLS-non*NIPBL*), such as *SMC1A*, *RAD21*, *SMC3*, *BRD4*, and *HDAC8*. Less severe phenotypes are also observed in the small number of patients with *NIPBL* intragenic deletions. Interestingly, beyond genotype-phenotype correlations, distinct episignatures among phenotypically heterogeneous patients with different genetic etiologies have been reported. The aim of our study is to clarify the possible genotype-epigenotype-phenotype correlation within CdLS-*NIPBL*.Methods: Phenotypic

and genotypic data were collected from CdLS patients with confirmed genetic diagnoses through exome or genome sequencing. Methylation profiles were obtained for patients with available genomic DNA extracted from peripheral blood.Results: Clinical data from 19 patients were analyzed: 12 with *NIPBL* variants, 2 with *SMC1A*, 2 with *RAD21*, 2 with *HDAC8*, and 1 with *SMC3* variants. Detailed variant information was available in 10 *NIPBL* variants: 7 truncating mutations, a missense mutation, a splice site mutation resulting in exon 2 skipping, and an intragenic deletion ranging 1.6kb including 3' part of exon 10. RT-PCR analysis confirmed that the intragenic deletion resulted in two aberrant transcripts: a complete (dominant) transcript and a partially skipping of exon 10 (Δex10), which escaped nonsense-mediated mRNA decay. Western blot analysis showed quantitatively intact mutant protein. Phenotypically, the patient with Δex10 (CdLS-*NIPBL*Δex10) exhibited typical CdLS features, albeit with relatively mild growth retardation and upper-limb malformation, showing moderate severity compared to CdLS-trunc*NIPBL* cases. DNA methylation analysis for 11 patients (6 with *NIPBL* mutations: 4 truncating, a missense, and an intragenic deletion; 2 with SMC1A; 2 with RAD21; 1 with HDAC8) revealed two distinct clusters: one for more severe CdLS with *NIPBL* point mutations and another for CdLS-*NIPBL*Δex10 and CdLS-non*NIPBL*.Discussion: CdLS-*NIPBL*Δex10 exhibited an intermediate phenotype between CdLS-non*NIPBL* and CdLS-trunc*NIPBL*. The episignatures, distinct from the more severe CdLS-*NIPBL* and resembling the milder CdLS-non*NIPBL,* suggest a genotype-epigenotype-phenotype correlation. These findings offer crucial insights into the epigenetic mechanisms underlying CdLS pathophysiology from cohesin dysfunction.

Session Title: Epigenetics Poster Friday Session

Board 1017F: High-resolution CNS-specific spatial architectures of chromosomes in Drosophila reveal higher order organization patterns of gene regulation.

Authors:

X. Yu; Michigan State Univ., East Lansing, MI

Abstract:

The spatial conformation of chromosomes in 3D space plays pivotal roles in coordinating gene expression and epigenetics in the central nervous system (CNS). However, due to the extremely high rates of missing chromatin contacts in current experimental Hi-C datasets, advanced computational models are desired to accurately reconstruct the CNS-specific 3D configurations of genomes at high-resolution. Using our recently developed algorithm, FLAMINGO, we successfully reconstructed the highest resolution (5kb-resolution) 3D architectures of the whole

Drosophila genome, revealing detailed and specific chromatin interactions. The accuracy of the predicted spatial conformations is systematically benchmarked, and complete chromatin contact maps are generated to impute missing contacts. Furthermore, the high-resolution predictions enable the discovery of ultra long-range chromatin loops, multi-scale TADs, and novel boundary elements. The reconstructed chromatin configurations provide the structural foundation to decode the spatial cooperation among distal regulatory elements, genes, and epigenetic marks in CNS, and will facilitate mechanistic insights into genome folding and molecular dysregulation in neurodegenerative diseases.

Session Title: Epigenetics Poster Friday Session

Board 1018F: Hyperosmotic stress induces complete rewiring of chromatin interactions ⭐

Authors:

J. Flores[1], A. Perreault[2], D. Cruz Alonso[1], Y. Wu[1], Z. Drum[1], Y. Quiroga-Barber[1], I. Sahasrabudhe[1], J. Demmerle[3], D. Cai[3], D. Phanstiel[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Elon Univ., Elon, NC, [3]Johns Hopkins Univ., Baltimore, MD

Abstract:

3D chromatin structure plays a critical role in gene regulation, cell fate decisions, and the response to external stimuli by connecting regulatory loci, such as enhancers, with gene promoters. While most 3D chromatin interactions are thought to be driven by the active extrusion of DNA loops via CTCF and the cohesin complex, recent work has suggested other mechanisms drive loop formation, including liquid-liquid phase separation. Cellular response to hyperosmotic stress has recently been shown to involve phase separation of specific transcription factors and the rapid clustering of open chromatin regions, which suggest large-scale changes in 3D chromatin structure. To map and understand these changes we conducted Hi-C, CUT&RUN, and RNA-seq in HEK293 cells responding to hyperosmotic stress (200 mM sorbitol). We observed a near complete rewiring of 3D chromatin contacts. All existing loops were lost or weakened. In addition, we identified 163 newly formed loops that appeared to be driven by CTCF/cohesin-independent mechanisms. Compared to the loops in resting cells, the anchors of these sorbitol-induced loops were less likely to be bound by CTCF, exhibited a more punctate pattern in Hi-C maps, and were associated with weaker contact domains. Interestingly, these loop anchors were enriched for overlap with gene promoters but were not associated with changes in gene expression. Further work is underway to determine the exact mechanisms and transcriptional impacts of these changes.

Session Title: Epigenetics Poster Friday Session

Board 1019F: Identification of the *hsa-circ-0018905* as new biomarker for Multiple Sclerosis

Authors:

V. Lodde[1], I. Zarbo[1], I. Campesi[1], M. Gorospe[2], M. Floris[1], F. Cucca[1], M. Idda[1]; [1]Univ. of Sassari, Sassari, Italy, [2]NIH NIA, Baltimore, MD

Abstract:

Introduction: Multiple sclerosis (MS) is a demyelinating autoimmune disease characterized by early onset and for which the interaction of genetic and environmental factors is crucial. Dysregulation of the immune system as well as myelinization-de-myelinization has been shown to correlate with changes in RNA including non-coding RNAs. Recently, a new class of non-coding RNAs, circular RNAs, has emerged as a key player in the complex network of gene dysregulation associated to MS. Despite several efforts, the mechanisms driving circRNAs regulation and dysregulation in MS still need to be properly elucidate.**Method**s: Here, we explore the panorama of circRNA expression in PBMCs purified from 5 newly diagnosed MS patients and 5 healthy controls (HC) using the Arraystar Human Circular RNA Array. Experimental validation was then carried out in a validation cohort and a possible correlation with disease severity was tested.**Results**: We globally identified 64 differentially expressed circRNAs 53 of which are downregulated in PBMCs purified from MS as compared to HC. The discovery set data has been subsequently validated using qRT-PCR with an independent cohort of 20 RRMS patients and 20 HCs. We validated nine circRNAs showing the lowest and highest levels of differential expression in the RRMS remission group versus the HC group. *hsa-circ-0000518*, *hsa-circ-0000517*, *hsa-circ-0000514*, and *hsa-circ-0000511* were significantly upregulated in the MS group. Moreover, *hsa-circ-0018904*, *hsa-circ-0048764*, and *hsa-circ-0003445* were significantly downregulated. The expression level of *hsa-circ-0018904 was significantly decreased* in patients showing a higher level of disability and in progressive form of MS.**Conclusion**: We described the circRNAs expression profile of PBMCs in newly diagnosed MS patients and propose *hsa-circ-0018905* as potential MS biomarker.

Session Title: Epigenetics Poster Friday Session

Board 1020F: Identifying Biomarkers Associated with Chronic Pain in Hidradenitis Suppurativa

Authors:

S. Shah[1], A. Vedangi[2], U. Radhakrishna[3]; [1]BJ Med. Coll. Inst. of Med. Post-Graduate Studies and Res., Ahmedabad, India, [2]2. Dept. of Clinical Res., KIMS ICON Hosp., A unit of ICON Krishi Inst. Med. Sci., Visakpatnam, India, [3]Green Cross Pathology & Molecular Lab., Ahmedabad, India

Abstract:

Background: Hidradenitis suppurativa (HS) is a persistent skin ailment that predominantly targets regions rich in hair follicles and apocrine sweat glands, notably encompassing the underarm, groin, buttocks, and lower breast areas. Until 85 pathogenic variations in the gamma-secretase gene are identified, documenting genetic heterogeneity. Intense pain and discomfort in HS have been commonly noted, primarily due to the lesions' effects on nearby tissues. Pain is one of several factors that can influence alterations in DNA methylation patterns. However, its precise role in HS is not fully understood. We aim to identify molecular markers for chronic pain in HS patients. **Methods:** We performed DNA methylome of peripheral blood DNA derived from a group of 24 patients with HS, and 24 healthy controls, using Illumina Methylation array chips. **Results:** We identified 173 significantly differentially methylated CpG sites across 173 distinct genes implicated in pain sensitization in HS. Among these, 151 sites were hypomethylated, while 22 were hypermethylated. Several of the identified genes are well-known for their crucial roles in inflammation (e.g., *ITGB2, PRMT2, RIPK4, BACE2, ADAMTS5, MX1, IFNAR2, NOS3, AHR, HLA-DRB1, NR3C1, CFTR, PTGER3, DLG2, PTGER1, LTA, SMAD3, PRDM16, MMP3, IL4, IKBKAP, CSNK1D, TGFB1, GNAQ, CAMK2A, PRX, PLCB3, ESR2, GABRB3*). Additionally, others are associated with cytokine functions (e.g., *AHR, HLA-DRB1, NR3C1, CFTR, LTA, SMAD3, IL4, TGFB1, GNAQ, STAT6, POMC, MMP2, CX3CR1, DPP4, FGF2, NRG1, DNMT1, BDNF, IL18R1, ABCA1, TNFRSF1B, TGFBR2, PPARA, IL1R2, RUNX1, ACE, MALAT1*), ion channels (e.g., *CLIC1, CACNA1C, SCN1A, KCNQ5, KCNMA1*), oxidative stress (e.g., *SCN8A, DNMT1, CSNK1D), and obesity (e.g., DRD2, FAAH, LMX1B, MYT1L, TGFB1*). The results of gene ontology and pathway enrichment analysis revealed the involvement of these genes in 43 pathways (FDR P-values ≤0.05). Notable pathways include the Calcium signaling pathway, Cocaine addiction, Nicotine addiction, Oxytocin signaling pathway, Gap junction, cAMP signaling pathway, Glutamatergic synapse, and Lipid and atherosclerosis. **Conclusion:** Our study uncovered key insights into how methylation variation in specific genes impacts pain susceptibility and perception in HS, suggesting novel potential therapeutic targets for pain sensitization. Understanding the epigenetic regulation of pain-related genes and functional pathways is crucial for developing targeted interventions for pain management.

Session Title: Epigenetics Poster Friday Session

Board 1021F: Inferring genome organisation and gene regulation from 6-base sequencing data

Authors:

P. Creed[1], R. Crawford[1], M. S. Hill[1], E. Lleshi[1], R. Sethi[1], W. Stark[1], J. Teyssandier[1], M. Wilson[2]; [1]biomodal, Cambridge, United Kingdom, [2]Princeton Univ., Princeton, NJ

Abstract:

A key challenge in genomics is the generation and integration of data across different modalities. This typically requires several assays to be performed, which is costly and time-consuming, and the subsequent integration of these data across assays is technically challenging. Here, we leverage an assay that enables sequencing the complete genetic sequence and the DNA modifications, 5-methylocytosine (5mC) and 5-hydroxy-methylocytosine (5hmC), from low nanogram amounts of DNA, to provide 6-base genomic data.
Given 5mC and 5hmC play key roles in gene regulation and chromatin organisation, we aimed to explore how these multimodal data could further elucidate key biological processes and yield novel insight. To this end, we trained and evaluated a series of machine-learning models to predict gene expression, chromatin accessibility, and enhancer state from 6-base sequence data.
We use 6-base data from a mouse embryonic stem cell-line, ES-E14TG2A, alongside publicly available polyA RNA-seq, ATAC-seq, TT-seq, and histone modification data of the same cell line as training data and evaluate the performance of the models on held-out test chromosomes.
We show that these models can generate highly accurate predictions of gene expression (polyA RNA-seq prediction: R2=0.75, Spearman's ρ;=0.86; TT-seq prediction: R2=0.85, Spearman's ρ; =0.91) and chromatin accessibility (ATAC-seq prediction: R2=0.83, Spearman's ρ;=0.93).
Importantly, we found stronger performance of our models for predicting TT-seq over polyA RNA-seq signal, suggesting that 6-base data offers a powerful window into nascent transcriptional activity. As well as being able to predict continuous expression and chromatin accessibility metrics, we show 6-base data predicts enhancer state (Active, Primed, Repressed), defined by histone modifications, with 91% accuracy. In all models the addition of resolved 5hmC signal over undifferentiated 5mc and 5hmC improved predictive performance.
We have shown that the combination of resolved methylation and genomic data combined with machine-learning can generate accurate inference of other data modalities which play key roles in gene regulation. Thus, there is a compounding effect whereby 6-base genomic assays not only yield direct data, but also the foundations for multiple other inferred modalities. Looking

ahead, these approaches can enable novel insights into core biological processes and accelerate the speed of iteration for experimental projects, where one can yield multifaceted insights and even conduct pilot experiments in silico using predictive models.

Session Title: Epigenetics Poster Friday Session

Board 1022F: Integrative analysis of omics summary data reveals putative causal involvement of cardiac DNA methylation in dilated cardiomyopathy

Authors:

K. Tan[1], D. Tay[1], W. Tan[2,3,4], H. Ng[1], E. Wong[4], M. P. Morley[5], G. K. Singhera[6,7], C. Lee[8], P. R. Jain[1], F. Tai[1], P. J. Hanson[7,9], T. P. Cappola[5], K. B. Margulies[5,10,11,12], R. Foo[2,13,14,15,16], M. Loh[1,4,17,18]; [1]Lee Kong Chian Sch. of Med., Nanyang Technological Univ., Singapore, Singapore, [2]Cardiovascular Res. Inst., Natl. Univ. Hlth.System, Singapore, Singapore, [3]Cardiovascular Disease Translational Res. Programme, Natl. Univ. of Singapore, Singapore, Singapore, [4]Genome Inst. of Singapore, Agency for Sci., Technology and Res., Singapore, Singapore, [5]Penn Cardiovascular Inst., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [6]Dept. of Med., Univ. of British Columbia, Vancouver, BC, Canada, [7]Bruce McManus Cardiovascular Biobank, UBC- Ctr. for Heart Lung Innovation, Vancouver, BC, Canada, [8]Cardiovascular-Metabolic Disease Translational Res. Programme, Natl. Univ. of Singapore, Singapore, Singapore, [9]Dept. of Pathology and Lab. Med., Vancouver, BC, Canada, [10]Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [11]Dept. of Physiology, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [12]Penn Muscle Inst., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [13]Natl. Univ. Hosp., Singapore, Singapore, [14]Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore, Singapore, Singapore, [15]The Clinician Scientist Academy, Natl. Univ. Hlth.System, Singapore, Singapore, [16]Ctr. for Translational Med., Natl. Univ. Heart Ctr., Natl. Univ. Hlth.System, Singapore, Singapore, [17]Dept. of Epidemiology and Biostatistics, Imperial Coll. London, London, United Kingdom, [18]Natl. Skin Ctr., Singapore, Singapore

Abstract:

DCM-linked methylation changes influence the expression of key genes in cardiac disease. We aimed to clarify DCM mechanisms influenced by CpG methylation using multi-omics and causal analyses in the largest existing DCM cohort of left ventricular tissues (n = 235 DCM, n =179 control).

We used the EPIC array to map DNA methylation at approximately 850k CpG sites in left-ventricular tissue samples from failing and non-failing hearts across two independent DCM

cohorts (discovery n = 329, replication n = 85). This surfaced 194 replicated, independent methylation signals robustly associated with DCM (discovery Bonferroni-corrected P <0.05), enriched in regions of active transcriptional regulation. Array-based genotyping and RNA sequencing were performed on the same samples (discovery n = 306, replication n = 85). We identified and replicated significant relationships between the methylation levels of 183 sentinel CpGs and the expression of 849 proximal genes (±1 Mb).
To explore the causal involvement of sentinel CpGs in DCM, we applied Summary data-based Mendelian Randomization (SMR), integrating publicly available GWAS and eQTL data with meQTL data from our DCM cohort. Two sentinel CpGs were causally linked to DCM, and 36 sentinel CpGs to the expression of 43 unique proximal genes (SMR P <0.05). Of the 36 sentinel CpGs causally linked to gene expression, three were supported by Bayesian colocalization analysis to have a single causal variant underlying methylation and target gene expression (PP.H4 > 0.4). Genes mapped to the three sentinel CpGs (IER5, ENPTD6, ABHD12, KCNC4) had functions relevant to DCM pathogenesis, warranting future investigation in a cardiac context. Beyond the limited CpG sites probed by the EPIC array, targeted methylation sequencing around DCM sentinel CpGs detected additional signals linked to cardiovascular traits such as hsCRP and diastolic blood pressure, highlighting aspects of DCM pathogenesis potentially regulated by CpG methylation.
Beyond single-loci associations, multiple co-methylation modules identified by WGCNA were enriched in gene sets related to cardiac development, inflammation, and ECM alterations. Specific co-methylation modules were enriched in muscle adaptation and metabolism gene sets, along with binding sites of transcriptional regulators of these pathways (THRA, PTRF, NSDD2, HOXA9, LZBTB, RXR; p < 0.001) and key effectors of the Hippo pathway for cardiac development (YAP/TEAD; p < 0.001).
To our knowledge, this study is the first to elucidate the causal involvement of cardiac methylation changes linked to DCM. Our findings nominate several targets for experimental studies to probe DCM pathogenesis.

Session Title: Epigenetics Poster Friday Session

Board 1023F: KMT2D-deficiency in chondrocytes induces hypoxia responses in normoxia and causes precocious differentiation

Authors:

S. Halldórsdóttir[1], H. Bjornsson[2]; [1]Univ. of Iceland, Reykjavík, Iceland, [2]Univ. of Iceland, Reykjavik, Iceland

Abstract:

Kabuki syndrome (KS) is a rare Mendelian disorder of the epigenetic machinery caused by heterozygous loss of function mutations in the gene encoding the histone methyltransferase KMT2D. Postnatal growth retardation is a defining characteristic in KS and is seen in majority of individuals but KMT2D involvement in chondrocyte differentiation remains unclear. We have previously observed precocious differentiation resulting in premature maturation of two disease relevant cell types in KS, KMT2D-deficient neurons and chondrocytes as well as elevated levels of HIF1A in neurons. Since hypoxia is known to be a vital physiological factor in chondrocyte differentiation we hypothesized that KMT2D deficiency in chondrocytes also increases HIF1A and plays a role in driving cells towards chondrocyte differentiation. scRNA sequencing performed over four differentiation time-points in KMT2D-deficient ATDC5 cells (*Kmt2d*$^{-/-}$) and WT revealed that even prior to induced differentiation cells show advanced states of maturation suggesting the lack of a differentiation barrier and the accelerated differentiation is sustained through the differentiation trajectory. The precocious differentiation is observed in normoxic (20%O_2) conditions ($p<0.01$) but in hypoxia (1%O_2), *Kmt2d*$^{-/-}$ cells fail to transition to hypertrophic state, resulting in no difference from WT (ns). Bulk RNA sequencing was performed on undifferentiated cells both in normoxia and hypoxia but we did not observe a global change in gene expression with hypoxia exposure. Undifferentiated *Kmt2d*$^{-/-}$ cells display elevated expression of *Hif1a* both in normoxia and hypoxia along with an increased hypoxia response seen with enrichment of upregulated HIF1A-targets both in normoxia (p-value 7.2e-20) and hypoxia (p-value 2e-07) but no enrichment in downregulated genes (p-value normoxia = 0.29, hypoxia = 0.071). To interrogate the role of histone modifications we performed CUT&RUN. Although we do observe changes of H3K4me1 and H3K4me3 at promoter/enhancers of *Hif1a*, this does not occur at KMT2D-bound loci suggesting an indirect role. Finally, using a Seahorse Cell Mito Stress Test we observe decreased oxygen consumption of *Kmt2d*$^{-/-}$ cells compared to WT in normoxia, indicating an altered mitochondrial function. This metabolic change is a likely culprit of the elevated HIF1A levels since Reactive Oxygen Species (ROS) are known to both upregulate transcription as well as stabilize HIF1A. Our results suggest that KMT2D-deficient chondrocytes are primed towards precocious differentiation with abnormal hypoxia responses and oxygen consumption rate, a likely cause of the precocious differentiation phenotype.

Session Title: Epigenetics Poster Friday Session

Board 1024F: MeSS: a DNA methylation smoking score to quantify individual-level long-term tobacco exposure and its health impacts

Authors:

C. Wang, L. Ding, H. Guo, T. Wu; Huazhong Univ. of Sci. and Technology, Wuhan, China

Abstract:

Smoking is a leading risk factor for many major chronic diseases, including lung cancer. However, self-reported smoking status often suffer from reporting or recalling bias. Smoking, on the other hand, has a widespread impact on DNA methylation (DNAm). Here, we propose the Methylation Smoking Score (MeSS) to objectively quantify an individual's exposure to tobacco. We first performed a large-scale epigenome-wide association study (EWAS) of smoking, consisting of 4739 Chinese with blood DNA methylation data at 739,645 CpG sites. We identified 2448 smoking differentially methylated positions (DMPs), from which 96 DMPs were selected by logistic LASSO regression to construct MeSS. We showed that the score is accurate in discriminating never and current smokers (AUC = 0.970), with former smokers having intermediate MeSS. Also, we found a significant nonlinear increasing trend of the score with pack-years among current smokers and a linear decreasing trend of the score with years since cessation among former smokers, supporting the health benefits of reducing or ceasing cigarette consumption. In a case-cohort study, we demonstrated that MeSS could predict incident lung cancer in addition to self-reported smoking status (HR = 2.48, $P<0.001$). In summary, we have constructed a DNAm-based score that can accurately quantify an individual's tobacco exposure and predict incident lung cancer risk better than self-reported smoking status. These findings highlight the potential of MeSS to serve as a clinical biomarker for smoking-related health risk assessment.

Session Title: Epigenetics Poster Friday Session

Board 1025F: Neuroepigenomic Profiling of Opioid Use Disorder in Human Neurons of the Orbitofrontal Cortex

Authors:

S. Nagamatsu[1], G. Rompala[2], J. Martínez-Magaña[3], D. L. Nuñez-Ríos[1], J. Wang[4], M. Girgenti[4,1], J. Krystal[5], J. Gelernter[6], Traumatic Stress Brain Research Group[5], Y. Hurd[7], J. Montalvo-Ortiz[3]; [1]Yale Univ., Sch. of Med.; VA Connecticut Hlth.care System, New Haven, CT, [2]Mount Sinai, New York, NY, [3]Yale Univ., Orange, CT, [4]Yale Univ., New Haven, CT, [56]Yale Univ Sch. of Med., West Haven, CT, [7]New York, NY

Abstract:

Introduction: Opioid use disorder (OUD) is an epidemic in the US associated with high mortality rates. It may cause impairment and disabilities, leading to overdose when taken in large

amounts. Although there is a genetic component associated with OUD, a missing heritability is still observed, and can be, in parts, explained by epigenetic mechanisms, which may modulate the interaction between environmental and genetic factors. DNA methylation (5mC) in CpG sites is the most studied mechanism. However, there are additional epigenetic modifications highly abundant in the brain, such as DNA hydroxymethylation in CpG sites (5hmC), and 5mC at non-CpG sites (mCpHs; H indicates Adenine, Thymine, or Cytosine). This work aims to evaluate the interaction between 5mC, 5hmC, and mCpHs in neurons of the human postmortem orbital frontal cortex (OFC) in the context of OUD. Methods: A total of 38 Postmortem OFC samples were obtained from the VA Brain Bank (OUD=12;Control=26). 5mC, 5hmC, and mCpHs were assessed using reduced representation oxidative bisulfite sequencing (RRoxBS) in neuronal nuclei. Differential analysis was performed considering age, ancestry, postmortem interval, PTSD, and smoking status as covariates using methylkit R package. Genome-wide significant (GWS) sites were set at q-value<0.05. Functional enrichment and drug-gene interaction analyses were performed for the annotated genes of all GWS. Results: Differential analyses identified 397 5mC GWS, 1740 5hmC GWS, and 2,352 mCpH GWS associated with OUD. We observed CpGs in 10 genes common across the three epigenetic mechanisms examined, including RP9, GALNT2, ARSA, RUFY2, MAP4K5, ONECUT3, RAI1, ISCA1P4, PRICKLE1. Further, we also observed an overlap of 218 annotated genes between 5hmC and mCpH. Functional annotation analyses revealed trending significance enrichment for opioid signaling pathways (FDR-adjusted p<0.1) in 5mC and 5hmC at mCpGs, and glutamatergic synapse (FDR 9E-4-2.1E-2) in the mCpHs. Drug interaction analysis detected one known interaction with opioids in 5mC, seven in 5hmC, and 10 in mCpH. Conclusion: This is the first study to examine neuron-specific 5mC, 5hmC, and mCpH alterations associated with OUD in the human postmortem OFC. We identified overlap between different epigenetic mechanisms including opioid signaling for 5mC and 5hmC. The findings reveal an important role of 5hmC and mCpHs in OUD in OFC neurons.

Session Title: Epigenetics Poster Friday Session

Board 1026F: Novel Angelman and Prader-Willi Syndrome Mouse Models with a 6 Mb Deletion homologous to 15q11.2-q13 Deletion generated by CRISPR editing

Authors:

X. Lu, S. Bai, W. Philbrick, T. Nottoli, Y-H. Jiang; Yale Univ. Sch. of Med., New Haven, CT

Abstract:

More than 70% of Angelman syndrome (AS) and Prader-Willi Syndrome (PWS) cases are caused by deficiencies resulting from maternal or paternal deletions of the ~6 Mb chromosome 15q11.2-q13 region, respectively. The deficiency of maternal *UBE3A* is the known cause of AS, while the molecular basis of PWS is less well elucidated. Patients with paternal *SNORD116* deletion present key PWS features, and paternal *MAGEL2* deficiency associated with Schaaf-Yang syndrome has overlapping features with PWS. There are 8 AS mouse models with targeted mutations in different coding exons (*Ube3a*$^{m-/p+}$) on the maternal chromosome. These *Ube3a*$^{m-/p+}$ models recapitulate <10 % of AS cases and do not model the haploinsufficiency of non-imprinted genes within the 15q11.2-q13 region. Similarly, none of the >10 mouse models with mutations in the homologous region of 15q11.2-q13 fully recapitulate the deficiency of all paternally expressed genes and key features of PWS. The ongoing effort to develop molecular treatments for AS and PWS demands better models that accurately recapitulate the genetic defects found in the majority of patients. We generated new AS and PWS mouse models carrying a 6 Mb deletion of 15q11.2-q13 on the maternal allele for AS and the paternal allele for PWS. Using a multi-step strategy combining CRE/loxP and FLP/Frt with CRISPR, we successfully created novel mutant mice with a 6 Mb maternal deletion (AS-m^{LD}/p^{+}) or paternal deletion (PWS-m^{+}/p^{LD}), homologous to the human 15q11.2-q13 deletion. The deletion was confirmed by detection of the joint fragment using PCR and Sanger sequencing. Expression analysis in the brain of AS-m^{LD}/p^{+} mice validated the deficiency of *Ube3a* and the haploinsufficiency of biallelically expressed genes (*Tubgcp5, Cyfip1, Gabrg3, Gabra5, Gabrb3* and *Atp10a*). As expected, paternally expressed genes (*Snhg14, Snord115, Snord116, Snrpn, IPW, Magel2, Mkrn3* and *Ndn*) were not affected. AS-m^{LD}/p^{+} mice showed reduced travel distance, motor incoordination, impaired short- and long-term memory, and contextual and cued learning. Consistent with human AS with a large deletion, AS-m^{LD}/p^{+} mice were more susceptible to seizure induction compared to *Ube3a*$^{m-/p+}$ mice. PWS-m^{+}/p^{LD} mice exhibited more severe perinatal lethality than the PWS model with a paternal deletion from *Snrpn* to *Ube3a* (PWS-m^{+}/p$^{\Delta S-U}$). 100% of PWS-m^{+}/p^{LD} pups (N=73) died within the first week, compared to ~80% of PWS-m^{+}/p$^{\Delta S-U}$ pups, which died within two weeks after birth due to poor feeding. Therefore, AS-m^{LD}/p^{+} and PWS-m^{+}/p^{LD} mice offer improved molecular and phenotypic validity for modeling AS and PWS, providing greater value for future pre-clinical studies.

Session Title: Epigenetics Poster Friday Session

Board 1027F: Novel chromatin QTL mapping approach provides insight into the cistrome regulation and complex traits

Authors:

L. Wang, B. Fair, C. Buen Abad Najar, X. Liu; Univ. of Chicago, Chicago, IL

Abstract:

Regulatory variants that modulate chromatin state facilitate understanding of disease mechanisms through context-specific gene regulations. However, identifying genetic variants associated with chromatin states (cQTLs) is nontrivial due to inaccurate peak calling and limited power at the current sample size. To overcome the obstacles, we introduce CACTI, a PC-based multivariate association method, which improves cQTL detection by leveraging correlations among nearby regulatory elements. More specifically, CACTI groups multiple adjacent peaks within fixed-width windows, and tests the association between the multiple nearby peaks in each window and genetic variants using the PC-based omnibus test. CACTI significantly outperforms the standard single-peak cQTL detection approach. For example, CACTI detects 1554 additional cQTLs of mark H3K27ac in macrophage, while replicating 99% of 1295 cQTLs detected with the standard single-peak method.
Applying CACTI to five histone marks (H3K27ac, H3K4me1, H3K4me3, H3K36me3, and H3K27me3) in six cell types and tissues (LCL, macrophage, brain, heart, lung, and muscle; sample size N=35-113), we constructed a comprehensive map of cQTLs. Our powerful method allows us to identify weak cQTLs that were previously missed. The large numbers of cQTLs and the shared definitions of windows helped us to examine the sharing of cQTLs across marks and cell types. We found extensive sharing of cQTLs across marks. For example, 39-65% of H3K27ac cQTLs are H3Kme1 cQTLs and 47-49% are H3K4me3 cQTLs, at 5% FDR in LCL and macrophage. The correlation of cQTL effects between LCL and macrophage ranges from 0.01 (H3K27ac) to 0.15 (H3K4me1), indicating high tissue and cell-type specificity of cQTLs. To interpret complex trait genetics with cQTLs, we performed colocalization analyses of cQTLs and GWAS loci of 36 blood traits. Overall, 72 GWAS loci on average colocalized with cQTLs of any mark in any cell type (18-101 loci per trait), including 11 loci additionally explained by new cQTLs found by CACTI. Notably, 45 GWAS loci only colocalize with cQTLs, but not with any eQTLs in the eQTLGen (whole blood, N=31,684), supporting that cQTLs can offer additional insights that are missed by steady-state eQTLs.

Session Title: Epigenetics Poster Friday Session

Board 1028F: Novel epigenetic mediators for Barker's hypothesis: A Mendelian randomization study

Authors:

J. Tang, N. Y. T. Ng, A. W. S. Kwok, C. C. Y. Mak, B. H. Y. Chung; The Univ. of Hong Kong, Hong Kong, China

Abstract:

Barker's hypothesis posits that adverse in utero exposures, often indicated by low birthweight, increase the risk of adult cardiometabolic diseases such as cardiovascular disease (CVD). However, the mechanism for Barker's hypothesis remains unclear and germline genetic variation alone is insufficient to explain this association. Epigenetics is a viable alternative as it dynamically affects gene transcription. DNA methylation (DNAm) is among the most studied epigenetic markers, but prior studies only used DNAm measured in adulthood and do not account for temporal ordering in Barker's hypothesis.

To address the gap, we used epigenetic Mendelian randomization (MR) with Steiger testing and colocalization as a primary analysis. For exposures, we used birthweight-associated CpG sites (BW-CpGs) from epigenome-wide association studies (EWAS), with data from the Norwegian Mother and Child (MoBA) cohort and the Pregnancy and Childhood Epigenetics (PACE) consortium. We used methylation quantitative trait loci (mQTLs) measured at birth to proxy BW-CpGs. Outcome data were from genome-wide association studies (GWAS) of lipid levels and risk of CVD outcomes. To identify persistent DNAm signatures from birth, childhood to adolescence, we repeated the primary analysis with mQTLs measured at these timepoints. We interrogated functional relevance with a scoring system based on data availability from other EWAS, GWAS, and expression quantitative trait methylation and expression quantitative trait loci (eQTLs) datasets.

We found 32 significant associations (FDR-corrected $P < 0.05$) between DNAm of 22 BW-CpGs at birth and adult lipid levels and risk of CVD outcomes. F-statistics of all mQTLs suggested strong instruments ($F > 10$). All associations indicated correct causal direction (Steiger $P < 0.05$). Colocalization ruled out horizontal pleiotropy for 16 of the 32 associations ($PP > 0.5$). We identified persistent DNAm from birth through adolescence near *ASGR1*, *KDM2B*, and *SLC12A4*. We used eQTL data, other EWAS and GWAS to prioritize ASGR1 as a potential druggable target, since two BW-CpGs near it had the highest scores. This concurs with a recent drug-target MR study and a Phase 1 clinical trial on a monoclonal antibody targeting ASGR1 to lower CVD risk. For the first time, causal evidence was established for DNAm as a mechanism in Barker's hypothesis using a robust, multistep, epigenetic MR framework, overcoming the limitations of prior studies. Our hypothesis-free approach identified candidate genes to mediate adverse CVD outcomes per Barker's hypothesis.

This study was supported by the Society of the Relief of Disabled Children.

Session Title: Epigenetics Poster Friday Session

Board 1029F: Schizophrenia methylation profile score in a Brazilian antipsychotic-naive first-episode psychosis cohort.

Authors:

R. Ormond Sampaio[1], V. Ota[1], L. Spindola[2], S. Le Hellard[2], A. Gadelha[1], C. Noto[1], G. Breen[3], J. Montalvo-Ortiz[4], S. Belangero[1], M. Santoro[1]; [1]Univ.e Federal de São Paulo, São Paulo, Brazil, [2]Univ. of Bergen, Bergen, Norway, [3]King s Coll. London, London, United Kingdom, [4]Yale Univ., Orange, CT

Abstract:

Schizophrenia (SCZ) is one of the most heritable mental disorders, with an estimated heritability of 0.8. However, the most recent genome-wide association study could only capture 7.3% of the variance of this heritability. Epigenetic factors, such as DNA methylation, can have a major impact on SCZ risk, and could help to understand the biological underpinnings of the disorder, as they can also reflect environmental changes, which also contribute to SCZ risk. Studies in the early stages of SCZ may be beneficial, since, although the disease is already established, there are no confounding factors acting, such as time of illness and treatment effect. Evaluating DNA methylation in multiple CpG sites, such as in Epigenetic Wide Association Studies (EWAS), can offer potential biomarkers of trait and state and allows the development of a Methylation Profile Score (MPS). The objective of this study is to verify whether MPS is associated with FEP and its changes after risperidone treatment. This study includes 117 antipsychotic-naive FEP which were evaluated at the baseline and after two months of risperidone treatment, and 114 controls were included. Blood was collected and DNA methylation was assessed using 450 K array or EPIC v1.0 array (Illumina). The MPS for each sample was calculated based on the two largest blood schizophrenia EWAS as training dataset: Hannon et al. 2021, for schizophrenia (SCZ-Hannon, N= 3264) and for psychosis (PSI-Hannon, N=4483) and Li et al. 2021, for schizophrenia (SCZ-Li, N= 999). We performed a logistic regression in our test sample for each dataset (450K or EPIC) to verify association between MPS and FEP (compared to controls), controlling for smoking, batch, age, sex and cell types, and mixed models to verify association with risperidone treatment, controlling for smoking, batch and cell types. We considered statistically significant results with p-value < 0.05. SCZ-Li MPS was significant between cases and controls (Epic array) (coefficient = -2.776; p = 0.031; OR = 0.0623 [95% CI 0.004-0.704]), indicating an inverse negative association with the phenotype. Of note, there were also statistically significant differences in smoking score (450K array), and CD8T cell type (Epic array) comparing FEP cases and controls. No association between any MPS and treatment was identified in the longitudinal analysis, but all cell types proportions changed after risperidone (450k). In conclusion, although we found a significant association, caution is needed and a next

step, we will evaluate additional MPS models and integrate them with Polygenic Scores (PGS) to enhance predictive accuracy.

Session Title: Epigenetics Poster Friday Session

Board 1030F: Single cell multiomic profiling of the human lung in health and disease to understand genetic risk of complex respiratory disease

Authors:

E. Griffin[1], J. Buchanan[1], Q. Yang[1], S. Corban[1], W. Elison[1], R. Melton[1], H. Mummey[1], Y. Lee[1], L. Tucciarone[1], C. McGrail[1], J. Chi[1], J. Verheyden[1], G. Pryhuber[2], L. Croty-Alexander[1], A. Wang[3], X. Sun[1], K. Gaulton[4]; [1]UCSD, La Jolla, CA, [2]Univ. of Rochester, Rochester, CA, [3]Univ. of California San Diego, La Jolla, CA, [4]UCSD, San Diego, CA

Abstract:

Genome wide association studies (GWAS) have identified a substantial number of genomic loci associated with complex respiratory diseases including asthma, idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), and SARS-CoV2 infection. The majority of variants associated with respiratory disease map to non-coding regions and likely affect gene regulatory programs in disease-critical cell types. Our work seeks to understand novel mechanisms and therapeutic targets for respiratory disease by identifying cell types in the lung impacting disease initiation and progression both broadly and at individual risk loci. In this study, we performed single cell multiome (paired RNA-seq and ATAC-seq) profiling of 98 human lung samples including donors without disease as well as 9 disease states including asthma, IPF, and history of COVID with and without pulmonary fibrosis. We created an integrated cell type map of the human lung consisting of 229,690 cells representing 37 cell types, which included rare or lower abundant cell types such as pulmonary neuroendocrine cells (PNECs) and proliferating AT2 cells. We use this integrated map to define cell type-specific gene regulatory programs in the lung including expressed genes, accessible chromatin sites, site-target gene links, and transcription factor networks. We used cell type-specific gene regulatory programs to perform genome-wide enrichment tests to identify cell types broadly involved in asthma, COPD, SARS-CoV2 infection, and IPF risk. These results highlight cell types that may play a role in disease etiology or may be affected by disease progression. Additionally, in combination with genome-wide genotyping data, we performed quantitative trait loci mapping of chromatin accessibility (caQTLs) and gene expression (eQTLs) in each lung cell type, and identified both shared, lineage-specific, and cell type-specific QTLs. We identified disease loci statistically colocalized with lung cell type QTLs, which revealed specific loci modulating disease risk and

their putative cellular mechanisms. Our findings highlight the power of single cell multiomics in uncovering cellular and genetic architecture of lung disease etiology and progression, as well as the importance of integrating multiomic data at single cell resolution for QTL mapping to interpret complex disease risk loci in individual cell types.

Session Title: Epigenetics Poster Friday Session

Board 1031F: Single-cell multi-omics analysis of transcription factor cooperativity in oligodendrocyte gene regulation

Authors:

J. Choi[1,2], J. Svaren[1,2], D. Wang[1,2]; [1]Univ. of Wisconsin - Madison, Madison, WI, [2]Waisman Ctr., Madison, WI

Abstract:

Oligodendrocytes are the myelinating cells within the central nervous system, playing a crucial role in myelination. The genes expressed in oligodendrocytes are typically associated with brain disorders, and are controlled by complex gene regulatory mechanisms involving transcription factors (TFs). Not only do individual TFs work, but co-binding TFs cooperatively work to co-regulate gene expression from the regulatory elements. Nevertheless, the regulatory mechanisms of cooperative TFs in oligodendrocytes remain uncharacterized.To address this, we integrated published scRNA-seq and scATAC-seq data to identify the cooperative TFs that co-regulate the expression levels of target genes (TGs) in oligodendrocytes. First, we identified co-binding TF pairs whose binding sites overlapped in oligodendrocyte-specific regulatory regions. Second, we trained a deep learning model to predict the expression level of each TG using the expression levels of co-binding TFs. Third, using the trained models, we computed the TF importance and TF-TF interaction scores for predicting TG expression levels by the Shapley-Taylor interaction index. Finally, we found co-binding TF pairs with high interaction scores as "cooperative" TFs for gene regulation in oligodendrocytes.We identified 8,101 co-binding TF pairs among 206 TFs for 445 TGs in 643 oligodendrocyte-specific regulatory regions via scATAC-seq peak-gene links. We found that the co-binding TF pairs involving key TFs, such as SOX10-TCF12, SOX10-MYRF, and SOX10-OLIG2, exhibited higher interaction scores compared to other co-binding TF pairs (t-test, p-value < 1e-4). Additionally, we found that some non-key TFs have high interaction scores, such as RBPJ-TCF12 and ELF2-TCF12, providing potential novel insights into regulatory mechanisms in oligodendrocytes. Furthermore, 153 out of 160 TGs that are the eGenes of oligodendrocyte-associated eQTLs are regulated by cooperative TFs, including key TFs, suggesting potential causal relationships between genetic variants and gene regulatory

regions where cooperative TF pairs are present. Additionally, we implemented co-enrichment analysis for binding sites to validate cooperative TF pairs, such as SOX10-OLIG2 and SOX10-NKX2.2, using rat ChIP-seq data.

Session Title: Epigenetics Poster Friday Session

Board 1032F: Sleep, Epigenetics, and Childhood Obesity: Unveiling Novel DNA Methylation Patterns

Authors:

E. Richter[1], P. J. Patel[2], Y. Y. Ozdemir[1], U. V. Nnyabbe[1], J. R. Babu[1], T. Geetha[1]; [1]Auburn Univ., Auburn, AL, [2]Emory Univ., Atlanta, GA

Abstract:

The growing recognition of sleep's profound impact on health and quality of life has ignited a surge in research exploring its potential to prevent chronic diseases through practical preventative strategies. Sleep plays a crucial role in shaping epigenetics, the study of heritable changes in gene expression without altering the DNA sequence itself, by emerging as a key factor influencing public health outcomes. Previous research has shown a link between unhealthy sleeping patterns and changes in DNA methylation of children, however, not much has been studied on understanding the role of sleep timings on epigenetic regulations. In this study, we aim to identify the methylation differences in the topmost associated genomic location with the sleep timing of children. We measured the epigenome-wide DNA methylation using Illumina MethylationEpic BeadChip Array from the saliva of children aged 6-10 years. Genomic location selection was performed by firstly screening all locations based on cross-validation error through logistic regression. Then, Random Forest and Lasso models were run on the screened locations, with notable overlap in selected loci, indicating potential significance in determining children's sleeping patterns. Additional variables such as race, sex, maternal education, and family income were adjusted. The results showed that 23 target IDs were differentially methylated in children with late bedtime (after 8:30 pm) compared to early (before 8:30 pm). Out of which 18 genomic locations were hypomethylated while only 5 were hypermethylated. Our KEGG analysis indicates that the most enriched genes from those (*HOMER2*, *HOMER3*, *RAP1A,* and *GLI3*) play a major role in processes such as Glutamatergic synapse, FoxO signaling pathway, and cAMP signaling pathway. All these signaling pathways play interconnected roles in regulating sleep-wake cycles, impacting neuronal activity, cellular metabolism, and gene expression. These findings highlight the intricate interplay between sleep

patterns and epigenetic modifications, offering new insights into the molecular mechanisms underlying sleep regulation and its impact on children's health and well-being.

Session Title: Epigenetics Poster Friday Session

Board 1033F: Structural Genomics Reveals the Effects of Rare Disease Mutations on RAG2 Epigenetic Reader Functions

Authors:

N. Haque[1], J. Wagenknecht[2], D. Jensen[1], J. Licklider[1], B. Smith[1], R. Urrutia[1], M. Zimmermann[3]; [1]Med. Coll. of Wisconsin, WAUWATOSA, WI, [2]Med. Coll. of Wisconsin, Franklin, WI, [3]Med. Coll. of Wisconsin, Milwaukee, WI

Abstract:

Molecular recognition plays a pivotal role in the regulation of intricate biological phenomena. One notable instance is the precise identification of methylated histone H3 by the PHD finger domain of RAG2, which is postulated to be a critical determinant in maintaining the distinctive gene recombination necessary for immunological function. The RAG1/2 V(D)J recombinase complex is instrumental in orchestrating the assembly of B-cell and T-cell receptor repertoires through antigen receptor gene rearrangement [1]. Thus, the epigenetic reader function of RAG2 is critical for transducing genome regulation signals into adaptive immunity. We previously characterized a broad repertoire of genetic variants identified from patients with severe combined immunodeficiency (SCID) [2-3] and affecting the RAG core [4]. In the current investigation, we examined the dynamic behavior of RAG2 PHD bound to methylated and non-methylated histone H3 peptides, and across distinct genetic variants also identified from SCID patients. We leveraged robust molecular modeling and dynamics simulations to calculate specific details of protein-peptide interactions and validated findings using experimental measurements of binding affinities. Our findings indicate that the PHD domain exhibits a partially rigid structure that is influenced by the N-terminal region of the histone H3 peptide. Upon interaction with the histone H3 peptide, the PHD domain adopts an extended β-sheet conformation, with specific residues (ALA1, ARG2, THR3, LYS4, GLN5, THR6, and ALA7) from the N-terminal of histone H3 interacting with the first β-sheet (residues 442-446) of the PHD domain. Importantly, the affinity and structural arrangement of the PHD-peptide complex vary depending on the presence and methylation status of specific residues. Notably, the RAG2 autoregulation segment (residues 397-407) exhibits negligible interaction with the PHD finger in the absence of methylated peptide. Conversely, in the presence of methylated LYS4 alone or in combination with symmetric and asymmetric ARG2 methylation, this segment demonstrates an

increased propensity for interaction with the PHD domain. Such alterations in the topology of the PHD domain induced by methylated residues enhance the affinity of the histone H3 peptide for the PHD domain. Consequently, our investigation provides insights into the molecular mechanisms underlying critical biological processes, the disruption of which can lead to SCID.

Session Title: Epigenetics Poster Friday Session

Board 1034F: The Multi-Ancestry Genomics, Epigenomics, and Transcriptomics of Alzheimer's (MAGENTA) Project: Understanding Alzheimer's Disease through DNA Methylation Analysis

Authors:

A. Griswold[1], T. Gu[1], L. Gomez[1], C. Kosanovic[1], D. Van Booven[1], M. Mews[2], K. Hamilton-Nelson[1], J. Sanchez[1], L. Adams[1], P. Mena[1], P. Whitehead[1], T. Starks[3], C. Silva[4], M. Illanes-Manrique[5], M. Cuccaro[1], J. Vance[1], M. Cornejo-Olivas[5], B. Feliciano-Astacio[4], J. Haines[2], M. Pericak-Vance[1], W. Bush[2]; [1]Univ. of Miami, Miami, FL, [2]Case Western Reserve Univ., Cleveland, OH, [3]Wake Forest Univ., Winston-Salem, NC, [4]Univ. Central Del Caribe, Bayamon, PR, [5]Neurogenetics Res. Ctr. - Inst. Natl. de Ciencias Neurologicas, Lima, Peru

Abstract:

Background: Effects of AD associated genetic loci can vary drastically between individuals of diverse ancestral backgrounds. The Multi-Ancestry Genomics, Epigenomics, and Transcriptomics of Alzheimer's (MAGENTA) project aims to determine how these differences in the genetic architecture of AD contribute to underlying functional outcomes such as gene expression and epigenetic features like DNA methylation. Such studies are critical to discover disease mechanisms that may inform early potential therapeutic interventions across diverse populations. **Methods:** The MAGENTA project has generated blood DNA methylomes with the Illumina MethylationEPICv2 on individuals from diverse ancestries array (N=626; 135 European, 204 African Americans, 205 Caribbean Hispanics, 82 Peruvians), divided between AD and cognitively unimpaired (CU). Differentially methylated sites and regions were determined using the SeSAMe R package using a linear model with covariate variables sex, age, methylation batch, global genetic ancestry, and estimated immune cell proportions. **Results:** When combining all samples, we identified 906 CpG sites with nominally significant differences (p-value <= 0.001) between AD and CU. Notably, when restricting this analysis within each cohort, the number of differentially methylated sites differed: European - 572 sites, African American - 1033 sites, Caribbean Hispanic - 448, Peruvians - 422. Notably, however, the markers across the ancestral groups did not overlap. Similarly, only one region (on chr20q13 near the *GNAS* gene)

overlapped across cohorts, though several other regions showed differences in an ancestry specific manner. **Conclusions:** Identifying the genes and biological pathways involved in AD is critical to develop effective therapies. The MAGENTA project analyses suggest the possibility of whole blood DNA methylation patterns as ancestry specific signatures of AD pathogenesis. While expansion into continental African and Pacific Islander populations is ongoing to further represent diversity in the MAGENTA project, our current data indicate prominent differences in the methylation architecture across ancestries that highlight the importance of inclusion of diverse individuals for a comprehensive understanding of epigenetic risk for AD.

Session Title: Epigenetics Poster Friday Session

Board 1035F: Tracing Early Fetal Chromatin Signatures in Schizophrenic Brains

Authors:

J. Bendl[1], K. Girdhar[2], A. Baumgartner[3], K. Therrien[4], S. Venkatesh[4], D. Mathur[4], P. Dong[5], S. Rahman[4], S. Kleopoulos[4], R. Misir[1], S. Reach[1], P. K. Auluck[6], S. Marenco[6], D. A. Lewis[7], V. Haroutunian[8], C. Funk[3], G. Voloudakis[4], G. Hoffman[4], J. Fullard[4], P. Roussos[4]; [1]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [2]Mount Sinai, New York, NY, [3]Inst. for Systems Biology, Seattle, WA, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Icahn school of Medcine at Mount Sinai, New York, NY, [6]Human Brain Collection Core, Natl. Inst. of Mental Hlth.-Intramural Res. Program, Bethesda, MD, [7]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [8]James J. Peters VA Med. Ctr., Bronx, NY

Abstract:

Non-coding variants increase the risk of neuropsychiatric diseases, but the cell-type-specific roles of the non-coding genome in these conditions are not well understood. We conducted population-scale chromatin accessibility profiling (N=1,393) of neurons and non-neurons from two neocortical brain regions: the anterior cingulate cortex and the dorsolateral prefrontal cortex. Across both regions, we found significant differences in neuronal chromatin accessibility between schizophrenia cases and controls. A per-sample disease pseudotime analysis showed a positive correlation with genetic liability for schizophrenia. By organizing chromatin into cis- and trans-regulatory domains, we identified a prominent neuronal trans-regulatory domain (TRD1) active in immature glutamatergic neurons during fetal development. Polygenic risk score analysis using genetic variants within the chromatin accessibility of TRD1 successfully predicted schizophrenia susceptibility in the Million Veteran Program cohort. Overall, we present the most extensive resource to date on chromatin accessibility in the human cortex, providing new

insights into the cell-type-specific etiology of schizophrenia. p { line-height: 115%; margin-bottom: 0.1in; background: transparent }

Session Title: Epigenetics Poster Friday Session

Board 1036F: Understanding the functional consequences of rapid evolution of sperm protamines

Authors:

R. Agrawal, L. Moritz, M. Rabbani, S. Hammoud; Univ. of Michigan, Ann Arbor, MI

Abstract:

Protamine-mediated DNA compaction in sperm, achieved by electrostatic interactions between DNA and the arginine-rich core of protamines, is essential for sperm chromatin structure, development and reproductive fitness of an organism. Protamines are subject to strong positive selection with an unusual form of purifying selection that tightly maintains arginine/serine-rich regions, but not strict sequences among the species. The significance of the rapid evolution of protamines and protamine protein sequence changes outside the arginine cores is not well understood. Here, we show that the single amino acid substitutions in the mouse P1 sequence with the residues that exist in other species in nature leads to impaired progressive motility of sperm, sperm structural abnormalities, and severe subfertility. However, mice lines replaced with the mulatta P1 protamines sequences show no fertility defects in natural mating. This suggests that potentially multiple residues co-evolve in a species to maintain the structure function relationship of protamine proteins. Overall, the rapidly evolving protein sequences, variation in protein composition protein net charge, and potential sites of post-translational modification across species raises the possibility of a species specific 'protamine code' and function.

Session Title: Epigenetics Poster Friday Session

Board 1038F: Variants at adipocyte open chromatin act in *cis* and interact in *trans* on abdominal obesity risk ⭐

Authors:

M. Sukhatme[1], A. Kar[1,2], U. T. Arasu[3], S. T. Lee[1], M. Alvarez[1], K. M. Garske[1], K. Z. Gelev[1], S. Rajkumar[1], S. Das[1], D. Kaminska[4,5], V. Männistö[6,7], H. Peltoniemi[8], S. Heinonen[9], U. Säiläkivi[10], T.

Saarinen[10], A. Juuti[10], K. Pietiläinen[9,11], J. Pihlajamäki[4,12], M. U. Kaikkonen[3], P. Pajukanta[1,2,13]; [1]Dept. of Human Genetics, UCLA, Los Angeles, CA, [2]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [3]A. I. Virtanen Inst. for Molecular Sci., Univ. of Eastern Finland, Kuopio, Finland, [4]Inst. of Publ. Hlth.and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, [5]Dept. of Med., Div. of Cardiology, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [6]Inst. of Clinical Med., Internal Med., Univ. of Eastern Finland, Kuopio, Finland, [7]Dept. of Internal Med., Kuopio Univ. Hosp., Kuopio, Finland, [8]Tilkka Hosp., Helsinki, Finland, [9]Obesity Res. Unit, Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [10]Dept. of Abdominal Surgery, Abdominal Ctr., Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, [11]Hlth.yWeightHub, Endocrinology, Abdominal Ctr., Helsinki Univ. Central Hosp. and Univ. of Helsinki, Helsinki, Finland, [12]Dept. of Med., Endocrinology and Clinical Nutrition, Kuopio Univ. Hosp., Kuopio, Finland, [13]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract:

Abdominal obesity (AOb) is the metabolically harmful form of obesity, likely impacted by both genetic and epigenetic factors. To elucidate the interplay between these factors, we first investigated the AOb heritability (h^2) enrichment of SNPs in subcutaneous adipose tissue (SAT) adipocyte open chromatin peaks using ATAC-seq data from differentiated human adipocytes and found that their AOb h^2 was strongly enriched (h^2 enrichment=17.7 ($p=1.52x10^{-28}$), percent h^2=75.7%) compared to all other SNPs (h^2 enrichment=0.254 ($p=1.51x10^{-28}$), percent h^2=24.3%). Based on this, we hypothesized that there are undiscovered risk variants at adipocyte open chromatin that act in *cis* and interact in *trans* via chromosomal interactions on AOb risk. To test this, we identified adipocyte marker genes specific to the SAT (MGSS) using single nucleus RNA-seq (snRNA-seq) data from the KOBS cohort (n=7) and evaluated the risk contributions of the SNPs within the ATAC-seq peaks in the *cis* regions of these genes for AOb in the UK Biobank (UKB). Through a longitudinal differential accessibility test of ATAC-seq data across human SAT preadipocyte differentiation using DESeq2, we identified peaks upregulated in adipocytes, preadipocytes, and shared between the two. To further investigate SNPs in these ATAC peaks, we constructed polygenic risk scores (PRS) for AOb in UKB for SNPs in the 3 peak sets in all individuals, males, and females. The incremental variance (R^2) of all 9 partitioned PRSs was consistently >0.5%, all of which were significantly enriched ($p_{perm1000}<0.05$). The PRS SNPs were also enriched for CTCF binding sites ($p_{perm1000}<0.05$), identified by VEP. To find *trans* regulatory SNPs, we focused on open chromatin peaks in the *cis* region of the SAT master TF of human adipogenesis, *SREBF1*, with 7 AOb GWAS SNPs. Using SAT snRNA-seq data from the RYSA cohort (n=68), we discovered >100 MGSS genes with altered adipocyte expression in *trans* by the carrier status of the AOb-increasing allele of the 7 SNPs, in line with their *cis* effects on *SREBF1* expression. For each *trans* gene set, we built regional PRSs in all, females, and males

using their *cis* regional open chromatin SNPs in PLINK, while removing the 7 SNPs and their LD proxies (R^2>0.1). We observed greater regional AOb PRSs (Wilcoxon $1.83x10^{-10}$<p<0.0001) in the carriers vs. non-carriers of the 7 *trans*-acting SNPs in all individuals or females. The majority of PRSs were also enriched for variance explained in AOb ($p_{perm1000}$<0.05). Overall, we find the first example of *cis-trans* interactions on the risk of AOb at adipocyte open chromatin enriched for CTCF binding, likely facilitating *cis-trans* interactions on AOb via chromosomal interactions.

Session Title: Epigenetics Poster Friday Session

Board 1039F: Whole-blood DNA methylome sequencing identifies genes associated with asthma in a family cohort

Authors:

B. Tremblay, A-M. Madore, C. Laprise; Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

Abstract:

Background: Asthma is a chronic inflammatory disease of the respiratory tract characterized by airway hyperresponsiveness and obstruction. It is part of the atopic march, a pattern of progression of allergic diseases in which young children having moderate to severe atopic dermatitis (MSAD) have 35% to 75% increased risk of developing food allergy, allergic asthma, and allergic rhinitis. A previous analysis has identified a polygenic risk score (PRS) with 88% and 71% discriminative values to identify children at risk of developing MSAD and asthma respectively. Epigenetic data as DNA methylation levels could improve this predictive tool. However, most epigenetic studies on asthma focused on children or adults and used microarray data. Thus, our objective is to identify CpG sites in immune regulatory regions, measured using targeted bisulfite sequencing, that are associated with asthma in a large French-Canadian family cohort of children and adults.**Methods:** A total of 676 children and adults (3-80 years) from the *Saguenay-Lac-Saint-Jean* asthma family cohort were recruited. Asthma was clinically confirmed by a doctor (416 asthmatics and 260 controls). DNA methylation data (n=5,210,165) were obtained from a custom sequencing panel targeting immune regulatory regions. Linear mixed effects models between DNA methylation levels and asthma status adjusted for fixed (age, sex, smoking status, cell counts, surrogate variables) and random (family ID) effects were computed using R software. **Results:** A total of seven CpG sites were associated with the asthma status (P<$1x10^{-6}$). CpG sites were located in intergenic regions and in *SNX29*, *CHL1*, *CTU1*, and *LINC00202-2* genes. Besides *LINC00202-2*, all genes have been previously associated with asthma and immune response in the literature, but not at the DNA methylation

level. **Conclusions:** These results reveal novel CpG sites associated with asthma in a family cohort. This provides potential insights on the underlying effects of DNA methylation on asthma. These results will be compared with ongoing analyses on atopic dermatitis, in the same cohort, to improve our predictive polygenic risk score for atopic dermatitis and other allergic diseases of the atopic march.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3002F: 3023 human genomes from mainland Southeast Asia disclose hidden genetic diversity and signatures of tropical adaptation

Authors:

B. Su[1], Z. Xiaoming[1], P. Min-Sheng[1], L. Yuchun[1], L. Kai[1], Z. Bin[1], M. Yujie[1], G. Yongbo[1], Z. Wangshan[1], Z. Yu[1], Y. Tian[1], L. Shen-Ao[2], C. Lu[2], W. Xuankai[3], M. Yafei[3], D. Shanshan[1], G. Zongliang[1], Y. Liqin[1], S. Tuot[4], L. Bunnath[4], J. Kampuansai[5], M. Srikummool[6], H. Quang[7], S. Singthong[8], S. Sochampa[9], U. Kyaing[10], Q. Kong[1], Y. Zhang[1], Y. He[1]; [1]Kunming Inst. of Zoology, Chinese Academy of Sci., Kunming, China, [2]Sch. of Life Sci., Fudan Univ., Shanghai, China, [3]Shanghai Jiao Tong Univ., Shanghai, China, [4]Dept. of Geography and Land Management, Royal Univ. of Phnom Penh, Phnom Penh, Cambodia, [5]Dept. of Biology, Faculty of Sci., Chiang Mai Univ., Chiang Mai, Thailand, [6]Dept. of Biology, Faculty of Sci., Naresuan Univ., Naresuan, Thailand, [7]Dept. of Immunology, Ha Noi Med. Univ., Ha Noi, Viet Nam, [8]The Natl. Univ. of Laos, Vientiane, Lao People's Democratic Republic, [9]Muang Sing District Hosp., Muang Sing, Lao People's Democratic Republic, [10]Field Sch. of Archaeology, Paukkhaung, Myanmar

Abstract:

Mainland Southeast Asia (MSEA) harbors rich ethnic and cultural diversity with a population of nearly 300 million. However, MSEA people are underrepresented in the current genomic database of global populations. Here we present the SEA3K genome dataset (Phase-I), consisting of the deep whole-genome sequencing (WGS) data of 3,023 individuals from 30 MSEA populations, together with the long-read WGS data of 40 representative MSEA individuals. We identified 71.91 million short variants and 43,462 structural variants, nearly half of which are novel. Using the long-read genome assemblies, we generated the first pangenome of MSEA populations, adding 339.36 Mb novel sequences to the current human reference genomes. We observed a high level of genetic heterogeneity across MSEA populations, reflected

by the varied combinations of genetic components inherited from ancestral hunter-gatherers and Neolithic farmers. Remarkably, we discovered strong signatures of Darwinian positive selection, suggesting a comprehensive adaptation to tropical environment that explains the distinct phenotypic traits in MSEA populations, including short stature, dark skin pigmentation, wide and snub nose, curly hair and prevalence of thalassemia. Furthermore, we observed different patterns of archaic Denisovans introgression in MSEA populations, supporting the proposal of at least two distinct instances of Denisovan admixture into modern humans in Asia. Importantly, we detected the genomic regions indicating "borrowed" fitness in MSEA populations, due to either Neanderthal or Denisovan adaptive introgressions. The reported multi-million novel genomic variants in MSEA highlight the necessity of studying regional populations that can help answer the key questions on prehistory, genetic adaptation and complex diseases.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3003F: A preliminary study on polygenic adaption in the Han Taiwanese people

Authors:

M-L. Kang[1], L. Jun-Hun[2], S. Yoko[3], L. Cheng_Jui[2], L. I-Hui[4], K. Wen-Ya[1]; [1]Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, [2]Mackay Mem. Hosp., Taipei, Taiwan, [3]Res. Ctr. for Integrative Evolutionary Sci., Sokendai, Kanagawa, Japan, [4]Taipei Veterans Gen. Hosp., Taipei, Taiwan

Abstract:

Despite the success in identifying a significant number of genes targeted by positive Darwinian selection in diverse ethnic groups, genomic signatures of polygenic adaptation have only been studied in a small number of populations. The complete adaptive genetic architecture underlying complex traits remains not fully understood, given that human populations are widespread across the Earth and have a complex migration history. Here, we studied polygenic adaptation in thirty-four cardiometabolic- or hematological-related traits in the Han Taiwanese people. For each trait, we first conducted genome-wide association studies (GWAS) in 74,984 individuals for a total of 586,528 SNPs (collected by the Taiwan Biobank). To detect subtle adaptive changes across numerous variants, the distribution of selection summary statistic (measured by single density score, SDS) for the SNPs of significant GWAS effects ($p<5x10^{-8}$) was compared with the null distribution of SDS from the background SNPs. The SDS scores were calculated by analyzing a subset of individuals (1,487) whose whole-genome sequencing data are available. As a result, we identified 14 (out of 34) traits showing a significant departure from

null distribution. Among them, eight traits displayed a tendency towards favoring trait-increasing alleles by selection while an opposite pattern was observed for the rest six traits. Particularly, our study reveals several novel findings of polygenic adaption. Natural selection appears to favor trait-creasing alleles for serum creatine, triglycerides, alpha-fetoprotein (AFP), glycated hemoglobin (HbA1c) as well as trait-decreasing alleles for albumin in serum, body-fat rate, heart-beat speed, and estimated Glomerular filtration rate (eGFR) in the Han Taiwanese people. Future research will focus on conducting LD score regression to distinguish between true polygenic signals and confounding factors in GWAS and the possible role of each adaptive polygenic trait in contributing to disease susceptibility.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3004F: Admixed X-linked hemophilia A patients from the Brazilian Study of Induction of Immune Tolerance (BrazIT) study do not show the pervasive African sex-ancestry bias of Latin American populations

Authors:

E. Tarazona-Santos, H. S. Sant'Anna, R. Tou, L. Farias, B. Miwa, J. Duarte, L. Zuccherato, R. M. Camelo, S. Rezende; Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract:

Hemophilia A, an X-linked bleeding disorder, presents significant treatment challenges due to its allelic heterogeneity and varying severity. We performed whole exome sequencing of the 193 patients from the Brazilian Immune Tolerance (BrazIT) cohort of hemophilia A patients high-responding inhibitors, identifying 222,314 genetic variants. BrazIT individuals have medians of 64%, 17% and 13% of European, African and Native American ancestries respectively, estimated from autosome SNVs. After controlling for appropriate covariates, we did not observe association between these continental ancestries and success of ITI treatment or inhibitor titers, but we observed that Native American ancestry was negatively associated with large deletions in the FVIII gene (Log-OR=-0.081, 95%CI: -0.144,-0.018, p=0.011). In most Latin American populations, sex-ancestry bias is observed: during the last 5 centuries of admixture, European ancestry has been preferentially mediated by males and African or Native American ancestries preferentially by females. Considering that hemophilia is an X-linked disease, we estimated X-chromosome ancestry to confirm the pervasive sex-ancestry bias in the BrazIT cohort. As expected, BrazIT individuals show a strong Native American ancestry bias (Autosome - X-chromosome: -44%, p<0.001), but exceptionally, no African ancestry bias (-0.006, NS). This result suggests that X-chromosomes of African origins may be under-represented among severe

hemophilia A patients with high-response inhibitors, exemplifying the unexplored issue of how sex-bias in Post-Columbian migrations from Europe and African to the Americas have differently shaped the genetic diversity of X-chromosome and autosomal linked mendelian diseases.
Funding: Brazilian Ministry of Health, FAPEMIG, CNPq, CAPES.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3005F: Ancient DNA provides evidence for related selection pressures at *ABO* and *FUT2* loci over the past 10,000 years in West Eurasia ⭐

Authors:

A. Barton[1,2,3], A. Akbari[2,1,3], D. Reich[2,1,3,4]; [1]Harvard Univ., Cambridge, MA, [2]Harvard Med. Sch., Boston, MA, [3]Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Howard Hughes Med. Inst., Boston, MA

Abstract:

Blood group traits represent an ancient source of variation in humans that has persisted since our great ape ancestors. While their distribution and possible role in infection has long been noted, how selection has shaped their frequencies in the recent past remains understudied. Using 8,505 ancient samples and a generalized linear mixed model approach, we detect evidence of selection on the *ABO* gene over the past 10,000 years in West Eurasia in addition to selection at *FUT2*, a gene with epistatic interactions with *ABO*, that is suggestive of their correlated selection.
Prior to 10,000 years before present (BP), the allele for the O blood type and the allele for secretor status at the *FUT2* gene were predominant across West Eurasia. However, starting around 10000 BP, the A allele (rs8176719) began to increase in frequency, and shortly after this, the nonsecretor variant for *FUT2* (rs601338) began to increase in frequency due to selection ($s=0.8\%$, $P=2x10^{-26}$). This timing suggests there may have been negative fitness consequences to the secretion of the A antigen into tissues other than the blood such as the saliva or intestinal mucosal, possibly due to its recognition as a receptor by certain pathogens. This is supported by the trajectory of the phenotypes at these loci as the A secretor phenotype began to decrease in frequency around 5700 BP as the A nonsecretor phenotype continued to significantly ($Z > 3$) increase in frequency prior to the arrival of the B allele to most of West Eurasia. Finally around 4000 BP with the arrival of Steppe populations to the region, the B allele (rs8176746) began to experience significant positive selection ($s=2.7\%$, $P=7x10^{-46}$) that coincided with continued selection for nonsecretor status and an observed increase in both the B secretor and

nonsecretor phenotypes but a decrease in all A phenotypes.
The environmental pressures behind this selection are likely related to those that have resulted in long term balancing selection on the *ABO* alleles. We observe significant associations of these alleles with a number of traits in the UK Biobank, including alkaline phosphatase which shows differential relationships with each blood type based on secretor status (A secretor β=-0.16, P=$2x10^{-12}$, A nonsecretor -0.14, $2x10^{-9}$, B secretor 0.01, 0.6, B nonsecretor 0.13, $2x10^{-7}$; O secretor 0.08, $7x10^{-4}$, O nonsecretor 0.21, $9x10^{-21}$), suggesting an interaction between these genes. Finally, other blood group loci, including at *GYPA* (s=1.0%, P=$4x10^{-56}$) and *FUT3* (s=0.8%, P=$9x10^{-28}$) show significant evidence of selection in West Eurasia, and further research is needed to elucidate the suite of selective pressures including through investigation into ancient pathogens.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3008F: Characterizing and elucidating the mechanistic basis of context-dependent CpG mutation rate variation

Authors:

S. Chandra, Z. Gao; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

The rate of mutation varies substantially across the human genome. Much of this rate variation is explained by the flanking nucleotides surrounding the polymorphic site (the sequence context). An extreme example of such context-dependency is CpG sites. Cytosines at CpGs exhibit a mutation rate substantially higher than the genome-wide average, and this hypermutability is related to methylation. CpG sites in different sequence contexts themselves exhibit mutation rate variation, but this variation is partially correlated with methylation. In this study, we aim to elucidate how mutation rates at CpG sites vary across contexts and at different methylation states. We also probe the mechanistic basis of this variation by testing the role of different genomic features in the interaction between context and methylation. To quantify the variation in mutation rates at CpG sites, we used polymorphisms from the 1000 Genomes Project as a proxy for mutations. We found that polymorphism probability increases with germline methylation level linearly until ~50% methylation, after which the slope starts to decrease, reflecting the effect of recurrent mutations. This indicates that direct usage of polymorphism data underestimates the mutation rate at hypermutable sites. To account for this effect of repeat mutations, we developed a regression framework that relates observed polymorphism to mutation rate based on a coalescent model. Using this regression framework,

we found that mutation rates at unmethylated and methylated CpGs are both significantly different across contexts and in distinct orders. We also observed largely independent effects of the upstream and downstream bases flanking the CpG. These findings indicate that methyl group and flanking sequence interact to contribute to mutation rate variation. Unexpectedly, we also found strand asymmetry between the two cytosines within the same methylated CpG: the 5mC flanked by adenine on the same strand is more prone to mutation than the 5mC on the opposite strand. We tested two potential mechanisms for the context-dependent mutation rate variation at CpG sites: binding by transcription factors (TF) and demethylation by ten eleven translocation (TET) enzymes. Overall, this study shows that flanking sequences and methylation at the focal cytosine interact in driving mutation rate variation across CpG sites, and these interactions are partially shaped by *trans*- factors such as TF binding and TET activity.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3009F: Cross-comparison of genomic approaches for ancestry estimation

Authors:

Q. Wang, Y. Wu, F. R. Schumacher; Case Western Reserve Univ., Cleveland, OH

Abstract:

Population stratification is a primary consideration in genetic research where failure to adjust can lead to the detection of false positive or false negative associations. Population stratification arising from geographic isolation, interbreeding, and adaptation will yield allelic frequency differences by ancestral groups. The reliance of self-reported ancestry, which can be inaccurate due to uncertainty or limited to broad categories, may introduce additional bias. Germline variation based on single nucleotide polymorphisms (SNPs) effectively captures ancestry information, however, this data may be unavailable. Although other whole-genome -*omics* may be available for quantifying ancestry, the robustness of these approaches remains unknown. Our study contrasts ancestry estimates based on three data types - self-reported data, whole-genome SNP, and whole-genome methylation data. Our aim is to evaluate the accuracy and reliability of these estimates, providing insights into their strengths and limitations.
Our study cohort of 864 individuals has self-reported ancestry and genome-wide SNP data, including a subset of 229 individuals with genome-wide DNA methylation data. Ancestry estimates the SNP data were calculated using the ADMIXTURE algorithm to estimate the proportions of European, Native American, Asian, and African ancestry. For the subset of 229 individuals with whole-genome methylation data we used the EPISTRUCTURE algorithm, which

utilizes principal component analysis (PCA) of methylation CpG sites highly correlated with nearby SNPs to infer ancestry proportions. These ancestry derived estimates based on SNP and methylation data will be compared, and contrasted with self-reported race, to identify consistencies and discrepancies. Correlation analysis and ANOVA will be conducted to evaluate the significant differences between each approach. Pairwise comparisons to identify specific differences between ancestry estimation methods will be performed and visualized with Bland-Altman plots. Intra-class correlation coefficients (ICC) will be calculated to quantify the consistency of ancestry estimates across data types. Visual tools such as PCA plots, bar charts, and heatmaps will be employed to illustrate the comparisons.
By contrasting these three approaches, our study will comprehensively assess the accuracy and reliability of self-reported, genetic, and epigenetic approaches in ancestry estimation and addresses a current scientific gap.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3010F: Endogamy and high prevalence of deleterious mutations in India: evidence from strong founder events ⭐

Authors:

P. Machha[1,2], D. Sowpati[1,2], K. Thangaraj[1,2]; [1]Ctr. for Cellular and Molecular Biology (CCMB), Hyderabad, India, [2]Academy of Scientific and Innovative Res. (AcSIR), Ghaziabad, India

Abstract:

Background/Objectives: India is a country of immense diversity, comprising more than 4500 anthropologically distinct groups, who strictly follow endogamy. Our earlier study showed that several Indian populations have experienced founder events stronger than the archetypal Finns and Ashkenazi Jews, groups known for a high prevalence of recessive disorders. Here, we selected four populations (Yadavs, Kallars, Kalingas and Reddys) with high IBD scores to explore population-specific recessive mutations, which might be responsible for various health conditions and disease.
Methods: We performed whole exome sequencing for 281 individuals and SNP-array genotyping of 96 samples across these 4 groups.
Results: Firstly, we observed a high rate of inbreeding in the four populations, with 59% of our samples showing parental relationships as first cousins, second cousins, or double-first cousins. We confirmed the presence of both founder events and inbreeding in these groups through the identification of both short and long stretches of runs of homozygosity (ROH) in almost similar measures, respectively. Upon comparison with existing population datasets, we discovered

1,284 unique novel exonic variants across the groups, highlighting the genetic underrepresentation of Indian populations. Of them, 23 were predicted to be deleterious, all occurring in heterozygous condition, and majority of them are inferred to impact the protein stability, thereby, emphasizing their potential role in disease biology. Notably, 40-68% of the identified deleterious variants (reported in ClinVar) showed significantly higher allele frequencies ($p<0.05$) than the expected. Additionally, our pharmacogenomic analysis revealed the occurrence of several actionable variants from the Pharmacogenomics Knowledgebase (PharmGKB) across all four groups. Further, demonstrating the impact of founder events and inbreeding on disease susceptibility, we report a higher prevalence of ankylosing spondylitis in the Reddys, linked to a higher frequency of HLA-B*27:04.
Conclusion: Our findings underscore the need for a comprehensive understanding of the impacts of founder events and the subsequent prevalence of inbreeding on disease susceptibility and medical risk profiles, aiming to reduce disease burden in understudied population groups.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3011F: Evaluating Genetic Diversity in the All of Us Research Program

Authors:

M. Gouveia[1], V. Borda[2], T. Peixoto Leal[3], C. Fortes-Lima[4], I. Mata[3], C. Schlebusch[5], T. O'Connor[6], E. Tarazona-Santos[7], A. Adeyemo[1], C. Rotimi[1], D. Shriner[1]; [1]NIH, Bethesda, MD, [2]Th Univ. of Maryland Sch. of Med., Baltimore, MD, [3]Lerner Res. Inst., Cleveland, OH, [4]Johns Hopkins Univ., Baltimore, MD, [5]Uppsala Univ., Uppsala, Sweden, [6]The Univ. of Maryland Sch. of Med., Baltimore, MD, [7]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract:

The All of Us Research Program (AoU) aims to accelerate biomedical research and address minority underrepresentation by enrolling over a million multi-ethnic individuals across the USA. A recent genetic analysis reported that Uniform Manifold Approximation and Projection (UMAP) recapitulated known patterns of population structure and that AoU participants clustered within discrete, self-identified race and ethnicity categories, sparking discussion on the validity and interpretability of these results. To contextualize the genetic diversity sampled by the program, we analyzed approximately 2 million common genome-wide variants in the whole-genome sequence data of 245,388 participants, along with public reference data from the 1000 Genomes Project (1000G), the Human Genetic Diversity Project (HGDP), and our

recently published panel of 79 European populations. Unsupervised Principal Component Analysis (PCA) yielded five significant principal components (PCs) with genome-wide variant loadings and statistically large eigenvalues. After projecting 1000G and HGDP onto these PCs, we observed that the genetic diversity reflected in the AoU participants recapitulated and exceeded the genetic diversity reflected in the external reference panels. AoU "African Americans" were distributed across the top PC. AoU "Hispanic or Latino" participants were distributed across the top two PCs, simultaneously representing two gradients of African *vs*. European diversity (PC 1) and European *vs*. Native American diversity (PC 2). We found multiple clusters encompassing Native American and Asian diversity (PCs 3 and 5). We also found that AoU self-identified "whites" covered most North-to-South European genetic diversity (~90%) of our European panel as well as Middle Eastern genetic diversity (PC 4). Moreover, analysis of the same data with UMAP produced artefactual visualization and lacked the strong interpretability of PCA, particularly in assessing admixture. In summary, the genetic diversity in the All of Us Research Program reflects known continental and subcontinental patterns, as well as genetic variation not identifiable through self-report or existing 1000G/HGDP and European diversity reference panels. Self-reported "Hispanic or Latino" and "African American" participants show greater genetic diversity than their 1000G counterparts, prompting further analyses to assess Native American and African subcontinental diversity in AoU. Finally, our results based on multi-ancestry individuals indicate that PCA is recommended over UMAP for assessing population structure and admixture in human populations.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3013F: Evaluating the evolutionary mismatch hypothesis in the electronic health record: A phenome and lab-wide investigation of known malaria variants in >35k individuals of primarily African ancestry in the Vanderbilt University biobank

Authors:

F. Blostein[1], P. Straub[1], N. J. Cox[2], J. Hirbo[3]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Univ Med Ctr., Nashville, TN, [3]Vanderbilt Univ. Sch. of Med., Nashville, TN

Abstract:

Infectious diseases such as malaria were strong selective pressures for much of human evolution, leaving signatures of genomic change in the human genome. Under the evolutionary mismatch hypothesis, genetic variants beneficial in high-pathogen environments may contribute

to disease phenotypes in low-pathogen environments. Since pathogen incidence varies across geographic locales, these “mismatched” genetic variants vary in frequency across genetic ancestry clines and may therefore be crucial for equitable disease or biomarker modeling. We investigated the evolutionary mismatch hypothesis by testing genetic variants related to malaria susceptibility and severity against phenome-wide medical conditions and laboratory traits in the electronic health record. Thirteen genetic variants previously reported to associate with malaria incidence/severity or signature of resistance to malaria parasite were identified from a literature review, extracted from whole genome sequencing data for 35,024 primarily African ancestry individuals, and tested for association with medical conditions derived from ICD codes and common quantitative laboratory tests using parallel logistic and linear regressions, respectively, while controlling for sex, median age across medical record, and the first ten principal components of genetic ancestry. The malaria variants were associated with 154 phenotypes after Bonferroni correction across phenotypes and variants. Variant-phenotype associations were enriched for associations with hematopoietic, genitourinary, and circulatory system phenotypes. As expected, rs2814778 (*ACKR1,* Duffy antigen) was strongly associated with white blood cell disorders and rs334 (*HbS,* sickle cell) was strongly associated with anemias. rs8176719 (*ABO*) and rs1050828 (*G6PD*) were associated with diabetes. At FDR-correct *p*-value<0.05, rs17047660 (*CR1*) was associated with tension headache and rs8176746 (*ABO*) with venous embolism. There were 193 Bonferroni significant variant-laboratory trait associations representing 6 malaria variants, with enrichment in labs related to blood, immune, and lung related traits. In future analyses, we plan to test if incorporating these variants can improve genetic prediction of blood traits and yield more equitable blood trait biomarker reference ranges. Understanding how malaria variants impact the medical phenome can illustrate mechanisms of action and help identify and explain gene by environmental interactions. Moreover, incorporating information on such variants into disease and biomarker modeling may improve health outcomes and help create more equitable biomarker ranges.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3014F: Evaluation of imputation methods for ancient African DNA

Authors:

R. Pornmongkolsuk, A. Ochoa; Duke Univ., Durham, NC

Abstract:

Ancient DNA provides unique insights into the ancestral genetic variation; this is particularly useful in the study of adaptations with which pre-selective sweep haplotypes are not available

in the modern day populations. However, DNA degradation in ancient specimens often results in low-coverage sequences which impacts downstream analyses, calling for an imputation of missing genotypes. As most imputation tools require high-quality reference haplotypes which are often obtained from modern genomes, imputing ancient haplotypes can be more challenging than modern sequences. Additionally, due to the unique demographic history of African populations that had allowed for high genetic diversity and low linkage disequilibrium, African sequences are especially difficult to impute. Here we show that GLIMPSE2, an imputation tool designed for low-coverage sequencing data, has better performance than Minimac4 and Beagle on ancient African sequences. We used the pseudo-haploid called variants, of which all loci are homozygous variants due to low coverage, that are publicly available at the Allen Ancient DNA Resource (AADR) as inputs to the other tools, but GLIMPSE2 is capable of imputing diploid genomes with high confidence, which is a distinct advantage, for which we downloaded aligned sequencing data (BAM files) from various source publications. We focus on the Duffy null polymorphism on the *DARC* gene (chr1:rs2814778) - a well-known adaptive variant protective against vivax malaria - to study its evolutionary history in ancient African populations. The AADR includes 171 ancient African genomes which cover the geographical and temporal spans relevant to the Duffy-null sweep in Africa. To evaluate these imputation tools, we performed cross-validation using, of the total 171 samples, 41 ancient African genomes of which Duffy genotypes are known and masked, and compared the PCA plots. The cross-validation error rates of Minimac4 and Beagle are dramatically higher than GLIMPSE2. We also benchmarked their performance using simulation of ancient DNA with varying coverages and levels of contamination. Overall, we find that fully utilizing the information in sequencing reads allows for improved imputation in the notoriously hard datasets of ancient African ancestry.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3015F: Exploring structural variants and adaptation in the Andean highland and coastal populations using long-read sequencing

Authors:

S. Roy[1], V. Borda Pua[2], M. Gouveia[3], R. McCoy[4], S. Yan[4], K. Cifuentes[1], C. Sanchez[5], C. Padilla[6], H. Guio[7], M. Santolalla[8], R. Juscamaita[8], C. Rotimi[9], O. Caceres[5], E. Tarazona-Santos[10], T. O'Connor[2]; [1]Univ. of Maryland, Baltimore, MD, [2]Univ. of Maryland Sch. of Med., Baltimore, MD, [3]NIH, Bethesda, MD, [4]Johns Hopkins Univ., Baltimore, MD, [5]3 Laboratorio de Innovación y Desarrollo, Inst. Natl. de Salud, Lima, Peru, [67]INBIOMEDIC, Lima 14, Peru, [8]Univ. Peruana

Cayetano Heredia, Lima, Peru, [9]NIH, Woodstock, MD, [10]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract:

The migration of human populations from the lowlands or coastal areas of South America to the high-altitude Andean regions began 10,000 to 12,000 years ago. These early settlers had to adapt to hostile high-altitude environments, which included lower oxygen levels and colder temperatures compared to the hot and humid conditions of the equatorial coasts. While physiological adaptations to such environments are well documented, the genetic basis for these adaptations is still not well comprehended, particularly beyond single nucleotide variants (SNVs). Structural variants (SVs ≥ 30 bp) are thought to play a crucial role in genome evolution and adaptation, yet remain largely uncharted.To better understand their role in genomic adaptation, we generated high-coverage, high-depth long-read sequence data on two Peruvian individuals with predominant Indigenous American ancestry (>99%) using PacBio HiFi sequencing. We integrated our data with 34 publicly available PacBio HiFi datasets from the Human Genome Structural Variation Consortium to construct an SV panel containing 200,341 SVs. We then employed a graph-genotyping method, PARAGRAPH, to apply our SV panel to short-read data from the Peruvian Genome Project, Human Genome Diversity Project, and 1000 Genomes Project, including three highland and two lowland populations, resulting in a set of 110,588 SVs.We employed population branch statistics (PBS) to measure the genetic divergence between the highland and coastal populations, with a West African population, Yoruba as an outgroup. This analysis identified two highly differentiated novel variants as candidates for highland adaptation (PBS value > 99th percentile). Both exhibit strong allele frequency (AF) differences between highland and coastal populations. The first variant is a 132 bp deletion spanning the first intron of *LINC01725* with an AF of 0.7 in the highlands and 0.2 in the lowlands. The second is a 344 bp deletion spanning the first intron of *TSGA10* with an AF of 0.86 in the highlands and 0.4 on the coast. These genes are implicated in modulating reticulocyte count, markers for hemolytic anemia, higher oxygen demand, and hemoglobin levels, typical consequences of highland-related hypoxia. We hypothesize these variants have enabled the Andean population to adapt to extreme hypoxic conditions. Our approach to exploring structural variation provides new directions to discover genetic variants under natural selection in harsh environments during the peopling of the Americas. These results underscore the importance of gene-environment interactions in population genetics and provide insights into the role of SVs in evolutionary genetics.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3016F: Exploring the demographic histories of rainforest hunter-gatherers and neighboring Bantu-speaking agriculturalists as well as the genomic basis of short stature in Africa ⭐

Authors:

A. Lisi[1,2], J. X. Liu[1], C. Battaggia[2], G. Destro Bisol[2], P. Anagnostou[2], M. C. Campbell[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]La Sapienza Univ. di Roma, Rome, Italy

Abstract:

West Central and Central Africa, which are characterized by vast rainforests, have been regions of anthropological interest due to the long-term presence of rainforest hunter-gatherers (RHG) and the more recent arrival of neighboring Bantu-speaking agriculturalists (BSA). Furthermore, BSA also expanded more broadly to eastern, central, and southern Africa, leading to a spread of Bantu languages across the continent. While many African populations transitioned to an agricultural lifestyle after their contact with these expanding BSA, other populations, such as the western RHG, maintained a hunting and gathering lifestyle. Given the paucity of fossil remains in West Central and Central Africa, information regarding the history of RHG and BSA populations within these regions has been limited. With recent advances in computational and molecular methods, we now have a unique opportunity to explore new and long-standing hypotheses regarding the demographic events that have shaped genomic diversity within and between populations.To reconstruct the past demographic history of RHG and BSA, we performed genomic genotyping in 144 western RHG (Baka, Bakola and Mbenzele) and BSA (Bassa, Ewondo and Ngoumba) from Cameroon and the Central African Republic. We further integrated these genotypes with publicly available genomic data from >1,400 individuals from 59 African populations, including other hunter-gatherer populations across the continent. Our data showed that the ancestors of Mbenzele RHG diverged from other western hunter-gatherers ~10,000 years ago (ya), well before the expansion of BSA across Africa beginning ~4,000 ya. Although the Mbenzele experienced a bottleneck event after their divergence, this population still maintained a larger effective population size compared to the other western RHG. Moreover, we uncovered a striking pattern of population structure among western RHG and differential gene flow between the ancestors of western RHG and BSA at different points in time. In addition, we examined the genetic architecture of short stature in RHG using a genetic association approach, focusing on western RHG with low levels of admixture from BSA. Based on this analysis, we identified several candidate SNPs for height on chromosomes 9 and 19 and observed unusually long haplotypes around these polymorphisms consistent with a selective sweep model. Overall, this study provides a more nuanced perspective on the genetic history of

RHG and BSA as well as new insights into the genetic architecture underlying short stature in rainforest hunter-gatherer populations.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3017F: Familial long-read genome sequencing identifies early embryonic new mutations and an accelerated SNV mutation rate in high identity segmental duplications.

Authors:

M. Noyes[1], W. T. Harvey[1], Y. Sui[1], K. M. Munson[1], K. Hoekzema[1], J. Kordosky[1], G. Garcia[1], J. Knuth[1], A. P. Lewis[1], M. C. Zody[2], E. E. Eichler[1,3]; [1]Univ. of Washington, Seattle, WA, [2]New York Genome Ctr., New York, NY, [3]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA

Abstract:

De novo mutations (DNMs) play an important role in human evolution and disease yet have been challenging to identify in repetitive regions. Typical studies of DNM use Illumina data, which are limited by their short length in their ability to both provide parent-of-origin information and align to the most repetitive regions of the genome, where we anticipate the mutation rate is higher. We used PacBio HiFi, Oxford Nanopore Technologies (ONT), and Illumina to sequence 90 individuals from 24 families (19 quads and 6 trios, n=42 transmissions) affected with simplex autism. DNMs were first identified in HiFi read data derived from peripheral blood, then validated by confirming their presence in a sample's ONT and Illumina reads and their absence in all parental read data. In total, we found 3,505 *de novo* SNVs on the autosomes, 32.3% (n= 1,131) of which had not been described by previous Illumina studies of the same samples. We were able to determine the parental haplotype of origin for 97.7% of all DNMs by examining a sample's HiFi and ONT data for SNPs that could be uniquely attributed to one parent. Long-read phasing data also allowed us to unambiguously discern germline (n=2,953) from postzygotic (n=552) DNMs that arose after fertilization by determining whether a *de novo* variant was present on all reads from a parental haplotype. Germline DNMs are enriched on paternal haplotypes, with a 3.85:1 paternal:maternal ratio, while postzygotic DNMs have a 1.07:1 paternal:maternal ratio, confirming their origins later in development. Additionally, these two classes of DNMs have significantly different mutational spectra, with postzygotic mutations enriched for A>C and A>T substitutions. Based on the depth of high-quality HiFi reads at every genomic coordinate, we estimate that we saturated germline DNM discovery in 92% of the autosomes. This allows us to calculate germline and postzygotic mutation rates of 1.30×10^{-8} and 2.38×10^{-9} substitutions/base pair/generation, respectively. Germline mutations are enriched

38.4% in segmental duplications (SDs) (t-test, p=0.0103), with a mutation rate of 1.80×10^{-8} substitutions/base pair/generation in SDs. There is a more extreme 3-fold enrichment of PZMs in centromeres (t-test, p=0.0310) and in SDs (t-test, p= 0.0000389). For both germline and postzygotic mutations, this signal is driven by SDs with >99% identity, which have a mutation rate more than double that of the lowest identity duplications, supporting our hypothesis that repetitive regions of the genome are hypermutable compared to unique regions.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3018F: Genealogy-informed variant filtering and genotype calling on large scale human Y chromosome call sets

Authors:

Y. Si, S. Zöllner, H. Kang, NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium; Univ. of Michigan, Ann Arbor, MI

Abstract:

Human Y chromosome has unique genetic properties due to its male specificity and lack of recombination. The genetic variation on the Y chromosome is therefore key to study human population history and sex-related diseases. In principle, due to the lack of recombination or heterozygotes in Y chromosomes, identifying true genetic variants and calling their genotypes should be more straightforward than autosomes. However, because of the Y chromosome's complex repetitive structure and lower depth, typical variant call sets contain many false positive variants and incorrect genotypes that do not fit to the genealogy. As short read sequencing data will remain the main source to study the genetic variation in the population until long read sequencing dominates the field, it is important to produce high quality variant call sets for the Y chromosome through more rigorous and chromosome Y-specific variant filtering. We propose a new approach that leverages the shared genealogy in the haploid non-recombinant region to improve the variant calling and genotyping on Y chromosomes in large sequencing samples. From germline short variant call sets for the 1000 Genome Project deep sequencing data and a subset of the TOPMed data, we evaluate our approach using variant quality metrics, within-family consistencies, and population genetics metrics. We demonstrate the improvement of the genotype quality over existing variant discovery pipelines that do not take into account the special properties of sex chromosomes. As an application, we construct the genealogy from thousands of deeply sequenced Y chromosomes, which provides much more detailed information of human evolution than previous manually curated haplogroups based on limited samples and a small number of SNPs. We expect that these high quality Y

chromosome call sets will become the basis of various chromosome Y-specific downstream analyses of disease association, male specific demographic history, germline mutation process and natural selection, and other human genetics questions.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3019F: Genetic Prevalence of PEX1, PEX3, PEX5, and PEX6 Variants in Nephrolithiasis and Nephrocalcinosis within Zellweger Spectrum Disorders

Authors:

Y-R. Huang[1,2], K. Lovrenert[1,2], J. Weaver[3], D. Bodner[2], F. Schumacher[1], C-H. Wu[1,2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. Hosp., Cleveland, OH, [3]Cleveland Clinic, Cleveland, OH

Abstract:

Peroxisome is an organelle in the cell responsible for the metabolism of glyoxylate, a major precursor for hepatic oxalate. Deficiency in alanine-glyoxylate aminotransferase encoded by AGXT, results in the overproduction of oxalate by the liver, leading to the formation of nephrolithiasis. Previous research reported high incidence of hyperoxaluria in generalized peroxisomal disorder, and nephrolithiasis and nephrocalcinosis were reported in Zellweger spectrum disorder (ZSD) patients with PEX1, PEX3, PEX5 and PEX6 pathogenic variants. We aim to estimate the genetic prevalence of ZSD associated with nephrolithiasis and nephrocalcinosis through these four genes.

The 1000 Genomes Project (1KG) was utilized to identify genetic variants in the general population. Clinvar was used to identify pathogenic variants of PEX1, PEX3, PEX5 and PEX6. Overlapping variants from both databases were evaluated. The pathogenic allele frequency, carrier rate, and affected rate were calculated and estimated based on Hardy-Weinberg equilibrium.

The 1KG Project contained 874, 820, 942 and 392 unique variants for PEX1, PEX3, PEX5 and PEX6, respectively. Clinvar contained 296, 18, 40 and 185 unique pathogenic variants of PEX1, PEX3, PEX5 and PEX6 respectively. The intersection between Clinvar and 1KG resulted in three pathogenic variants of PEX1 across four heterozygous carriers (pathogenic variant allele frequency of 0.08%) and no pathogenic variants of PEX3, PEX5 and PEX6. The expected carrier rates were 0 for PEX3, PEX5 and PEX6, and 1 in 625 for PEX1 and for the cumulative rate combining the four genes. The expected affected rates were 0 for PEX3, PEX5 and PEX6, 1 in 1,567,398 for PEX1 and for the cumulative rate combining the four genes.

Our results reveal 4 PEX1 heterozygous carriers and 0 PEX3, PEX5 and PEX6 carriers in the

healthy, unrelated 1KG population. ZSD is the most common peroxisomal disorder with an overall incidence of 1 in 50,000 to 100,000 live births. Clinically, 58.9% of ZSD is caused by pathogenic variants of PEX1, 15.9% by PEX6, 1.7% by PEX5, and 0.5% by PEX3. These four genes account for 77% of ZSD with a clinical prevalence estimate of 1 in 64,935 to 129,870, which is 12-24x higher than our genetic estimate. There is a gap between the clinical observation and genetic estimate, which requires persistent research endeavors to further explore the phenomenon.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3021F: Inference of recent population histories by combining genetic and archeological data

Authors:

G. Kennedy[1], S. Gopalan[2], A. Goldberg[1]; [1]Duke Univ., Durham, NC, [2]Clemson Univ., Clemson, SC

Abstract:

Fine-scale temporal resolution of human population size over time informs our understanding of the causes and consequences of population growth. A variety of methods to reconstruct population size over time have been developed independently across fields such as genomics and archeology, each with their own benefits and limitations. For example, interpretation of population size from genomic data generally has poor temporal resolution and accuracy on recent timescales, and may be difficult to interpret as direct census size. In contrast, archeological methods to infer population sizes—which often use the distribution of radiocarbon dates as a proxy for human census size under the general assumption that more people leave more material behind—have fine-scale spatial and temporal resolution. However, these archeological methods are sensitive to taphonomic and sampling biases, and have difficulty differentiating between endogenous growth versus migration. Here, we present an Approximate Bayesian Computation framework to infer human population size over time using genomic and archeological data simultaneously to leverage their respective strengths. We simulate genetic data in SLiM and and radiocarbon data rcarbon, and calculate summary statistics for both types of data to perform both model choice and parameter inference. We compare performance of each data type alone versus the combined archeological and genetic inference under for models of growth: exponential, linear, logistic, and constant growth. Preliminary tests suggest that logistic growth parameters are better captured by archeological data, while exponential and linear growth can be well-estimated from both types of data. This model framework shows potential to improve our understanding recent human history by

combining the population specificity of genetic data with the temporal resolution of archeological data. We apply this method to test the effect of agricultural developments on population growth in the South American Andes.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3022F: Insights into copy-number evolution and polygenic score transferability from *LPA* ⭐

Authors:

R. Mukamel[1], S. Rubinacci[2], N. Kamitaki[3], R. Handsaker[4], M. Hujoel[5], D. Tang[6], S. McCarroll[1], P-R. Loh[7]; [1]Harvard Med. Sch., Boston, MA, [2]Brigham and Women s Hosp., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA, [4]Broad Inst., Cambridge, MA, [5]Brigham & Women s Hosp./Harvard Med. Sch., Boston, MA, [6]Harvard Univ., Boston, MA, [7]Brigham and Women s Hosp., Boston, MA

Abstract:

The 5.5kb kringle IV-2 (KIV-2) repeat in *LPA* is one of the largest and most size-polymorphic repeats in the human genome. Here, we analyzed whole-genome sequencing (WGS) data from >700,000 participants in UK Biobank (UKB) and *All of Us* to learn about the mutational processes that have shaped KIV-2 alleles and the way in which sequence variation in *LPA* determines serum lipoprotein(a) (Lp(a)) concentration, a major cardiovascular risk factor. We identified 5 common KIV-2 haplogroups and dated their divergence times by counting mutations flanking KIV-2. The deepest divergence (~2 Mya) partitions KIV-2 alleles into two groups that have rarely recombined throughout human history (0.03% frequency of recombined alleles in UKB), consistent with KIV-2 being a recombination desert (0.003 cM/Mb). Five paralogous sequence variants (PSVs) within each repeat unit of each KIV-2 allele segregate nearly perfectly with these two groups. By analyzing allele-specific WGS read-depth, we accurately estimated KIV-2 allele lengths in UKB (RMSE=0.4 repeat units) and characterized the mutational dynamics of the KIV-2 repeat. KIV-2 copy-number mutations occur at a rate of 1 per ~500 allele transmissions—three orders of magnitude more frequently than recombination within KIV-2—implicating sister chromatid exchange (SCE), rather than meiotic recombination, as the dominant mechanism of KIV-2 length mutation. Most mutations modify KIV-2 copy number by 1 or 2 repeat units, suggesting that asymmetric exchanges during DNA replication tend to utilize homologous sequence nearby on the nascent sister chromatid.
Analyzing KIV-2 lengths together with *LPA* sequence variation showed that all genetic ancestries in UKB share a common genetic architecture for Lp(a) driven by coding and splice

polymorphisms. Genetic predictions incorporating KIV-2 length and the relative haplotype phase of 27 fine-mapped Lp(a)-modifying SNPs explained >80% of heritable variation in log-transformed Lp(a) within European, African, East Asian, and South Asian ancestries. Six coding variants that inactivated short, usually-pathogenic KIV-2 alleles were commonly found in non-European ancestries (AF=0.1-1%) but were too rare for their effects to be detectable in European-ancestry data, underscoring the importance of diversity in genetic analyses. Applying the genetic predictor to *All of Us*, high predicted Lp(a) (>300 nmol/L) associated with consistent increases in myocardial infarction risk across African and European genetic ancestries (OR = 1.53 (1.11-2.11) and 1.70 (1.34-2.15) relative to predicted Lp(a)<30 nmol/L), demonstrating its potential value for genomic medicine.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3023F: Integrated haplotype score IHS and signals of positive selection in suicide associated SNPs.

Authors:

E. Pranckeviciene[1,2], J. Kasperiuniene, Dr.[2]; [1]Vilnius Univ., Vilnius, Lithuania, [2]Vytautas Magnus Univ., Kaunas, Lithuania

Abstract:

Suicide trait varies across countries from very high to almost non existent. Suicide mortality rate (SMR) world development indicator(WDI) data published by world bank over a period of time (2000-2019) for each country shows that Eastern and Northern European countries have highest SMR while Middle Eastern , North African and Caribbean countries have lowest. We investigated which of the 1000 Genome Project (1KGP) populations have high frequency of suicide alleles in 11 SNPs taken together that were identified by GWAS. We noticed a correspondence between populations with high/low suicide associated alleles frequency and high/low SMR in countries from which those populations come from.
The integrated haplotype score (IHS) was high for several 1KGP populations in the database of human recent positive selection signals (RePoS) for the three of the suicide associated SNPs: rs336284, rs3741042 and rs6913400. A high negative IHS of a derived allele relative to ancestral shows a positively selected sweep which means that a genetic variant rapidly increases in frequency in population. Similarly, a high positive IHS informs the same about ancestral allele.
A suicide associated allele A in the rs336284 A>G (mapped to genes *TBX20, HERPUD2*) is a derived allele. A high negative IHS (< -2) for this SNP is in AFR ESN (Esan in Nigeria), AFR GWD (Gambian in Western Division, The Gambia) and SAS ITU (Indian Telugu in the UK) and SAS PJL

(Punjabi in Lahore, Pakistan) populations.
In rs3741042 C>T (mapped to genes *LYVE1, IRAG1*) a predisposing allele is C which is a derived allele. In this SNP a high negative IHS (< -2) marks AFR GWD (Gambian in Western Division, The Gambia), AFR LWK (Luhya in Webuye, Kenya), AFR YRI (Yoruba in Ibadan, Nigeria) populations while high positive IHS (> 2) is characteristic for EAS CHB (Han Chinese in Bejing, China), EAS CHS (Southern Han Chinese) and EAS KHV (Kinh in Ho Chi Minh City, Vietnam).
A predisposing allele in the rs6913400 G>C,T (mapped to *GRM1* gene) is C which is a derived allele. In this SNP a high positive IHS (>2) is characteristic for AMR PUR (Puerto Rican from Puerto Rica), EAS CDX (Chinese Dai in Xishuanagbanna, China) and EAS CHB (Han Chinese in Bejing, China) populations.
Taken together in 1KGP African and several South Asian populations the positive selection signals show direction towards favoring the suicide associated alleles while in East Asian Chinese and Vietnamese populations selection signals indicate direction against the suicide associated alleles. The average statistical suicide mortality rate per 100,000 population in 2000-2019 in African countries is fluctuating around 7 while in China it is close to 12 and steadily decreasing.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3024F: Long-read sequencing on diverse populations identifies an association of a haplotype with two CAA interruptions in the CAG trinucleotide repeats of the ATXN2 gene

Authors:

B. Lee[1], K. Wang[2], D. Amado[1], Y. Song[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]CHOP, Philadelphia, PA

Abstract:

The CAG repeat expansion in exon 1 of the ATXN2 gene is associated with multiple diseases, including amyotrophic lateral sclerosis (ALS), spinocerebellar ataxia 2 (SCA2), and Parkinson's disease (PD). One of the suggested mechanisms that separates susceptibility to PD versus SCA2 is the presence of CAA interruptions in the expanded CAG repeats. CAA interruption disturbs repeat-associated non-AUG translation that results in translation of poly-glutamine proteins. While the effect of the presence of the CAA interruption has been studied, the effect of the number of CAA interruptions is not well understood. Using publicly available long-read sequencing datasets from 1KG ONT Vienna Consortium and 1KGP ONT Sequencing Consortium, we examined the number of CAA interruptions in healthy control data (n = 1728 alleles). We

found that 36%, 63% and 1% of alleles had one, two and three CAA interruptions, respectively. Frequency of two CAA interruptions was highest in the South Asian population (67.16%), while it was the lowest in the East Asian population (56.77%). Surprisingly, more than half of alleles that had three CAA interruptions had the expanded number of polyQ (> 27). We also examined two previously described SNPs (rs695871 G > C, rs695872, G > A, haplotype = CA) that were shown to be correlated with 3 CAA interruptions and expanded CAG repeats in SCA2 (haplotype = GG). We found that the frequency of these alternate SNPs is higher for two and three CAA interruptions when compared to one CAA interruption (odd ratio = 3.32 for CA), and had higher frequencies in European, South Asian and American populations when compared to East Asian and African populations. We also found SNPs in chr12: 111582410 (G > C), 111585197(A > G), 111588597(A > G), 111598993 (C > T), and 111599722 (A > G) that we found to be associated with two CAA interruptions when compared to one CAA interruption, with an odds ratio of 4.85 for haplotype CGGTG. This haplotype was associated with two CAA interruptions in all population groups with a similar odds ratio. In summary, we have identified haplotypes that correlate with the number of CAA interruptions within the CAG repeat region of ATXN2 and population-based studies demonstrate a decreased prevalence of CAA interruptions in the East Asian population. Further work in ALS, SCA2 and PD populations is underway to clarify disease mechanism and therapeutic relevance of these findings.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3025F: Neanderthal introgressed variants are associated with lung function

Authors:

J. Li[1], W. Gu[2], R. M. Salem[1], J. Li[2], S. P. Bhatt[3], A. L. Non[1], T. S. Simonson[1]; [1]UC San Diego, La Jolla, CA, [2]Univ. of California, San Francisco, San Francisco, CA, [3]Univ. of Alabama, Birmingham, AL

Abstract:

Background: Many modern human genomes have a small proportion of ancestry from archaic populations, including Neanderthals. While Neanderthals exhibited various distinct traits relative to modern humans, recent studies provide evidence for Neanderthal-specific contribution to present-day phenotypic variation in key physiological factors such as immune function and metabolism. While a relatively large thoracic volume has been noted in Neanderthals, it is unknown whether Neanderthal-introgressed alleles are associated with lung size and/or function in modern humans.
Methods: We compiled four publicly available multi-ancestry lung function genome-wide

association studies (GWAS) summary statistics of lung function measurements, including forced expiratory volume in one second (FEV1), forced vital capacity (FVC), FEV1/FVC ratio, and peak expiratory flow (PEF), which provided 1020 previously published genome-wide significant signals (p-values ≤ 5e-8). We then extracted a consensus set of ~250,000 Neanderthal-introgressed alleles from the published literature and performed a linkage disequilibrium (LD) score regression, using the 1K Genome Project, to determine the enrichment of Neanderthal-specific alleles in lung-function-associated loci from the top GWAS summary statistics.
Results: LD score regression indicated enriched heritability of Neanderthal-introgressed alleles in lung-function-associated loci, particularly in loci associated with FEV1 and FEV1/FVC. A post-hoc analysis identified 15 significant regions, all positively associated with the respective trait; notably, rs79966207, an exonic variant in the *PLXNB2* gene associated with FEV1. Eight other intronic or intergenic regions were associated with the FEV1/FVC ratio, including *ME3*, *NXNP1*, *FCGR3A*, *FCCR2C*, *RREB1*, *MEGF6*, *SNRPF-DT*, and *FLT1*. *DUSP10* and *TSPAN9* were associated with FVC, while PEF-associated genes are comprised of *RREB1*, *PDE3A*, *CRELD2*, and *PIM3*.
Conclusion: This is the first study that demonstrates regions of the modern human genome with Neanderthal introgression are associated with lung function measurements. These findings suggest that archaic introgression may play a key role in shaping the genetic architecture of lung function in modern humans.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3026F: Non-parametric representations of structure in genetic ancestry through time

Authors:

A. Selim[1], N. Hermosilla-Albala[2], L. Anderson-Trocmé[1], J. Novembre[1]; [1]Univ. of Chicago, Chicago, IL, [2]Inst. of Evolutionary Biology (UPF-CSIC), Barcelona, Spain

Abstract:

Many methods describe how genetic ancestry is structured either in the present-day or through time through a population history model. These methods often model the structure in terms of discrete panmictic units (populations) often with some migration/admixture between them. Alternative approaches conceptualize structure in a more continuous manner (e.g. due to geographically structured mating / “isolation by distance”). Currently, there is the need for new tools that can reveal the structure of genetic ancestry in a time-varying way, without strong assumptions about whether the structure has discrete units or continuous gradients. Here, we

present progress on this problem, by developing a method that provides a non-parametric description of genetic ancestry through time by leveraging reconstructed Ancestral Recombination Graphs (ARGs).
The method considers the genetic ancestors of present-day sampled sequences at different time points in the past, and attempts to reveal structure in how the genetic lines from the present coalesce into the genetic ancestors alive at each time in the past. The method proceeds by defining a matrix describing the amount of genetic material inherited by each present day sample from each genetic ancestor living at a specified point in the past. Factor analysis of this matrix (e.g. using PCA) produces a low-dimensional summary of the structure at each time point, and a temporal stack of the low-dimensional summaries reveals how the structure of genetic ancestry changes through time.
In simulations of populations diverged from a common ancestral population, the method nicely shows an increasing signal of genetic structure as one moves from the present into the past. As one moves further ancestrally past the divergence time, the signature of genetic structure nicely disappears on a time-scale proportional to the coalescent rate in the ancestor population. In the case of isolation-by-distance simulations, the method reveals a gradient of population structure that decays in magnitude on a time-scale proportional to the migration rate between populations, as expected.
Overall, the method is a promising alternative to other non-parametric methods that provide a single snap-shot of population structure (e.g. PCA)

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3027F: Overlap of adaptive genetic signatures between Andeans and Tibetans demonstrates convergent signals of natural selection

Authors:

E. Lawrence[1], W. Gu[2], R. Bohlender[3], H. Hu[3], C. Huff[3], F. Villafuerte[4], T. Simonson[5]; [1]Duke Univ., Durham, NC, [2]Univ. of California, San Francisco, San Francisco, CA, [3]U.T. MD Anderson Cancer Ctr, Houston, TX, [4]Univ. Peruana Cayetano Heredia, Lima, Peru, [5]UC San Diego, La Jolla, CA

Abstract:

Background: Human populations have thrived at high altitudes for generations despite environmental stresses from hypoxia, intense UV radiation, and aridity. Genetic adaptations, particularly in hypoxia-inducible factor (HIF) pathway genes, have been identified in modern groups residing in high-altitude areas of Peru and Tibet, despite distinct population histories. A study of adaptive signatures using whole genome sequence analysis in both Tibetan and

Andean populations shows evidence for convergent adaptation and underscores the significance of regulatory and protein-coding variants in hypoxia adaptation.
Methods: We performed a Composite of Multiple Signals (CMS) selection scan analysis with high-coverage (30x) whole genomes of 40 Andean natives from Cerro de Pasco, Peru (4340m). This involved a combination of five different selection tests. Regions with strong selection signals (Bayes factor > 6) were compared with those previously identified in Tibetans. We also explored regulatory landscapes via enrichment tests of genetic markers within DNase I hypersensitivity and transcription factor footprint regions, augmented by pathway analysis using Cytoscape and BiNGO.
Results: Strong selection signals were detected in less than 0.62% of the entire Andean genome, with 178 regions overlapping with strong selection scan targets in Tibetans, including key genes like *EPAS1* and *EGLN1* as well as novel regions like *EP300* and *VDR/SENP1*. In Andeans, markers with strong selection were significantly enriched in DNA regulation and repair areas, emphasizing their roles. Pathway analysis identified genes linked to immune response and inflammation (*IL17A*, *IL17F*), cell cycle (*MCM3*), and DNA repair (*UNG*, *ALKBH2*). Parallel analysis in Tibetans showed no enrichment of adaptive genetic markers near these regulatory areas.
Conclusion: Our findings validate and reveal new genetic convergences in adaptation mechanisms between Tibetan and Andean highlanders. This study further emphasizes the roles of regulatory factors in adaptive genetic signatures, including those involved in inflammation, DNA regulation, and repair in Andean genomes.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3028F: Paleolithic gene duplications primed adaptive evolution of human amylase locus upon agriculture

Authors:

F. Yilmaz[1], C. Karageorgiou[2], K. Kim[1], P. Pajic[3], K. Scheer[3], Human Genome Structural Variation Consortium, C. Beck[4], A-M. Torregrossa[3], C. Lee[1], O. Gokcumen[3]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]Univ. at Buffalo, Buffalo, CT, [3]Univ. at Buffalo, Buffalo, NY, [4]UConn Hlth.and The Jackson Lab., Farmington, CT

Abstract:

Starch digestion is a cornerstone of human nutrition, driven by the enzyme amylase encoded by the amylase genes. The salivary amylase gene, *AMY1*, is particularly significant, with previous studies suggesting that its copy number increased in response to agricultural diets. However, a lack of nucleotide-level resolution of the amylase locus has limited comprehensive evolutionary

analyses—until now. In this study, we performed a detailed nucleotide-resolution analysis of the amylase locus in 98 present-day humans, uncovering 30 distinct haplotypes for the first time. Our findings revealed that the coding sequences of all amylase gene copies are under negative selection. Phylogenetic reconstruction demonstrated that haplotypes with three *AMY1* gene copies, representing approximately 70% of observed haplotypes, originated prior to the out-of-Africa migrations of early modern humans. These prevalent haplotypes with three *AMY1* gene copies have driven non-allelic homologous recombination events at one of the highest rates observed for tandem repeats in the human genome. By analyzing unique 25 base pair sequences in 73 ancient human genomes dating from 300 to 45,000 years ago, we discovered that *AMY1* copy number variation predates agriculture, setting the stage for an adaptive surge in higher amylase gene copies with the advent of farming. Moreover, we identified additional *AMY1* gene copies in four archaic hominin genomes, suggesting that the initial duplication occurred before the human-Neanderthal split, approximately 800,000 years ago. Our study provides a comprehensive population-level analysis of the amylase locus, shedding light on the mechanisms and evolutionary history of *AMY1* duplication and copy number variability.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3029F: Phenotypic and genotypic distribution of ABO and RHD blood group antigens in Senegal.

Authors:

O. Kane[1], M. Gadji[2], A. R. S. Sene[3], R. Ndiaye[3]; [1]Service of Biological Hematology and Oncology-Hematology (BHOH), Faculty of Med., Pharmacy, and Odontology (FMPOS), Cheikh Anta Diop Univ., Dakar, Senegal, [2]Service of Biological Hematology and Oncology-Hematology (BHOH), Faculty of Med., Pharmacy, and Odontology (FMPOS), Dakar, Senegal, [3]Service of Human Genetics, UCAD, FMPO, Dakar, Senegal, Dakar, Senegal

Abstract:

Background
The frequency and the distribution of ABO and RHD blood group antigens among various Senegalese populations, remain a poorly understood. We undertook this study to assess the distribution and phenotypic and genic frequencies of A, B, O, and D antigens within different population groups in Senegal.
Methods
A total of 852 participants from 13 different Senegalese ethnic population groups, were

recruited across all 14 regions of Senegal, and were enrolled in this study. Blood grouping was performed using both Beth Vincent and Simonin methods on opaline plates by two different operators. ABO gene frequencies were calculated using the Bernstein method, and RHD system frequencies, considered as a di-allelic gene, were determined using the Landsteiner and Wiener method.

Results

We observed a phenotypic predominance of blood group O (51.3%), followed by group A (23.4%), group B (21.6%), and group AB (3.8%) considering the ABO blood system meanwhile, 95.70% of participants were RH1 (D positive), while 4.30% were RH: -1 (D negative) in the Rh system. For the genotypic frequencies, 48.7% of participants had the O allele, followed by 14.6% for the A allele and 13.5% for the B allele. The D gene (RH1) was with a frequency of 79.2%, and the d gene (RH -1) represented 20.7% of the population.

Conclusion

Our results reveal a lack of significant variation in ABO and RHD blood groups among the various ethnicities and regions of Senegal. Furthermore, we observed a marked similarity in phenotypic and genotypic frequencies with the sub Saharan African ethnic population groups in line of the Negroid characterization of Senegalese population.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3030F: Selection signatures in Pacific Islanders and associations with Native American ancestry

Authors:

J. Sokol[1,2], O. Garcia[3], J. Blanco-Portillo[1], M. Penjueli[4], S. Kasser[5], C. Shanks[3], F. Liu[1], S. Sacco[3], M. Feldman[1], G. Wojcik[6], K. Fox[7], A. Ioannidis[3]; [1]Stanford Univ. Sch. of Med., Stanford, CA, [2]Boston Univ. Sch. of Med., Boston, MA, [3]Univ. of California, Santa Cruz, Santa Cruz, CA, [4]New York Univ. Abu Dhabi, Saadiyat Island, United Arab Emirates, [5]Univ. of St Andrews, St Andrews, United Kingdom, [6]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [7]Univ. of California, San Diego, La Jolla, CA

Abstract:

The underrepresentation of diverse populations in genetic research hinders our understanding of the genetic basis of disease. This is particularly true for Oceanian populations, who account for less than 0.02% of samples in the NHGRI-EBI GWAS catalog. In the present study we analyze genetic variants from 3,944 individuals from Hawai‘i, 296 individuals from Central Eastern Polynesia (CEP), 165 individuals from Micronesia, and 605 individuals from Melanesia using the

iHS, IBDLD, Ohana, PBS, B0 statistics, and local ancestry deviations. We identify several genetic risk loci specifically associated with these populations. Notably, we discovered a significant Native American ancestry-associated selection signature linking GDF5 polymorphisms with osteoarthritis in Polynesian populations. We also show that 20 HLA genotypes vary in frequency among Melanesian, Micronesian, and Polynesian populations, providing new evidence of selection among specific Pacific Islander communities.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3031F: Selective dynamics of interruptions at short tandem repeats differ between coding and noncoding loci

Authors:

M. Goldberg[1], H. Dashnow[1,2], K. Harris[3,4], A. R. Quinlan[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]Univ. of Colorado, Aurora, CO, [3]Univ. of Washington, Seattle, WA, [4]Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Abstract:

Short tandem repeats (STRs) are hotspots of genomic instability that mutate at rates orders of magnitude greater than non-repetitive loci due to frequent replication slippage. Expansions at some STR loci are linked to Mendelian diseases, while variation at other noncoding loci may affect complex traits, possibly by altering transcription factor occupancy of nearby binding sites. Accordingly, some STRs are inferred to be under purifying selection, regardless of their instability. One or more 'interruptions', or bases that disrupt the locus's canonical repeat, significantly decrease an STR's mutation rate. For example, the onset of Huntington's Disease, a neurodegenerative disorder associated with somatic expansions of a trinucleotide coding STR, is delayed in individuals whose inherited alleles contain interruptions. Thus, decreasing mutation rate at some coding loci may broadly protect against deleterious phenotypes associated with locus instability. However, interruptions may themselves be deleterious at constrained loci, particularly at noncoding loci in gene regulatory elements, possibly disrupting the formation of secondary structures key to their function. We therefore hypothesized that the frequency of interruptions could depend on a locus's purifying selection, where the fitness effects of expansions but also interruptions could be more deleterious than at neutral loci. To test this hypothesis, we examined the distribution of interruptions at ~650,000 autosomal STRs. In the ~2,500 3 or 6bp motif STRs that fall within coding exons, we find that the density of interruptions in coding STRs increases with purifying selection on the gene, provided those interruptions lead to no amino acid change. In contrast, noncoding STRs under purifying

selection harbor fewer interruptions than those evolving neutrally. Our findings indicate that the abundance of interruptions may be partially explained at coding STRs by the benefit of a lower mutational burden at their linked loci. In contrast, maintaining a minimum core stretch of uninterrupted repeat may be key to the function of noncoding STRs that fall within regulatory elements, outweighing the benefits of lowering the mutation rate.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3032F: SNP heritability estimation in admixed populations with Haseman-Elston regression

Authors:

N. Kleman[1], J. Huang[2], S. Basu[1], A. Zaidi[1]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Penn State Univ., State College, PA

Abstract:

SNP heritability estimation in admixed populations can be inflated because of confounding due to linkage disequilibrium (LD) and/or shared environment. We showed previously that a common SNP heritability estimator - genome-wide restricted maximum likelihood (GREML) - can be biased in admixed populations even in the absence of confounding. Here, we describe the behavior of another commonly used estimator of SNP heritability - Haseman-Elston (HE) regression.

HE regression estimates SNP heritability from the correlation between the phenotypic similarity and the genotypic similarity amongst unrelated individuals using a method of moments approach. We show that the HE estimator is also biased in admixed populations but this bias is in the opposite direction of that observed with GREML and for different reasons. GREML essentially estimates the genic variance because it fails to capture the LD contribution, leading to under- and over-estimates of heritability for traits with positive and negative LD. HE regression, on the other hand, exaggerates the LD contribution, resulting in over- and under-estimates of heritability for traits with positive and negative LD.

We describe the quantitative behavior of the bias in HE heritability as a function of genetic architecture and admixture history and provide potential solutions to overcome this bias. We show that standardizing the genotypes by the covariance across loci, i.e., LD, rather than their variance, corrects for the bias in both HE regression and GREML. However, this solution is limited to situations where the number of individuals is greater than the number of markers. There is a push in human genetics to analyze data from diverse cohorts, given that most existing studies are conducted on individuals of European ancestry, which limits our understanding of

the genetic basis of disease. While this is a welcome shift, there is a clear need to understand the generalizability and potential limitations of statistical genetic methods in diverse multi-ethnic cohorts, which can be highly structured. We provide a clear mechanistic interpretation of SNP heritability estimates in admixed populations, their implications for genome-wide association studies and polygenic risk prediction, and potential solutions moving forward.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3033F: Temporal estimation of Gene-specific ω identifies genes for the emergence of novel SARS-CoV-2 strains having super infectivity and virulence

Authors:

A. Maiti; mydnavar, Southfield, MI

Abstract:

The SARS-CoV-2 genome, like any other viral genome, continuously accumulates mutations by replicating mistakes and retaining the favorable mutations to achieve optimum functions for its proteins to evade the host immune system. This continuous accumulation of functional (nonsynonymous) mutations creates an evolutionary stabilized, or saturated virus for maximum adaptation with the host systems. The proportion (d_n/d_s, ω) of nonsynonymous (d_n) to the synonymous mutation (d_s) over time indicates the level of stringency of the host immune system that the virus is encountering. An estimation of ω of the SARS-CoV-2 genome would indicate the evolutionary dynamics necessary to evolve into novel strains with increased infection, virulence, and vaccine neutralization. A temporal estimation of ω of the whole genome, and all twenty-nine SARS-CoV-2 genes of major virulent strains of alpha, delta, and omicron demonstrates that the SARS-CoV-2 genome originally emerged (ω ~ 0.04) with a strong purifying selection (ω < 1) and reached (ω ~ 0.85) in omicron towards diversifying selection (ω > 1). A marked increase in the ω occurred in the spike gene from alpha (ω = 0.2) to omicron (ω = 1.97). The ω of the replication machinery genes including *RDRP, NSP3, NSP4, NSP7, NSP8, NSP10, NSP13, NSP14*, and *ORF9* are markedly increased, indicating that these genes/proteins are yet to be evolutionarily stabilized and are contributing to the evolution of novel virulent strains. The delta-specific maximum increase in ω in the immunomodulatory genes of *NSP8, NSP10, NSP16, ORF4, ORF5, ORF6, ORF7A, and ORF8* compared to alpha or omicron indicates delta-specific vulnerabilities for severe COVID-19 related hospitalization and death. The maximum values of ω are observed for spike (*S*), *NSP4, ORF8,* and *NSP15*. Notably, the ω of three genes NSP2, NSP6, and ORF10 are mostly unchanged from the original strain to omicron

indicating that these genes are highly adapted and optimized for their function in the human system. Thus, the drug or vaccine design against these genes/proteins would not need emerging strain-specific repeated drug/vaccine design and testing. Overall, the gene-specific temporal estimation of ω identifies specific genes for its super-infectivity and virulency and unchanged genes that could be targeted for drug/vaccine development.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3034F: The battle of the sexes in humans is highly polygenic

Authors:

J. Cole[1], S. B. Carly[1], M. M. Johnson[2], P. Golightly[1], J. Carlson[1], M. J. Ming[1], A. Harpak[1], M. Kirkpatrick[1]; [1]Univ. of Texas at Austin, Austin, TX, [2]Fred Hutchinson Cancer Ctr., Seattle, WA

Abstract:

Sex-differential selection (SDS), which occurs when alleles have different fitness effects in females and males, profoundly impacts the maintenance of genetic variation, disease risk, and other demographic properties of natural populations. Quantifying SDS using conventional population genetic approaches is not possible since the sexes mix their autosomal genomes each generation. Here, we present a novel method that leverages sex differences in haplotype frequencies in the current generation to assess the strength of SDS. Using data from over 300K individuals in the UK Biobank, we find highly polygenic signals of SDS acting on both viability and fecundity. We uncover a life-history tradeoff: alleles that increase viability more in one sex increase fecundity more in the other. Lastly, we find that SDS on fecundity in males acts on alleles that affect arm fat-free mass. Our results connect the widespread evidence for SDS on human phenotypes with its genomic impact.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3035F: The effect of long-range linkage disequilibrium on allele frequency dynamics under stabilizing selection

Authors:

S. Negm, C. Veller; The Univ. of Chicago, Chicago, IL

Abstract:

Stabilizing selection on a complex trait reduces the trait's genetic variance by (i) generating negative correlations (linkage disequilibria) between trait-increasing alleles throughout the genome and (ii) selecting against the rarer allele at polymorphic loci that affect the trait. Here, we characterize the impact of the negative linkage disequilibria, which stabilizing selection generates on a rapid timescale, on the subsequent allele-frequency dynamics at individual loci, which occur on a slower timescale. We obtain expressions for the expected per-generation change in minor-allele frequency at individual loci, as functions of the strength of selection on the trait, its heritability, and the linkage relations among loci affecting the trait. Using whole-genome simulations of polygenic traits under stabilizing selection, we show that our expressions predict allele-frequency dynamics more accurately than the expressions that have previously been used for this purpose. Our results have implications for understanding the genetic architecture of complex traits, as well as for the rate at which the loci that underlie variation in a complex trait turn over---one possible cause of the poor portability of association-study results across human populations.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3036F: The genomic population history of the Aleutian Islands before and after European contact

Authors:

S. Hay[1], B. Sanderson[1], K. G. Beaty[1], J. Tackney[1], M. Thomas[2], R. Macleod[3], D. Orourke[1], J. Raff[1]; [1]Univ. of Kansas, Lawrence, KS, [2]Univ. Coll. London, London, United Kingdom, [3]Homerton Coll., Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

Historical and ethnographic sources convey the devastating effects of European contact and subsequent colonization on Native American groups. These sources indicate a drastic overall decline in Indigenous populations across the two continents due to war, famine, and introduced disease. The Unangax of the Aleutian Archipelago have a distinctive history of European colonization as genetic diversity evolves uniquely within island populations and initial contact, with subsequent gene flow, did not occur before 1741. Previous studies attempting to detect the genomic impact of European colonization on Indigenous groups in the Americas, such as the magnitude of population loss, used only contemporary genomes to infer pre-contact diversity. Additionally, these studies were focused on broad-scale population histories and not geographically local groups. Recent analyses show using ancient genomic data in tandem with contemporary genomes increases the resolution of demographic inference. In this study, we

compared ancient and contemporary genomes from the same Unangax population to improve our ability to infer past demographic events. Our dataset was comprised of genome-wide data from 33 contemporary individuals captured on the Axiom Human Origin Array, and from 12 ancient individuals captured on the 1240K array. We analyzed the structure of our data using principal-component and admixture analyses, inferred the wide-ranging ancestral relationships present in the contemporary individuals, and inferred features of the demographic history of this population using maximum composite-likelihood methods based on coalescent simulation. Our findings are consistent with our knowledge of the ethnographic history of gene flow in the Aleutian Islands that has occurred since initial contact. Due to these past demographic events, the contemporary Unangax have a distinct pattern of inter-population variation in genetic structure that is unlike anywhere else in the North American arctic and sub-arctic.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3037F: The impact of background selection on deleterious alleles and its consequences on complex trait variation

Authors:

X. Li[1], J. Novembre[2], J. Berg[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ Chicago, Chicago, IL

Abstract:

In humans, background selection has been estimated to be a dominant factor shaping levels of neutral variation. However, the consequence of background selection on quantitative traits is poorly understood, and yet potentially relevant to understanding the genetic architecture and phenotypic distribution of complex traits.
To address this open question, we use theoretical models and simulations to study how background selection impacts polygenic traits in two settings: complex traits under stabilizing selection and complex diseases under directional selection.
First, we study how background selection impacts the evolution of corresponding single-site models for two trait settings. We found that for variants under additive selection and under-dominant selection, background selection only reduces the genetic diversity of weakly selected variants but has negligible effects on strongly selected variants. This implies that background selection shifts the genetic variance contribution more towards rare variants. Further, we find that, surprisingly, background selection can increase the genetic diversity of variants with moderate under-dominant selection.
Next, we extend our results to the multi-locus mutation-selection-drift-balance models. Consistent with our single-site results, background selection reduces the heritability for traits

with a substantial contribution from weakly selected variants. Different from the single-site prediction, for diseases under the liability threshold model, the reduction in genetic variance due to background selection is compensated by an increase in the realized strength of selection against liability. This results in an increase in disease prevalence, validated by both mathematical models and simulations. Using plausible parameter values that generate a diversity reduction of 17% due to background selection (as estimated in humans), we find background selection leads to an increase in disease prevalence of a trait with 50% heritability by ~15%.
Overall - Our work addresses the role of background selection as a factor impacting complex traits genetics, and reveals how its effects on complex traits depend on the distribution of fitness effects for their causal loci.

Session Title: Evolutionary and Population Genetics Poster Friday Session

Board 3038F: Whole genome sequencing of 1,795 individuals from large pedigrees in Mongolian populations using Ultima Genomics' UG100

Authors:

S-b. Lee[1], D-s. Lee[2], H. Kim[3], S. Pollock[4], I. Soifer[4], C. Kim[1], J-i. Kim[5], J-S. Seo[6]; [1]Macrogen Inc, Seoul, Korea, Republic of, [2]Univ. of Seoul, Seoul, Korea, Republic of, [3]Nanyang Technological Univ., Singapore, Singapore, [4]Ultima Genomics, Newark, CA, [5]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of, [6]Seoul Natl. Univ. Bundang Hosp., Seoul, Korea, Republic of

Abstract:

Whole Genome Sequencing (WGS) has revolutionized human genetics research, yet significant disparities persist in genomic data representation, particularly among non-European populations. From 2003 to 2009, as part of the GENDISCAN project, we collected DNA samples and comprehensive pedigree information from the Buryats, one of the Mongolian ethnic groups native to southeastern Siberia and Mongolia. Utilizing Ultima Genomics' UG100 sequencing platform, we conducted high-coverage (>30x) WGS on 1,795 individuals with familial information for 365 pedigrees. In our preliminary analysis, we sequenced 300 individuals, including 97 trios (291 individuals in total). This study represents the first large-scale WGS analysis of the Buryat population, revealing critical insights into the ancestry and historical migrations that shaped the genomic landscape of contemporary Northeastern Asians, including those in Korea, Japan, and China. The UG100 platform demonstrated cost-effectiveness ($1 per GB), rapid turnaround (less than 24 hours), and a fully automated workflow, highlighting its potential for widespread use in population genomics research. Our findings underscore the

necessity of including underrepresented populations in genomic studies to enhance personalized medicine and achieve a comprehensive understanding of human genetic diversity. Affordable WGS is vital for the broad implementation of precision health, facilitating tailored medical treatments for diverse populations.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7052F: A study of practical application of Instrument to Assess Children's Assent (IACA): Explanation support ICT materials and other factors related to assent of children participating in genome analysis research

Authors:

I. Ishiyama[1], T. Kobayashi[2]; [1]Tokoha Univ., Shizuoka-city, Japan, [2]Tohoku Med. Megabank Organization, Tohoku Univ., Sendai-city, Japan

Abstract:

[Background]Aiming for practical application of genomic medicine, genome analyses are conducted on patients and healthy people, including children. It is the duty of medical researchers to try to obtain appropriate assent from children, but it is not easy to explain toward children who have not received education in human genetics. In order to improve children's genomic literacy and to contribute the system with respect of children's human rights, we developed an ICT material that assist in explaining genome and support children’s decision making. To evaluate usefulness of the material, we used Instrument to Assess Children’s Assent (IACA), which has been used only for healthy children before. [Objective]The purpose of this study is to use IACA to evaluate the usefulness of Monokuro-Genome®, the ICT material developed for first-to-third grade elementary school children for support informed assent, and to clarify other factors related to children's assent process. [Methods]A double-blind randomized controlled trial was conducted on 14 children who had been proposed to participate in a genome analysis research of rare and intractable disease families with prior consent from their legal guardians. The children were divided into two groups: those who used Monokoro-Genome® during the informed assent process and those who did not. The assent process was evaluated using a semi-structured interview survey using IACA. In addition to the scores, conversations and behaviors of the children during interviews were recorded in the IACA forms. Quantitative analyses were conducted using IACA scores, age, gender, intellectual

development status, the presence or absence of prior explanations from the family about the study, and use or not use of Monokoro-Genome® as variables. Additionally, a qualitative analysis based on the records of the IACA forms was also conducted. [Results]The results suggested that intellectual development status, prior explanations from the family, the child's own acceptance of the disease, and logical thinking that can explain the reasons for decision-making were more closely related to IACA scores than the use of Monokoro-Genome®. [Discussion]Concerning patients in the first-to-third grades of elementary school who had been proposed to participate in a genome analysis study, it was suggested that intellectual development status, explanations from the family, acceptance of the disease, and logical thinking were factors to achieve appropriate assent process. From now on, we will conduct research on children in the fourth-to-sixth grades of elementary school and on healthy children to verify the usefulness of Monokuro-Genome®.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7053F: A systematic approach to community-engaged recruitment for genetics research

Authors:

J. Yap[1], M. M. Horstmann[2], Y. Bombard[3], R. Feys[4], H. C. Howard[5], A. S. Kokayi[1], K. M. Meagher[1], M. Michie[6], K. Musunuru[7], K. E. Ormond[8], K. A. Riggan[1], S. Rubeck[2], R. Isasi[9], M. A. Allyse[10]; [1]Mayo Clinic, Rochester, MN, [2]Case Western Reserve Univ., Cleveland, OH, [3]Univ of Toronto, Toronto, ON, Canada, [4]Univ. of Miami, Miami, FL, [5]Lund Univ., Lund, Sweden, [6]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [7]Univ. of Pennsylvania, Philadelphia, PA, [8]Stanford Univ and ETH-Zurich, Zurich, Switzerland, [9]Univ. of Miami Miller Sch. of Med., Miami, FL, [10]Mayo Clinic, Jacksonville, FL

Abstract:

Multiple international statements have called for including people with lived experience of genetic conditions in the governance of the translation of genetic therapies. Previous studies have suggested that some reluctance toward community-engaged research is due to extensive time investment and difficulty recruiting community participants. Our study, eliciting perspectives on prenatal genetic therapies, is currently recruiting hundreds of participants from 10 genetic condition communities, along with scientists and policy-makers. Here we describe the systematic approach our team has developed to successfully conduct community-engaged recruitment for genetics research.

Our research plan established a dedicated community engagement outreach team to foster relationships with community organizations and recruit individuals with lived experience of each of the 10 conditions in our study. Contact with organizations and response rates were recorded and reviewed weekly. In parallel, researchers partnered with professional societies to engage their expert membership. We anticipated variable responses based on prior studies. Factors influencing the success of external partnerships were organizational bandwidth, accessibility to community members, and baseline interest in research partnerships.
The outreach team has also collaborated with interested community organizations to craft optimized recruitment strategies specific to their condition and community. We have demonstrated that outreach to individuals across diverse genetic communities can be successful, especially when the study team adopts a bi-directional approach to community partnership and direct community recruitment activities. Although this approach requires dedicated personnel and time investment, it results in a mutually positive experience, encouraging future research involvement by all members of communities impacted by gene editing and effective study completion.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7054F: Adaptation of the Witness Project® to include genetic testing for hereditary breast and ovarian cancer risk

Authors:

K. Leopold[1], E. Shoulders[2], L. Carter-Bawa[3,2]; [1]Hackensack Meridian Sch. of Med., Nutley, NJ, [2]Cancer Community Outreach and Engagement at John Theurer Cancer Ctr. at Hackensack Meridian Hlth., Hackensack, NJ, [3]Cancer Prevention Precision Control Inst. at the Ctr. for Discovery & Innovation (CDI) at Hackensack Meridian Hlth., Nutley, NJ

Abstract:

Many patients are not aware of the potential impact of genetics on cancer risk and the availability of genetic testing to determine risk. Patients who have a high school education or less, who have a low income, and who are Black, Hispanic or Asian are less likely to be aware of cancer screening for heritable cancers. Community leaders and researchers initially designed the Witness Project® to increase awareness of breast self-examination and mammograms among women at Black churches. Role models (“witnesses”) from each church community described their motivation to receive screening and their experience with screening resulting in a significant increase in uptake. The Witness Project® has been successfully adapted for

different types of cancer and diverse populations. Community-based participatory research allows organizations and community members to come together in a Community Advisory Council (CAC) and collaborate with researchers in the development and dissemination of equitable research projects. This project aims to increase awareness of genetic testing for hereditary breast and ovarian cancer (HBOC) guided by the Witness Project® framework using a community-based participatory research model. Past materials, expert input, and a literature review informed initial development of an educational presentation for use during role model-led educational sessions. To facilitate review by our institution's CAC, the team developed a list of open-ended qualitative questions using the Health Belief Model to assess the acceptability of this presentation and make improvements before the study's pilot period. Results from CAC review identified meaning and flow of the presentation as strengths and accessible language and engaging imagery as areas-for-improvement in the planned intervention. CAC members recognized the importance of the presentation's subject matter for helping spur proactive management of HBOC risk. CAC members highlighted presentation storytelling and flow as helpful in understanding the complex concept of hereditary cancer. Though simple language used in the presentation was helpful to understand basic genetic principles, CAC members noted that the words "heritable" and "hereditary" were unfamiliar and above fourth grade reading level. CAC members recommended use of more pictures and brighter colors to improve audience engagement. Results show that an adaptation of the Witness Project® could be acceptable to the intended population, and CAC feedback will be used to inform further development of a pilot.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7055F: Breaking barriers: identifying challenges in accessing genetic diagnostics for at-risk Puerto Rican individuals

Authors:

E. Albino[1], P. Giusti-Rodriguez[2], S. Carlo[3], F. Velez-Bartolomei[4], A. Cornier[5], C. Buxo[6]; [1]Univ. of Puerto Rico - MSC, San Juan, PR, [2]Univ. of Florida Coll. of Med., Gainesville, FL, [3]Ponce Hlth.Sci. Univ., Cabo Rojo, PR, [4]Hosp. Pediatrico Univ.rio, San Juan, Puerto Rico, [5]San Jorge Children s & Women s Hosp., San Juan, PR, [6]Univ. of Puerto Rico Med. Sci. Campus, San Juan, PR

Abstract:

Genetics research involving individuals at risk of inherited conditions—particularly those with limited diagnostic information—often elicits varied responses from potential participants. Our

goal is to explore the perceptions and challenges these individuals face in accessing genetic diagnostics, focusing on their concerns, motivations, and systemic barriers. This mixed-methods study involved 40 Puerto Rican trios identified as being at risk for conditions with limited diagnostic information. We utilized structured surveys to gauge general perceptions and feelings toward genetic research and in-depth interviews to capture personal narratives and experiences. All sessions were recorded, transcribed, and analyzed using Dedoose qualitative analysis software. Thematic analysis was employed to identify recurring themes and sentiments. The study revealed significant challenges faced by individuals in Puerto Rico at risk for inherited conditions in accessing genetic diagnostics. Financial barriers emerged as a primary concern for 91% of participants, highlighting the need for more affordable healthcare options. Additionally, 73% of respondents emphasized the necessity for increased availability of medical geneticists, while 100% reported delays in scheduling timely medical appointments. Access to specialized laboratories was limited for 8% of participants, underscoring the need for expanded testing facilities. Furthermore, 27% expressed the importance of better coordination of health services closer to their homes. Participants' narratives vividly described the emotional and logistical struggles in navigating the healthcare system, emphasizing the urgency for systemic improvements to support those on their diagnostic journeys. These findings highlight critical areas for intervention to enhance the accessibility and efficiency of genetic diagnostics for at-risk populations. Understanding the barriers to accessing genetic diagnostics is crucial for improving the participation and outcomes of genetic research. This study provides a comprehensive overview of the systemic challenges faced by individuals at risk for inherited conditions. Our findings highlight the need for enhanced financial support, increased availability of medical geneticists, expedited appointment scheduling, expanded access to specialized laboratories, and better coordination of health services. Addressing these issues can improve outreach and education strategies, ensuring potential participants are well-informed, their concerns are addressed, and they are empowered to make decisions aligned with their values and interests.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7056F: BYOB (Beat Your Own Bots): Mitigating automated responses in genetics survey research

Authors:

A. McClellan, E. Charnysh, K. Sanghavi; The Jackson Lab. for Genomic Med., Farmington, CT

Abstract:

In web survey studies, malicious actors can exploit bots to submit multiple responses and unfairly access incentives designed to boost participation. A bot is an automated software that performs repetitive tasks on behalf of a user. Since bots can be programmed to mimic human behavior, distinguishing their responses from actual ones poses significant challenges for researchers seeking authentic data. Here, we outline strategies used to address unanticipated issues with bots in an international study of genetic counselors conducted via a Qualtrics survey. The goal of this incentivized, quantitative study was to assess genetic counselors' views on somatic genetic testing and germline cancer risk assessment in different practice settings. As bot security measures, a reCAPTCHA v2 question and a Qualtrics feature to deter multiple submissions were applied to the survey prior to distribution. The survey was initially circulated as part of a National Society of Genetic Counselors email blast. To enhance recruitment, the study information, along with the incentive details, was posted on a public social media account, resulting in a questionable surge in responses. After closing the survey, data collected during this surge (N=481) were scrutinized for suspicious characteristics as informed by the current literature: a) unrealistic completion time (i.e., 5 minutes or fewer), b) poor-quality free-text responses (e.g., inconsistent with the question), c) unanswered free-text responses (e.g., selecting "other" but not specifying further), d) contradictory information (e.g., indicating having 15 or more years of experience but also that they are age 20-29), or e) initiation or submission at the same time as other responses. Responses with less than 80% completion (n=33) were excluded from analysis. Responses scoring 3 or more (n=322) were deemed to be likely bot generated or fraudulent. Responses scoring less than 3 (n=126) were manually reviewed for other "red flags" (e.g., having the same open-ended answer as another submission). Ultimately, 432 (89.8%) responses were suspected to be illegitimate. In the era of bots and other advanced artificial intelligence applications, employing robust pre- and post-survey bot mitigation procedures is crucial for ensuring the fidelity of research data collected through web surveys.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7057F: Catalyzing biobank expansion: Empowering clinical staff champions to amplify biospecimen collections for a population-scale biobank

Authors:

A. Tanaka[1], E. C. Hearst[1], J. L. Martin[2], K. W. Hess[1], E. Casteel[2], V. Curreri[2], C. S. Greene[2], S. Jones[2], D. P. Kao[2], D. Koffenberger[2], E. L. Shalowitz[2], C. Aquilante[2], The Colorado Center for

Personalized Medicine; [1]UCHlth., Aurora, CO, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Population biobanks are a critical resource to facilitate research and discoveries in genomic medicine. In some biobanks, patients consent to participate in the study and biospecimens are collected at a subsequent healthcare visit by clinical staff. However, maintaining desired sample collection rates over time can be challenging due to competing demands on clinical staff time. To this end, we describe the design, implementation, and impact of a clinical staff member education, incentive, and engagement program at an institutional population-scale biobank. Participants in the biobank have a blood sample collected by clinical staff during routine blood draws. We developed a Biobank Champions Program, whereby department leaders selected clinical staff members to serve as biobank "Champions." Champions were provided intensive education about the biobank via monthly virtual training sessions. They then developed training materials that were individualized at the clinic level and provided peer-to-peer education. Champions were tasked with increasing their staff's understanding of the biobank, with the goal of increasing the site's sample collection rates by 10-20% within the first three months of the program. Non-monetary (e.g., continuing education credits) and monetary (e.g., gift cards) incentives were given to Champions. For this evaluation, Aug-Oct 2023 and Feb-Apr 2024 served as the pre-implementation (baseline) and post-implementation periods, respectively. Data were analyzed using descriptive statistics and Wilcoxon Signed Rank tests. In total, 2,697 biobank samples were collected across 21 sites, including 10 labs (2,023 samples, 75%) and 11 primary or specialty care clinics (674 samples, 25%). During the pre-implementation period, 943 samples were collected, with a median (IQR) of 30 (13.5, 56) samples per site over 3 months. Implementation of the Biobank Champions program resulted in an 86% increase in the number of collected samples. Specifically, 1,754 samples were collected over the post-implementation period, with a median (IQR) of 65 (24-119) samples per site over 3 months, $p<0.001$. The percentage increase in collected samples was greater in clinics (135.3%; samples pre=201 vs post=472, $p=0.004$) compared with labs (72.6%; samples pre=742 vs post=1,281, $p=0.008$), indicating differences in program impact based on the type of department. Our findings demonstrate that implementing an individualized clinical staff incentive and engagement program results in an immediate and significant increase in sample collection rates across various departments, with the greatest impact in primary and specialty care clinics.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7058F: Clients' experiences of empathy in genetic counseling for hereditary breast and ovarian cancer: A qualitative study in Japan

Authors:

C. Tomozawa[1,2,3], M. Kaneko[4], M. Sasaki[2,3], H. Miyake[2,3]; [1]Natl. Cancer Ctr. Hosp., Chuo-ku, Japan, [2]Graduate Sch. of Humanities and Sci., Ochanomizu Univ., Tokyo, Japan, [3]Inst. for Human Life Sci., Ochanomizu Univ., Tokyo, Japan, [4]The Jikei Univ. Hosp., Tokyo, Japan

Abstract:

Cancer genetic counseling needs have been rapidly evolving with the expansion of clinical comprehensive genomic profiling and genetic diagnosis approaches for hereditary cancers. Empathy is a significant element in genetic counseling for building relationships with the clients and addressing their issues. However, there are few reports on the experiences of the clients about their perceived empathy in genetic counseling. Therefore, this study aimed to reveal empathy perceptions of the clients during cancer genetic counseling. Semi-structured interviews were conducted, and a grounded theory approach was used for data analysis. A total of 13 participants were recruited from organizations for patients with cancer, among whom 11 were patients with hereditary breast and ovarian cancer (HBOC) and two were relatives of patients with HBOC. Data analysis was organized into five categories related to experiences with empathy: (i) *prior context to perceive empathy* (ii) *understanding and consideration*, (iii) *bedside manner*, and (iv) *impacted area of perceived empathy*; and (v) *no empathy*. While most participants in the present study perceived empathy during their genetic counseling, some did not experience empathy. Perceiving empathy from genetic counseling practitioners can play a significant role and impact the feelings of security and satisfaction of the clients during the care process. However, practitioners' behaviors can cause discomfort or disappointment if their assessment and approaches do not fit the needs or values of the clients. Understanding the empathy perspectives of clients can provide suggestive information for healthcare providers to develop better counseling approaches and systems for genetic counseling/testing.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7059F: Clinical genetic testing outcomes of 189 cancer patients: Importance of no-cost patient germline genetic testing access

Authors:

B. Sjoding, P. Brause, C. Handford, M. Laurino; Fred Hutchinson Cancer Ctr., Seattle, WA

Abstract:

INTRODUCTION
Access to germline genetic testing for hereditary cancer syndromes has improved with changes to insurance policy criteria for coverage, reduced cost, and self-pay options of $250 at multiple laboratories. However, barriers remain for those who are uninsured, underinsured, or who do not meet insurance criteria despite a cancer diagnosis and find $250 a financial burden. Since 2016, Fred Hutchinson Cancer Center (Fred Hutch), a merger of Seattle Cancer Care Alliance and Fred Hutchinson Cancer Research Center has had the opportunity to offer no-cost genetic testing to patients who meet the financial eligibility criteria.
METHODS
From July 2016 - May 2024, 189 patients who found the cost of genetic testing a burden and who met financial eligibility were offered 30-gene next-generation sequencing panel testing at no additional cost. We report on the clinical testing outcomes of the 157 patients who completed testing and received results.
RESULTS
Of the 189 individuals offered no-cost genetic testing, 32 (16.9%) did not obtain a result due to a variety of reasons such as sample failure, not sending a saliva collection kit back, etc. Of the 157 individuals who received results, 110 (70.1%) had negative results, 29 (18.5%) had a pathogenic variant identified, and 18 (11.5%) were found to have a variant of uncertain significance. Most samples came from patients identifying as female (77.2%), of European ancestry (49.7%), and were known to have health insurance (77.7%). Pathogenic variants were identified in the following genes: *BRCA2* (17.2%), *MUTYH* heterozygotes (17.2%), *CHEK2* (13.8%), *ATM* (13.8%), *BRCA1* (10.3%), *BRIP1* (6.9%), *PALB2* (6.9%), *MITF* (6.9%), *NBN* (3.4%), Other/Mosaic *TP53* (3.4%).
CONCLUSIONS
Despite having insurance, many patients in this program face out-of-pocket costs due to coverage criteria. Without no-cost genetic testing, 18.5% of tested individuals would have missed crucial pathogenic variant results. Access to genetic testing is vital for managing health, especially for those unable to afford it, impacting not only their own health but also that of their family. Policy updates are needed to reduce patient financial burden.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7060F: Complex Medical Management and Ethical Considerations in Foster Care Twins with Congenital Disorder of Glycosylation Type 1A

Authors:

E. Lopez[1], P. Suhas[1], O. Mirabella[2,3], E. Sperry[2,3], K. Widmeyer[2], G. Hijazi[2,3], Y. Wu[2,3], A. Shillington[2,3], R. Hopkin[2,3]; [1]Coll. of Med., Univ. of Cincinnati, Cincinnati, OH, [2]Div. of Human Genetics, Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [3]Dept. of Pediatrics, Univ. of Cincinnati, Cincinnati, OH

Abstract:

We describe premature twins who presented with complex care needs related to intrauterine growth retardation, severe congenital diarrhea, progressive kidney disease, hypotonia and a degenerative course who passed from complications due to severe type 1a PMM2-CDG glycosylation disorder. We discuss their journey in the healthcare and foster care system; highlighting the extensive resources needed and barriers to best care. This will be reviewed in the context of the core principles of medical ethics. This is a more complex issue with the advent of precision medicine and the rapidly advancing diagnostic tools and interventions for rare genetic disease. In this case there were at least five separate therapeutic options that were considered for the infants. However, there were many barriers to care leading to delayed decision making which may have contributed to increased suffering and diminished access to treatment for the twins. These delays and barriers also caused considerable moral and emotional distress to those caring for them in the hospital setting. Therapeutic considerations involved use of either novel genetic therapies, comfort care with limited invasive procedures, or standard of care with attempts to prolong survival. All of these are currently part of the usual care in most large children's hospitals which implies, in the spirit of equity, they should also be made available to children with difficult marginalized social challenges as well as those with more resources. We argue that shortcomings of the system resulted in inequitable access to resources and worsened the outcome for all involved in this case. Through this report, we highlight a need for improved integration between the US healthcare and foster care systems to support better health outcomes of pediatric patients in the foster system or other nontraditional care arrangements with rare, complex genetic disease.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7062F: Elective genome screening, clinical experience in 1300 healthy individuals

Authors:

M. León Otegui[1], R. Bernad Galán[1], B. Palao[1], E. Ordoñez[2], I. Castilla[2], M. Lechuga[2], M. Moreno[1], I. Perea[1], E. Fernandez[1], N. Lobato[1], L. García[2], V. Cirigliano[2], L. Izquierdo[1]; [1]Veritas Intercontinental, Madrid, Spain, [2]Veritas Intercontinental, Barcelona, Spain

Abstract:

Background/Objectives: Elective genome sequencing (EGS) it the genome analysis of individuals without a medical indication. This screening strategy enables the study of actionable conditions with the goal of preventing. To provide clinical utility to this approach, all the variants found should be curated properly. We present the results of the first 1300 patients performed with the only EGS test available with clinical utility.
Methods: The test was performed in over 1300 people after a medical consultation where myGenome (Veritas Intercontinental) test was prescribed. Extracted genomic DNA from blood or saliva is sequenced at approximately 30X average coverage. Interpretation was performed with a proprietary genomic pipeline that analyzes 600 genes based on ACMG-SF list and ACOG recommendations, together with pharmacogenomics and multifactorial diseases. Clinical variants were curated by a team of expert geneticists.
Results: Clinically relevant genetic variants were identified in 21% of individuals. From all the variants detected, 18.5% are included in ACMG-SF v3.2 list. Divided by medical area, the variants were related to hemochromatosis (42.6%), hereditary cancer predisposition (26.6%),hereditary cardiovascular disease (12.8%) and other diseases (17.8%). Additionally, 64% patients were carriers of one or more recessive conditions, 87% presented risk alleles for multifactorial diseases and relevant pharmacogenomics associations were detected in 99% of the patients, 32% of them were severe gene-drug interactions.
Conclusion: The inclusion of EGS in the medical setting allows an effective identification of relevant genetic findings related to clinical diseases, carrier status, pharmacogenomics, and risk alleles, that provides an outstanding preventive approach for clinicians.
Grants: None significant to the purpose of this research.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7063F: Exploring Uterine Fibroid Disparities among Black and Hispanic Women: Insights from the *All of Us* Workbench

Authors:

L. Gerido[1], M. Gillentine[2], K. L. Malone[3], Y. H. Lee[4], M. E. Bates[5], S. Holechek[6], H. Sharifan[7], I. B. Van Den Veyver[8]; [1]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Seattle Children s Hosp., Seattle, WA, [3]Baylor Coll. of Med.cine, Houston, TX, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Fayetteville State Univ., Fayetteville, NC, [6]Arizona State Univ., Tempe, AZ, [7]Univ. of Texas at El Paso, El Paso, TX, [8]Baylor Coll. of Med., Houston, TX

Abstract:

Uterine fibroids (UF) are the most common benign neoplasm of the uterus. UF are documented to cause abnormal uterine function, which can disrupt the quality of life. The prevalence of UF in the overall population approaches 70%, and an increased prevalence of 80% in African Americans has been well documented. Although the Hispanic population is the largest ethnic minority in the United States, few studies report prevalence and risk specific to the population group. Often disparities research focuses on racial groups, but the Hispanic ethnic minority group spans all racial categories. The *All of Us* Research program provides a unique opportunity for health equity research because it actively recruits individuals belonging to demographic categories that have been underrepresented in biomedical research.
Our preliminary analysis of the *All of Us* sample population indicates 287,012 participants who released electronic health records (EHR), of which 15,374 had at least one clinical diagnosis of UF (prevalence of 5.4%). Those self-identifying as Black or African American had the highest prevalence of UF (5.7%), followed by those self-identifying as Hispanic or Latino (4.6%). We will conduct a phenome-wide association study (PheWAS) analysis with logistic regression models to observe variability in risk of diagnoses associated with race and ethnicity. Firth logistic regression will be performed using Python within the *All of Us* Researcher Workbench with covariates including current age, age at diagnosis, dates and types of clinic encounters, and family history. We will calculate the odds ratios (OR) or relative risk for our comparisons of our data with those in existing literature. Additionally, geospatial neighborhood analyses for environmental factors (i.e. exposure to endocrine-disrupting chemicals) will be conducted using the Zip Code Socioeconomic Status Data included in the *All of Us* Researcher Workbench.
We recognize that racial and ethnic identities are not scientifically, genetically, or biologically valid. The *All of Us* platform includes rich and extensive data to study the complex nature of disparities and to develop targeted interventions for diverse populations.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7064F: Genetic Counseling Services for Hereditary Breast and Ovarian Cancer: Patients' Experience and Satisfaction with Different Service Models ⭐

Authors:

H. Nabi, The CMonGene Study Group; Université Laval, Québec, QC, Canada

Abstract:

In a context of scarce resources and increasing demands, some patients do undergo genetic testing for hereditary breast and ovarian cancer (HBOC) through different service models, which may or may not include a formal genetic counseling session. We assessed whether the genetic testing service model and attendance at a formal genetic counseling session impact patients' experience and satisfaction with the genetic testing process for HBOC. A random sample of patients who underwent genetic testing for HBOC at the *CHU de Québec-Université Laval,* Quebec City, Canada between 2017 and 2021 was invited to respond to a 35-item survey over the phone or online. The survey included a validated French version of the Genetic Counseling Satisfaction Scale and the Decision Regret Scale. Patients' experience with the testing process was assessed using a modified version of the Royal Marsden Satisfaction Questionnaire. Descriptive statistics and Fisher exact tests were used for the analysis. Of the 501 patients who completed the survey, 162 (32.3%) received services within a traditional genetic-led testing model, 258 (51.5%) within a collaborative oncology-led model, and 81 (16.2%) within a fast-track oncology-led model. Among them, 378 (75.4%) reported attending a genetic counseling session, while 109 (21.8%) did not attend, and 14 (2.8%) could not recall. Four aspects of the genetic testing experience consistently varied across different service models and according to attendance at a genetic counseling session: i) reception of patient information material prior to the test, ii) information regarding the possibility of additional discussions with the genetic team to support the decision, iii) details on the timeline for receiving the test result, and iv) information on the method by which the result will be provided. The type of oncogenetic service model and attendance at a genetic counseling session appear to influence the quality of the care for patients undergoing genetic testing for HBOC. Despite these differences, patients' satisfaction remained remarkably high, and decision regret was consistently low across all service models. These findings are encouraging for the implementation of models aimed at increasing access to genetic service.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7065F: Genomic newborn screening, first clinical experience in Spain

Authors:

R. Bernad Galán[1], M. León Otegui[1], E. Ordoñez[2], B. Palao[1], L. García[2], I. Perea[1], E. Fernández[1], N. Lobato[1], M. Lechuga[2], I. Castilla[2], M. Moreno[1], V. Cirigliano[2], L. Izquierdo[1]; [1]Veritas Intercontinental, Madrid, Spain, [2]Veritas Intercontinental, Barcelona, Spain

Abstract:

Background/Objectives: The advances in genomic sequencing and the growing knowledge in variant interpretation is facilitating the implementation of genomic screening in different medical areas. Genomic newborn screening is becoming a priority, with several countries implementing programs to evaluate the cost-effectiveness of integrating this determination.Here we present the results of the first 800 newborn genomic screening test performed by a Spanish private laboratory.
Methods: Saliva, cord blood or venous blood sample was obtained from the newborn. Extracted genomic DNA is processed by a capture-based assay and sequenced on a next-generation sequencer (illumina). Exome sequencing data was processed using a proprietary bio-pipeline designed for this intended use. Only childhood onset diseases were analyzed.
Results: Around 8.6% of newborns presented variants in any of the genes included on the panel. The variants detected were related to hematological conditions (68.1%), syndromic (5.8%), Neurologic (5.8%), Cancer (5.8%), Musculoskeletal (4.4%), Metabolic (2.9%), Cardiovascular (2.9%), Immunological (1.5%), Endocrine (1.5%) and Urogenital (1.5%). More than 80% of the infants were newborns (0 months), the mean age was 3 month.
Conclusion: Even though the conditions included in the analysis are considered rare, the subset of pathologies tested could affect up to 8% of the newborns. Early detection lead to an adapted medical management that may involve symptoms prevention or the reduction of the impact of the condition in the child. Besides, rare conditions can take years before being diagnosed, this approach not only enables prevention in most of the cases, but also avoiddiagnostic odyssey.
Grants: None significant to the purpose of this research

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7066F: How to become a genomics GENie: Presenting GenQA's online education platform that offers training and ongoing competency assessment

Authors:

Z. Deans[1], A. Kolozi[1], R. Treacy[1], R. Hastings[2]; [1]GenQA, Edinburgh, United Kingdom, [2]GenQA, Oxford, United Kingdom

Abstract:

Background: There is an increasing need for individuals to demonstrate their competency to perform data analysis and report genomic test results. As variant classification guidance is evolving and variant nomenclature being updated, it is becoming more challenging to keep abreast of changes and evidence up to date knowledge. To support the genomics workforce, GenQA introduced the Genomics Individual Education which builds on GenQA's extensive expertise in providing individual competency assessment already used by 50,000 participants. **Methods:** The GENie platform comprises of four modules for classification of single nucleotide variants (SNVs) and three for copy number variants (CNVs). Participants classify the variants, provide the evidence used and are given real time feedback on their performance. GENie also supports the use of HGVS nomenclature (routine and advanced cases) and the change to the ISCN during 2024 through 10 modules for Constitutional Karyotyping, Neoplasia Karyotyping, Constitutional FISH, Neoplasia FISH, Constitutional Microarrays, Neoplasia Microarrays, Region-Specific Assays-PCR, Region-Specific Assays-MLPA, Sequencing and genome mapping. Similarly to variant classification modules, participants choose the correct nomenclature and are given real time feedback on their performance along with an extensive explanation for all the incorrect nomenclature. **Results:** Overall, over 1,100 individuals completed the scenarios for classification of SNV/CNV whereby a random set of three to six variants were provided for classification each time a participant accessed GENie. All the cases were generated from a vast bank of variants and included prenatal, postnatal, diagnostic and predictive clinical scenarios. In addition, there were over 1,300 participations in the nomenclature scenarios. The range of variant classification and nomenclature and all the GENie trial scenarios will be presented. **Conclusion:** As more genomic data is generated then the Scientists and Clinicians interpreting the results need to be up to date with variant classification frameworks and the use of standardised nomenclature. The uptake in the use of the GENie platform proves there is a demand for continued education and the results obtained by participants evidences that there is a variability in the level of competency to accurately determine the pathogenicity of variants and to apply international standardised nomenclature.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7067F: Increasing the Diversity of the Genomic Workforce : Designing Evaluation Plans to Maximize Impact

Authors:

K. Roberts[1], W. Odbagiorgis[2], A. Sy[3], S. Williams-Blangero[4], E. MANUSOV[4], L. Stewart[5], E. Madden[6]; [1]American Coll. Of Med. Genetics and Genomics, Bethesda, MD, [2]NIH, Bethesda, MD, [3]Univ. of Hawaii at Manoa, Honolulu, HI, [4]Univ. of Texas Rio Grande Valley, Brownsville, TX, [5]Meharry Med. Coll., Nashville, TN, [6]NIH/NHGRI, Bethesda, MD

Abstract:

There is a paucity of interventions that address the lack of workforce diversity in the field of genomics. The National Human Genome Research Institute (NHGRI) recently funded the Diversity Centers for Genome Research (DCGR) Consortium to address this critical gap. The DCGR are based at Minority Serving Institutions and are designed to enhance the diversity of the genomics workforce through a combination of training and opportunities to participate in state-of-the-art genomics research. To assess the impact of the Consortium, we developed and will use common evaluation methods, metrics, and milestones to capture outputs, ensure accurate reporting, facilitate continuous quality improvement, and adjust strategies to meet consortium goals. These metrics will be housed in an internal grantee database, managed by the Consortium coordinating center located at NHGRI. Each funded DCGR Center is comprised of three cores (Administrative, Workforce Development, and Community Engagement) and two or three research projects. Centers developed specific milestones aligned with consortium goals. We utilized content analysis methods to develop a uniform evaluation framework to assess progress towards individual and overall consortium goals and enable comparisons between centers. Content analysis methods are iterative and were applied within each core and across cores, to ensure the common data elements (CDEs) reflected each center's milestones. The analysis occurred over three months with input from NHGRI staff throughout the development of the CDEs. The CDEs evaluate trainee demographics, scientific productivity, genomic curriculum development, and community impact and engagement at each Center. Final CDEs were used to develop survey instruments that will collect monthly progress toward each of the CDEs for the Centers. We are currently in beta testing of the database with investigators from each of the centers pilot testing functionality and reviewing monthly survey questions. Qualitative interviews with pilot testers will be conducted at the end of the beta testing period, and their feedback will be applied for the final iteration of the CDEs to be used in consortium evaluation. The prospective creation of CDEs that align Consortium goals and Center

milestones ensures consistent evaluation of both short- and long-term progress across the DCGR. This innovative approach facilitates ongoing assessment and adjustment of strategic interventions to effectively achieve the Consortium goal of enhancing the diversity of the genomics workforce.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7068F: Limitations in access to genetic counseling training programs across the globe

Authors:

S. Meyer[1], A. Scott[2], C. Gabriel[3], T. Wenger[4]; [1]RWTH Aachen, Aachen, Germany, [2]Seattle Children´s Hosp., Seattle, WA, [3]Univ. Clinic, Heidelberg, Germany, [4]Univ. of Washington, Seattle, WA

Abstract:

The need for genetic counselors is ever increasing across the globe, with widespread increases in genetic testing and decreasing cost. However, the global pipeline for genetic counselors is critically dependent on access to training programs. Genetic counseling training programs are not evenly distributed across the globe, posing barriers to the future genetic counseling workforce. The purpose of this study was to determine the number and distribution of genetic counseling training programs across the globe.
Data was collected from publicly available websites and confirmed with different Societies including NSGC and TAGC.
There is substantial inequity in access to genetic counseling training programs across the globe. There are 56 accredited programs in the United States of America, while the rest of the world combined has only 67 programs. Within the continents themselves, disparities persist, with only a few countries allowing limited access to genetic counseling programs. One major barrier for international students who wish to pursue training in genetic counseling in the United States is financial, with most American universities requiring the cost to be paid upfront with limited or no access to financial support. Additional barriers include the highly competitive nature of genetic counseling training programs and variable recognition of undergraduate training in other countries, visa issues, and different processes for clinical rotations and examinations, making it challenging for individuals who train in one country to practice in another.
The current uneven distribution of genetic counseling programs hinders prospective students' access to this career path, which also results in an uneven distribution of genetic counseling

professionals. This systemic inequity compounds the limited pool of genetic counseling professionals in countries outside of North America. There was a notable absence of international online training programs. The creation of accessible accredited international online programs would allow students across the globe to benefit from a more universal training experience, provide more access equity to certification, and provide more prospective students worldwide with opportunities to pursue a degree in genetic counseling. There are substantial barriers to the creation, sustainability, and funding of such programs. However, by reducing barriers of relocation and culture adaptation, online programs could lead to more successful outcomes and a more representative genetic counseling workforce with individuals from all the countries in which genetic counseling services are needed.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7069F: Modernizing a comprehensive measure of genetic literacy

Authors:

L. Barna, G. Ramirez-Renta, C. Gunter; NIH, Washington, OR

Abstract:

Measuring genetic literacy (GL) is a complex task, and current methods to quantify it lack modernity and consistency. This is of particular importance as research shows the United States general population is lacking critical GL skills. Higher GL levels are needed as genetic and genomic medicine moves to the forefront of care, as GL levels correlate with confidence and autonomy in medical decision-making. The Genetic Literacy Survey (GLS), originally developed by the NHGRI's Social Behavioral Research Branch in 2013, has measured GL in a variety of contexts and through a variety of lenses. It is the only commonly used GL measure to include knowledge comprehension as a cognitive domain, despite this domain generally being cited as an important component of GL. Our previous results have shown within the general population, the average GL score for conceptual knowledge was 9.05 out of a possible 17, which is lower than ideal levels. We have also shown that populations involved in genetic research have only a slightly higher GL score, and that a number of social and psychological factors have direct bearing on GL scores, including education, religious beliefs, and feelings on control and responsibility for medical conditions. As genetics education shifts from Mendelian concepts to more complex conditions and our understanding of gene x environment models improves, new questions are needed to accurately gauge GL levels amongst today's adult general population. In this study, we updated the GLS measure to reflect the rapid advancement of genomic research

while prioritizing inclusivity and community preferences. We used multiple statistical validation tests on a general population sample of 1000 individuals, representative of the US population: item response theory and confirmatory factor analysis to examine the validity of specific questions; correlation coefficients and Cronbach's alpha testing to examine other aspects of validity; and correlation matrices to visualize any directional relationships found. The modernized GLS is available for use and will be given to large US samples in the near future, including an examination of urban and rural needs for improved GL.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7070F: Online Support for Individuals at Risk of Genetic ALS: A Digital Platform for Education, Counseling, Testing, and Peer Support

Authors:

P. Short, P. Wicks, L. Wahlstrom-Edwards, H. Holt; Sano Genetics, London, United Kingdom

Abstract:

Background: Individuals at risk of genetic ALS face many challenges; diagnosed patients may feel guilt about passing on their genes to their descendants, while asymptomatic individuals may experience worry about developing ALS. However, there is inconsistent access to education, genetic counseling, and testing in the US and globally. While pre- and post-test genetic counseling are important, there are not enough genetic counselors, necessitating novel approaches. Methods: With the support of scientific advisors and people at risk of ALS, we developed a digital research platform to offer education, counseling, testing, peer support, and connections to research for those at risk. Initially deployed in the United States and United Kingdom in both English and US Spanish, the platform features a dynamic flow of content modules. All participants are offered a “Genetics 101” education program written by a genetic counselor, as well as signposting to peer-led social support. For those who want to understand their personal risks of developing ALS (if pre-symptomatic) or understand the cause of their condition (if already diagnosed), participants are also offered 1:1 pre-test genetic counselling via videoconference or telephone. Participants who desire DNA testing are provided with an at-home saliva collection kit for the completion of a 40-gene ALS panel on an exome backbone plus targeted assays for repeat expansions (PCR amplification). Genetic results and their implications are returned via the genetic counselor at a post-test session, all at no cost to participants. Those with VUS will have their results reviewed on an annual basis and be recontacted in the event of a change in classification. A series of brief psychological distress

screeners (PHQ-9, GAD-7, Hopelessness, perceived benefits / regrets) is fielded longitudinally at baseline and at key milestones to describe the impact of the program. Individuals can share these results with their healthcare providers. Results: During the first 6 months of the program in the US (English only), the platform provided genetic education to 60 individuals, provided genetic counseling to 21, is in the process of genetically testing 18, and received prior positive genetic test reports uploaded for 11 individuals. Conclusions: Genetic forms of ALS affect families far beyond local boundaries, and with the support of the community we will develop a sustainable platform to offer education, support, and fellowship to those in need.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7071F: Parents' responses to their children's PRS-based genomic risk results ⭐

Authors:

J. Purcell[1], I. Allen[2], S. Morgan[1], C. Nelson[2], M. Myers[2], C. Prows[3], P. Maripuri[4], S. Terek[5], M. Harr[1], J. Connolly[6], J. Lynch[7], I. A. Holm[8], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [3]Children s Hosp Med Ctr, Cincinnati, OH, [4]The Children's Hosp. of Philadelphia, Philadelphia, PA, [5]The Children s Hosp. of Philadelphia, Philadelphia, PA, [6]Children s Hosp. of Philadelphia, Philadelphia, PA, [7]Univ. of Cincinnati, Cincinnati, OH, [8]Boston Children s Hosp, Boston, MA

Abstract:

Background: The eMERGE (electronic MEdical Records and GEnomics) network is returning Genome Informed Risk Assessments (GIRAs) to parents or guardians of children ages 3-17 years of age for asthma, obesity, type 1 diabetes, and type 2 diabetes. Polygenic Risk Scores (PRS) are the source for determining risk status (high or not high) based on study-defined thresholds for each condition. High-risk (HR) GIRAs also contain family history information provided by a parent, clinical variables extracted from the electronic health record (EHR), and healthcare recommendations. While all participants have their GIRA returned passively through the EHR, those who are HR meet with trained study staff to discuss results. Understanding parents' perception of their children's disease risk and its subsequent impact on health outcomes is important before expanding genomic risk assessments into clinical care.
Methods: Parents of pediatric participants enrolled in the eMERGE IV study from the Children's Hospital of Philadelphia and Cincinnati Children's Hospital Medical Center were invited to participate in a semi-structured interview 3-6 months after disclosure of results. Participants

were invited based on risk (HR vs. not HR) and disease status (affected vs. not affected) for asthma, obesity, type 1 diabetes, and type 2 diabetes. Interviews focused on the parents' experience with, and understanding of, their child's GIRA.
Results: Results from 34 interviews showed that all parents of HR participants understood that GIRAs are used for risk screening and are not deterministic, and many felt empowered to make the recommended lifestyle changes. However, some parents of not HR participants expressed confusion due to a personal or family history of the respective condition. All parents stated they followed the recommendations for diet and exercise when their child was at HR for type 2 diabetes and/or obesity, regardless of their child's current BMI. There was a common misconception by parents of children who were HR for type 1 diabetes that diet and exercise could lower the risk. Parents of children who were HR for asthma without a diagnosis of asthma often did not recall the recommendations to reduce asthma risk in their child. Cited barriers to uptake of recommendations were lack of time, lack of cooperation from their child, and lack of extended family support. Potential facilitators mentioned were a detailed plan from the doctor and access to affordable, healthy food.
Conclusion: The results highlight the potential utility of PRS-based GIRA reports for four pediatric conditions. Addressing these findings has the potential to improve patient understanding and health outcomes.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7072F: Piloting anew resiliency training program to support biomedical graduate students: The BioGREAT curriculum

Authors:

D. Dluzen, C. Neely, P. Perez Mejias, C. Hunt, N. Taylor, L. Ledebur, T. Inoue; Johns Hopkins Univ., Baltimore, MD

Abstract:

Recent studies have shown that mental health concerns and stress are on the rise among biomedical doctoral students. There is an essential need to further support doctoral student mental health and resiliency challenges given the rigors of graduate training. To better support the trainees of our institution's Biochemistry, Cellular and Molecular Biology (BCMB) PhD graduate program, we developed and piloted a workshop curriculum to promote wellness and resiliency among current BCMB students. The BioGREAT (Biomedical Graduate REsiliency and Adaptability Training) curriculum uses a flipped-classroom approach to support diverse

learning styles and promote in-depth understanding of workshop content and discussed strategies. We identified three time points to offer a workshop to proactively prepare students as they matriculate through the program. The first workshop, 'Reframing Setbacks and Imposter Syndrome', was designed for first-semester, first-year BCMB students. This workshop helped students recognize the challenges of graduate school, identify and utilize effective time management strategies, and develop and practice ways to reframe potential setbacks into growth experiences. The second workshop, 'Self-Advocacy and Communicating Needs', was piloted by our 2nd year BCMB students in their second semester. This workshop taught students different strategies for improving their self-advocacy skills, how to navigate power dynamics in the laboratory or with mentors, and how to build a mentoring network to support their scientific and career development. The final workshop, 'Designing Your Path Forward', was piloted by 4th year students. They learned how to use a backwards design approach to develop a graduation plan and use negotiation strategies to feel empowered when communicating their needs with their committee or mentors. Through these workshops, we observed student gains in their confidence and ability to navigate and reframe setbacks during school, as well as their ability to develop and set actionable research and mentorships goals. These three workshops form the BioGREAT curriculum that all BCMB students will take in the coming years. Importantly, BioGREAT; 1. Is adaptable to support the needs of biomedical doctoral programs, including genetics-focused training; 2. Can be scaled to support programs of different sizes; and 3. Leverages assessment and evaluation tools allowing for feedback and better alignment with department goals. In summary, this presentation will overview practical implications, lessons learned, and acknowledge encountered constraints, paving the way for replication in other contexts.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7074F: Policymaking in the Real-World: Assessing Key Informant Perspectives on Prenatal Gene Editing

Authors:

R. Isasi[1], M. Allyse[2], Y. Bombard[3], M. Horstmann[4], H. C. Howard[5], A. Kokayi[6], M. Michie[7], K. Ormond[8], K. Riggan[9], J. Yap[9], R. Feys[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Mayo Clinic, Jacksonville, FL, [3]Univ of Toronto, Toronto, ON, Canada, [4]Case Western Reserve Univ., Cleveland, OH, [5]Lund Univ., Lund, Sweden, [6]Mayo Clinic, BioMed. Ethics Res. Program, Rochester, MN, [7]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [8]Stanford Univ and ETH-Zurich, Zurich, Switzerland, [9]Mayo Clinic, Rochester, MN

Abstract:

It is predicted that before long, research on prenatal gene editing will be introduced, paving the way for its eventual clinical translation. With these expectations, it is critical to proactively consider the values that should form the foundational basis for policy and governance in this area. To assess the socio-ethical and scientific arguments justifying a given policy approach, we are conducting semi-structure key informants' interviews with professionals (e.g., clinicians, scientists, bioethicists) with experience in policymaking (e.g., laws, professional guidelines) in areas germane to potential prenatal gene editing applications. We use examples of real and proposed science governance mechanisms to explore individual attitudes towards scientific innovation, governance, risk tolerance, and socio-ethical values. To grasp how policymaking works in the real world and how individual attitudes could shape the direction of science and policy, we identify and compare socio-ethical and moral values stated in gene editing and prenatal therapy policies against professionals' individual perspectives. We consider that engaging affected/interested individuals and communities in collective deliberation is essential for the development of value-based governance for prenatal gene editing.In this presentation, we provide preliminary results of our key informants and comparative policy study. We further outline a value-based governance framework based on such results.
Acknowledgment:PAtient Supported Approaches to Gene Editing - PASAGE, National Human Genome Research Institute (grant R01HG011461)

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7075F: Polygenic-Risk Score Based Identification and Tracking of Pediatric Patients at High Risk for Type 1 Diabetes in the eMERGE Network ⭐

Authors:

J. Connolly[1], J. Roizen[1], P. Maripuri[1], F. Mentch[2], H. Qiu[1], S. Terek[3], M. Harr[4], M. Hess[1], J. Purcell[1], J. Glessner[5], H. Qu[2], J. Thayer[1], D. Karavite[6], B. T. Luo[1], R. Grundmeier[2], H. Hakonarson[6]; [1]Children s Hosp. of Philadelphia, Philadelphia, PA, [2]Childrens Hosp. of Philadelphia, Philadelphia, PA, [3]The Children s Hosp. of Philadelphia, Philadelphia, PA, [4]Children S Hosp. of Philadelphia, Philadelphia, PA, [5]Childrens Hosp. of Philadelphia/PENN, Philadelphia, PA, [6]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Background: Type 1 Diabetes (T1D) is an early-onset autoimmune disorder with a strong genetic predisposition, affecting over 1.6 million individuals in the United States. Advances in polygenic risk score (PRS) development have made it possible to identify individuals at high risk for T1D before clinical symptoms appear. Early identification and longitudinal tracking of high-risk individuals provide critical opportunities for timely intervention and improved disease management.
Objective: To utilize advanced genomic screening methods to identify and longitudinally track pediatric patients at high risk for developing T1D, aiming to better understand disease progression and inform early intervention strategies.
Methods: Pediatric participants (N=~3200) from the Children's Hospital of Philadelphia (CHOP) were recruited through the electronic medical records and genomics (eMERGE) Network. All participants underwent genomic assessment using a validated polygenic risk score (PRS) algorithm to determine their risk of developing T1D. High-risk individuals (N=~100) were identified and are followed longitudinally to track the progression of T1D-related biomarkers, such as HbA1c, C-peptide levels, and autoantibody profiles, over time.
Results: We demonstrate a comprehensive outcomes tracking infrastructure built around T1D risk in high-risk pediatric individuals. Ongoing electronic health record (EHR)-based analysis tracks and monitors changes predictive of disease progression, including clinical onset of T1D among high-risk individuals. The study is also evaluating the correlation between PRS and clinical outcomes to refine risk prediction models.
Conclusion: This study enhances our understanding of T1D pathogenesis and progression in pediatric patients at high genetic risk. By longitudinally tracking these individuals, we aim to identify critical windows for therapeutic intervention, including disease prevention, and optimizing predictive models for T1D. The insights gained will inform early preventive strategies and aims to reduce the incidence and burden of T1D in the pediatric population.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7076F: Population Grouping Schemas related to race and ethnicity in a large, contemporary young American population: Comparing Medical Record, Caregiver Survey, and Genetic Similarity

Authors:

B. Chaiyachati[1], H. Qu[2], J. Catalano[2], J. Xu[2], H. Qiu[2], F. Mentch[2], M. Khan[2], A. Thomas[2], N. O'Connor[2], J. Glessner[1], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Background: Population grouping schemas are constructs that reflect complex, past and current socio-political forces. Interrogation of grouping schemas is warranted to support meaningful advances for all persons, particularly considering disparities related to social constructs, as well as generalization across generations with rapidly changing demographic groupings within the U.S. Thus, we sought to describe population grouping schemas in a large, contemporary, young, American population. **Methods:** The Center for Applied Genomics has 64516 participants with electronic medical record (EMR) and Health Survey Data available born after 1980 without multiracial self-identification. Race/ethnicity were extracted from survey data (n=11 categories) and EMR (n=16 categories). Illumina SNP-array genotyping (imputed across HH550/610, 660, 1M, 2.5M, GSA) was used to generate genetically-similar continental groups by probabilistic multi-dimensional scaling (MDS) with reference to 1000 Genomes. "Consistent" groupings were identified by data-driven category mapping as the most common group connection. Additional interrogation was completed with Admixture. **Results:** Limiting to individuals without self-identified multiracial identities, survey to EMR race/ethnicity categories demonstrated a broad range of consistent identification (26% to 97%). The consistent ranges from survey or EMR categories to MDS-genetically similar groups were the same (Survey-to-MDS: 39% to 98%; EMR-to-MDS: 37% to 100%). MDS-probability regressed on birth year showed statistically significant decline since 1980 in the two most frequent MDS groupings (EUR and AFR, e.g., EUR beta=-3.8x10-4, $R^2 < .01$, $p < .001$) with decline trend in all groups. Individuals with any variability across three comparison populations (n=5,592; 8.7%, Survey-to-EMR, EMR-to-MDS, Survey-to-MDS) had greater computed admixture by lower single continental-group maximum assignment (mean 78% vs 85% for participants with consistent groupings). **Conclusions:** Population grouping schemas incorporating race and ethnicity in a large contemporary young American population demonstrated important variability across constructs that warrant continued methodologic advancement to ensure research results are generalizable and robust over time.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7077F: Precision Oncology Workflow Modifications in a Community Cancer Center Led to Greater Adoption of Genetic Counseling Referral and Testing in Patients Diagnosed with Ovarian Cancer ⭐

Authors:

S. Darabi[1], J. Homer[1], L. Richards[1], M. Demeure[1,2], T. Beck[1], A. Mendivil[1]; [1]Hoag Family Cancer Inst., Newport Beach, CA, [2]Translational Genomics Res. Inst., Phoenix, AZ

Abstract:

Background: Approximately 18% of patients with ovarian cancer (OC) harbor a pathogenic/likely pathogenic germline mutation in a known cancer predisposition gene. Guidelines recommending universal genetic testing for ovarian cancer (including fallopian tube and peritoneal carcinoma) were established in 2010; however, *o*nly 33-39% of patients are tested in the US. We evaluated whether the rate of genetic referral and testing improved after educational and process improvement efforts. **Methods:** We evaluated the rate of genetic counseling (GC) referrals and genetic testing for ovarian cancer in our cancer center from 2020 through 2023. We provided physicians and staff with education regarding guidelines and lack of adherence to universal testing. Our gynecologic oncologists (Gyn/Onc) partnered with our GC team to devise and implement internal workflow changes toward facilitating referrals. The GC team now receives medical record alerts on newly diagnosed ovarian cancer patients from the pathology department and notifies the Gyn/Onc office. Our team tracked and submitted appropriate referrals. Additionally, our GCs also dedicate one day a week to the Gyn/Onc medical office to accommodate patients during their regular doctor visits. We analyzed the number of cases referred and assessed pathogenic or likely pathogenic variants. **Results:** In 2020, 56 patients were diagnosed with OC, and 24 (43%) were referred to the GC program, of whom 22 were tested (5 were positive, of which 4 were in *BRCA1/2* genes). In 2021, 59 patients were diagnosed, and 41 (69%) were referred to the GC program, of whom 36 were tested (11 were positive, of which 5 were in *BRCA1/2* genes). After implementing the new workflow in 2022, the referral rate went up to 73% (47 of 64 patients diagnosed were referred). Of these, 37 were tested (4 were positive, 3 of which were in *BRCA1/2* genes). In 2023, 52 patients were diagnosed with ovarian cancer, 41 (79%) were referred, and 35 were tested (4 were positive, 2 of which were in *BRCA1/2* genes). Some patients were not referred because they had already done genetic testing, or they were referred to an institution outside our practice due to insurance requirements. **Conclusion:** Active education and process improvements resulted in an improvement in the rate of GC referral and testing for patients with ovarian cancer and should be implemented more widely. Our future direction is to continue improving our workflow so that all patients who are diagnosed with ovarian cancer are referred to GC so that they and their families can benefit from germline testing.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7078F: Public perceptions of the development and use of polygenic scores for social and behavioral traits

Authors:

J. Cadigan[1], K. M. Meagher[2], A. Long[1], M. J. Kucmanic[3], K. Kuczynski[1], M. Waltz[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Mayo Clinic, Rochester, MN, [3]Univ. of Iowa, Iowa City, IA

Abstract:

GWAS researchers often seek to identify the degree of heritability for a designated phenotype by using data from large-scale genomic databases or biobanks. When the phenotype in question is value-laden or socially sensitive, genetic research can be taken out of context by third parties. One example of unintended third-party use is the 2022 Buffalo mass shooter's invocation and misuse of genetic studies to justify a white supremacist-motivated hate crime. Increasingly, researchers are using methodologies originally used to develop polygenic scores (PGS) to predict risk for medical traits, like coronary artery disease, to measure associations between genomic variants and social and behavioral traits, such as educational attainment and smoking behaviors. This growing line of research is controversial. Proponents of PGS for social and behavioral traits argue the scores will improve understanding of gene-environment interactions and strengthen social science methodology. But others worry about misinterpretation of findings, including how oversimplification might play out in third-party use. Although PGS have low predictive values, they may be seen as deterministic and lead to stigmatization of already marginalized groups. We sought to understand how supportive the public may be of the development of PGS for social and behavioral traits and the potential use of their own personal data to conduct this research. We surveyed two diverse groups: over 400 people who donated to a U.S.-based biobank and over 1000 members of the “general public” whose inclusion in a biobank or other genomic database is unknown. Overall, biobank participants were more likely than the general public to say they trusted researchers to consider the social implications and potential harms of their research. However, both groups were more likely to support research on medical traits than social traits (87% vs 36% of biobank participants, 64% vs 42% of the general public), and be more comfortable with the use of their own data to develop PGS for medical than social traits (80% vs 40% and 52% vs 34%, respectively). We will analyze respondents’ perceptions of PGS by education, age, race/ethnicity, gender, and disability status. These survey results will assist both administrators of biobanks and genetic researchers to anticipate the ethical, legal, and social implications of their work, assess expectations of biobank donors, and to identify opportunities for more participatory research designs and dissemination strategies.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7079F: RareBoost: European Horizon 2020 Project for Rare and Undiagnosed Disease Research Capacity Building by an Interdisciplinary Approach

Authors:

U. Ozbek[1], B. Utz[1], N. Atabey[1], G. Cakan Akdogan[1], K. Diril[1], E. Erdal[1], S. Erkek Ozhan[1], B. Gungor[1], I. Gursel[1], M. Gursel[1], M. Inan[1], S. Kale[1], S. Kalyoncu[1], G. Karakulah[1], Y. Oktay[2], G. Ozhan[1], G. Wingender[1], B. Yilmaz[1]; [1]IBG-Izmir Biomedicine and Genome Ctr., Izmir, Turkey, [2]Izmir Biomedicine and Genome Ctr. / Dokuz Eylul Univ., Izmir, Turkey

Abstract:

In Europe, diseases afflicting less than 1 person in 2000 are defined as rare diseases. In total, there are more than 7000 different rare diseases which together cause a major burden on human wellbeing and the health systems. Over 90% of all rare diseases are currently without an approved treatment and approximately 80% of them are of genetic in origin. However, the research on rare diseases and the development of new diagnostic and therapeutic approaches is often hampered by limited resources, including patient material and biological models. The incidence of rare diseases (RD) increases in populations with a high frequency of consanguineous marriages. Within European Research Area (ERA), Turkey has the highest rate of consanguineous marriages and, consequently, rare diseases are about twice as prevalent in Turkey than in other ERA countries. Izmir Biomedicine and Genome Center (IBG), the largest and best-equipped public biomedical research institute in Turkey, developed a strategic goal to become an internationally recognized research and innovation facility for RDs.
RareBoost is a European Horizon 2020 "ERA Chairs" project awarded to IBG to guide its development towards a rare disease research center of excellence. ERA Chairs projects support research organizations to recruit an outstanding academic researcher (the ERA Chair) to implement structural changes necessary to achieve excellence in a specific field. With the direction of the RareBoost ERA Chair, IBG has established a research team, Rare and Undiagnosed Diseases Platform (RUDiP) and Unsolved/Undiagnosed Patient Board, to drive innovation and research excellence in the field of rare and undiagnosed diseases (RUDs) through a multidisciplinary approach. RUDiP consists of 15 faculty members and research groups with expertise in different areas of basic, translational, technological and industrial research. RUDiP tackles RUDs by integrating genomics and computational biology with functional studies using both in vitro and in vivo models (including organoids, zebrafish, and mouse models), covering the entire patient experience from diagnosis to treatment. This

collaborative effort aims to advance our knowledge of RDs and lead to innovative diagnostic methods and treatment options for affected individuals. IBG-RUDiP is open to international partnerships to collaborate on our mission to combat RDs and improve patient outcomes via scientific discovery.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7080F: RareInsight: From diagnosis to dashboard; innovations in rare disease reporting

Authors:

O. Awe[1], F. Zemzem[2], E. Akurut[3], G. Wiafe[4], K. C. Coetzer[5]; [1]ASBCB, Cape Town, South Africa, [2]Univ. of Monastir, Monastir, Tunisia, [3]Makerere Univ., Kampala, Uganda, [4]Univ. of Cape Town, Cape Town, South Africa, [5]Stellenbosch Univ., Cape Town, South Africa

Abstract:

Following a successful diagnosis, rare disease (RD) patients and their families typically face substantial hurdles in accessing information and support specific to their diagnosis. There is also a growing expectation for patients and non-specialist healthcare providers to comprehend and analyse the outcomes of these genetic tests. At present, there is no established standard for reporting these outcomes, and existing recommendations overlook resource-limited populations where genomic literacy hinders the accurate sharing of genetic test results. Utilizing an interactive dashBoard for reporting these results enhances accessibility, comprehension, and collaboration among patients, clinicians, and researchers in the rare disease field. While dashboards have been tested in various fields of medicine and research, their applicability has not yet been tested in the context of RDs. This research introduces RareInsight, an effort focused on developing an open-source, interactive dashBoard tailored to both clinicians and patients. Its primary aim is to transform complex genetic variant data into simplified and understandable resources to ease the dissemination of genetic test results and encourage future research to fill in identified knowledge gaps. RareInsight was designed using shinydashBoard to process genetic variant data into customizable and interactive reports. The dashBoard allows users to input and view VCF files, ideally from nf-core's raredisease pipeline. It is also equipped with a searchable ClinVar variant summary table with advanced filtering options. User-defined searches from this table give access to resources for patients, researchers, and clinicians, empowering them to explore relevant detailed data and collaborate with peers

for further analysis and interpretation. The development and testing of this dashBoard utilized whole exome sequencing data from patients with skeletal dysplasia. RareInsight offers users various resources for further research and support upon receiving a genetic diagnosis. In conclusion, RareInsight serves as a pilot study in disseminating genetic test results for RD patients through the use of interactive dashboards.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7081F: State Medicaid coverage for genetic counseling with CPT Code 96040

Authors:

K. Stoll; Genetic Support Fndn., Olympia, WA

Abstract:

Background: The American Medical Association (AMA) introduced the Current Procedural Terminology® (CPT) code, 96040 in 2007 for genetic counseling services offered by non-physician providers. The Center for Medicare and Medicaid Services (CMS) categorizes CPT 96040 as a bundled code and does not provide reimbursement for the code individually. Some State Medicaid programs have allocated a stand-alone fee for CPT 96040. Additionally, some states have published policies that explicitly recognize genetic counseling provided by genetic counselors for separate reimbursement under CPT 96040. We conducted a comprehensive review of Medicaid fee schedules and policies across the 50 United States and the District of Columbia to assess coverage for genetic counseling services utilizing CPT code 96040.Methods: We examined published Medicaid fee schedules and published policies from all 50 states and the District of Columbia to ascertain if CPT code 96040 was associated with a fee and to gain insights into the reimbursement landscape for genetic counseling services delivered by genetic counselors. This review was conducted between May and June 2024.Results: Among the 50 states and the District of Columbia, 22 assigned a nonzero value to CPT code 96040. The reimbursement rates ranged from $24.64 to $55.98 per unit, with a median of $36.39 and a mean of $36.47. Furthermore, ten states include specific language in their published policies indicating Medicaid coverage for services provided by genetic counselors.Conclusions: Despite CMS classifying CPT code 96040 as a bundled code, a subset of State Medicaid programs recognizes it with a designated fee in their fee schedules. Moreover, a minority of states explicitly acknowledge Medicaid coverage for genetic counseling services provided by genetic

counselors, indicating variability in policy interpretation and implementation across jurisdictions.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7082F: Summary and insights from the first genetic services clinic in public sector pediatric healthcare in Sindh, Pakistan. A joint collaboration

Authors:

M. Hanif[1], M. N. Ibrahim[2], A. Furqan[3], J. Raza[4], S. A. Ahmed[5]; [1]Dow Univ. of Hlth.Sci., Karachi, Pakistan, [2]Natl. Inst. of Child Hlth., Karachi, Pakistan, [3]Stanford Hlth.Care, Fresno, CA, [4]Sindh Inst. of Child Hlth.and Neonatology, Karachi, Pakistan, [5]Kaiser Permanente, Los Angeles, CA

Abstract:

Introduction: A tele-genetics clinic was started at National institute of Child Health (NICH) in collaboration with US based experts from Pakistani society of Medical Genetics and Genomics in which patients received free of cost genetic counseling services. **Objective:** This report details provision of genetic services and summarize data from multi-gene panel testing at a specific clinic. This study aims to assess the impact of consanguinity on autosomal recessive condition. **Methods:** These weekly zoom based clinic was started at NICH a in 2020 and comprised of pediatricians and trainee genetic counselor from NICH and United States based geneticist and genetic counselor. Patients were referred for a suspected genetic diagnosis based on their clinical presentation primarily from NICH and SICHN, Karachi, Pakistan. This clinic provided genetic counselling services to 37 families, and as part of the process, the designated trainee at NICH got genetic counselling training. Next generation sequencing was performed on 138 patients after proper pre-test genetic counseling. Result were disclosed and post-test counseling was provided and experiences were recorded through exit surveys. **Results:** A total of 208 samples were examined, of which 138 patients were patients. Parental consanguinity was found in 124 affected persons (89.8%), sixty-eight patients (49.2%) had a positive family history. Panel testing through NGS identified pathogenic or likely pathogenic variants in 94 cases (68.11%), while 28 (20.28%) cases had VUS. We successfully obtained 83 confirmed diagnoses (60.14%) from panel testing using next-generation sequencing. These results highlight the effectiveness of panel testing as a first-line approach in suspected genetic diseases, especially in regions with a high prevalence of consanguinity. All of families expressed satisfaction with overall process and usefulness of genetic counseling. Trainee genetic counselor also greatly

benefited from this learning experience and was able to provide genetic counseling to some families on her own. **Conclusions:** This is the first-of-its kind tele-genetic clinical model in a public sector hospital in Pakistan that provided free genetic counseling services to the underserved population along with training of a prospective genetic counselor. Our limited data underscores the significant genetic disease burden and efficacy of panel testing. Data highlights the prevalence of autosomal recessive disorders and role of consanguinity and provides valuable insights for improving genetic services through proper counseling and training.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7083F: Technical and ethical considerations and guidelines for return of RNA-seq results for rare disease research participants

Authors:

R. Ungar, M. Halley; Stanford Univ., Stanford, CA

Abstract:

In the last decade, many large research initiatives have been launched in the United States and globally focused on enrolling undiagnosed patients with the goals of providing individual diagnoses, identifying novel gene-disease associations, and promoting innovation of new methods for genetic medicine. Practices around return of results vary widely within and across research consortia, with some sites returning variants of uncertain significance (VUSs) identified in research settings and others requiring robust orthogonal validation and verification prior to providing any information to a participant or family (Curic et al 2023, Orphanet Journal of Rare Diseases). In such consortia RNA-seq is increasingly being used as a tool for causal gene identification and variant validation, however this is cutting-edge and no comprehensive framework exists for proper use in a diagnostic setting. Developing consistent practices and protocols is essential; however, the existing literature focuses primarily on return of known pathogenic variants and provides limited guidance regarding return of omic results that are not already recognized in clinical practice.

Determining best practices for return of research results in this context requires consideration of multiple factors, including participant (and family) tolerance for uncertainty, the technical certainty of RNA-seq results, and balancing the patient and family desire for information with the potential harms of sharing incorrect information.

To begin to address this gap, we are conducting a narrative literature review and normative analysis to develop guidelines and recommendations for return of results in research with

undiagnosed patients. This work, which is ongoing, draws on the concept of ancillary care obligations as the overarching ethical framework. We identify the range of ethical, scientific and technical challenges encountered in this research context including, for example, challenges in validating a data type that is dynamic in nature and that can be greatly influenced by technical covariates. We present a decision-making matrix to guide researchers in determining when and how results should be disclosed to patients and families. Finally, we identify key gaps in knowledge needed to evaluate the appropriateness of these guidelines, including evidence of potential benefits and risks of sharing different types of results with participants and families.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7084F: The Clinician-reported Genetic testing Utility InDEx for Newborn Screening (C-GUIDE NBSeq): Development and content validity

Authors:

S. Shickh[1], S. Luca[1], K. Fooks[1], E. Poole[1], A. Pichini[2], J. Ziff[2], K. Stone[2], B. Xiao[1], D. Bick[2], P. Chakraborty[3], W. J. Ungar[4], R. Z. Hayeems[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Genomics England, London, United Kingdom, [3]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [4]The Hosp. for Sick Children Peter Gilgan Ctr. for Res. and Learning, Toronto, ON, Canada

Abstract:

Introduction: Genome sequencing (GS) is emerging as a tool for expanding the scope of newborn screening (NBS) and has been integrated into research initiatives in several countries. Evidence of utility is required to assess the acceptability and performance of this strategy and guide future policy and funding decision-making. We previously developed and validated a consensus-based measure of clinical utility called the Clinician-reported Genetic testing Utility InDEX (C-GUIDE), for indication-based genetic testing in rare disease pediatric populations. A C-GUIDE measure tailored to the unique parameters of NBS is warranted.We aimed to develop an expert-informed, consensus-based version of the C-GUIDE tool to quantify the clinical utility of genomic NBS (*C-GUIDE NBSeq*).**Methods:** Using the original C-Guide as a starting point**,** we identified putative domains of utility relevant to NBS and developed a preliminary 11-item *C-GUIDE NBSeq.*Cognitive interviews were conducted with clinicians involved in genomic NBS, to assess the relevance, wording, feasibility and timepoint of *C-GUIDE NBSeq* administration. Interviews were transcribed verbatim. We quantified the number of items clinicians accepted,

modified, or rejected. A revised version of the tool was generated based on this feedback.Subsequently, interview participants completed a content validity questionnaire on the revised tool. Content validity scores for relevance and clarity were calculated for each item; a score of 0.7 was the threshold for item inclusion. A third iteration of *C-GUIDE NBSeq* was developed for a Delphi consensus process.**Results:** Twenty-two clinicians participated in the cognitive interviews. Participants endorsed the majority of items: 8/11 items were as acceptable "as-is" by at least 50% of respondents, two items required modification, two were combined, and one was rejected.Fifteen of the twenty-two participants completed the content validity questionnaire, which led to minor revisions in the wording. A Delphi consensus process is underway for the 10-item tool.**Conclusions:** No standardized methods currently exist for capturing the clinical utility of genomic NBS. Our preliminary, expert-informed, *C-GUIDE NBSeq* has undergone two rounds of revision informed by interviews and content validity analysis and provides a novel option for quantifying the clinical utility of genomic NBS. Following the expert consensus process and validation testing, *C-GUIDE NBSeq* will be available for licensed use.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7085F: The feasibility of returning reclassified genetic test results through electronic patient portals

Authors:

S. Makhnoon, M. Lee, T. Prasad, C. Jennifer, C. Kao, A. Badalamenti, C. Mai, J. Mersch; UT Southwestern Med. Ctr., Dallas, TX

Abstract:

Introduction: Over a 5-10-year period, between 6% and 15% of germline variants undergo reclassification. Recontact after reclassification is potentially beneficial to patients and can facilitate up-to-date medical management. As the frequency of actionable and inactionable reclassifications increases, a primary challenge today is efficiently relaying them to patients. A persistent question in the field is: will unidirectional, low-touch modalities of recontact (e.g., electronic patient portal message) suffice? We attempted to understand the feasibility of using patient portals in returning reclassified results. **Methods:** We conducted a retrospective study of enrollment in and use of MyChart, a patient poral for Epic electronic medical record system, among patients with reclassified genetic variants seen at a comprehensive cancer center. We identified patients using MyChart by extracting data based on cancer genetics department-type

appointments with an encounter between January 1, 2007 (when MyChart was first available at the institution) and December 31, 2023. We queried three different MyChart data tables for patient data, access patterns, and message actions. Epic demographic data were matched to MyChart data using unique patient medical record numbers. **Results:** In total, 801 patients with reclassified variants (79.2% female, mean age 53 years) were included: 28 had multiple reclassifications and 52 (6.6%) variants were clinically actionable. Over time, the proportion of patients with MyChart accounts increased from 68% to 84%. In total, 551 (78%) patients were enrolled in MyChart with active accounts and most (85%) were frequent users (number of MyChart use ≥ the median value of 361 log-ins). The most common patient portal actions were scheduling appointments (38.1%), updating medical history (26%) and viewing and responding to clinic messages (16.2%). Increased portal use was associated with older age but not with sex, or race/ethnicity. Return of reclassified results (total and clinically actionable) through MyChart did not alter patients' portal use behavior with regards to number of messages sent from patient to provider (total or to seek medical advice). **Discussion:** Among patients with reclassified variants who receive care at a cancer center, portal use is prevalent and frequent but inconsistent. Portal use more than doubled in recent years and was shown to serve as an efficient modality for returning reclassified results, especially clinically inactionable results. Returning reclassified results through patient portals is a feasible option for healthcare systems to manage the increasing number of reclassified results.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Friday Session

Board 7086F: Women's healthcare providers' knowledge and practices surrounding fragile-X associated primary ovarian insufficiency (FXPOI)

Authors:

E. Allen[1], A. Singleton[1], B. Poteet[2], E. Peery[1], H. Hipp[1], S. Sherman[1]; [1]Emory Univ., Atlanta, GA, [2]Northside Hosp. Cancer Inst., Atlanta, GA

Abstract:

An estimated 1 in 178 women carry a premutation (PM), putting them at risk for fragile X-related conditions, including primary ovarian insufficiency (FXPOI), tremor/ataxia syndrome (FXTAS), and neuropsychiatric disorders (FXAND). We have previously conducted semi-structured qualitative interviews in women with a PM to better understand their healthcare experiences. Our study of 24 women explored the diagnostic experience of PM women with FXPOI. Common themes that were identified included hopes for broader clinician awareness of

FXPOI, clearer guidelines for clinical treatment, and proper fertility workups to expand reproductive options prior to POI onset. Participants also spoke of difficulty finding centralized sources of care. Overall, our results indicated a lack of optimal care of women with a PM. In the current study, we have investigated the knowledge gaps about FXPOI among women's healthcare providers. The objective of this study was to pinpoint these gaps and explore demographic factors influencing FXPOI knowledge in women's healthcare practitioners. An online survey was used to assess familiarity with primary ovarian insufficiency and FXPOI knowledge among 107 women's healthcare providers and 14 medical students in the United States. Knowledge Scores, ranging from 0 to 16, were assigned, and demographic data, including healthcare provider type, specialty, and genetics exposure in education or training, were collected. Participants scored an average of 6.92 (±2.19) out of 16 (42%) despite 88% of participants reporting genetics exposure in training. Maternal fetal medicine (MFM) and reproductive endocrinology (REI) providers significantly outperformed general obstetrics and gynecology (OBGYN) practitioners ($p = 0.0186$ and 0.0125, respectively). Participants with a genetic counselor in their clinic scored 8% higher ($p = 0.0083$) than those without. Additionally, medical school graduation year was a significant predictor for knowledge score ($p = 0.0397$). This study underscores limited FXPOI knowledge among women's healthcare providers, aligning with patient reports. Notably, medical specialty and the presence of a genetic counselor impacted knowledge, emphasizing the urgency for broader education in the field, particularly among OBGYNs, the initial point of contact for patients with FXPOI symptoms.

Session Title: Genetic Therapies Poster Friday Session

Board 8081F: A drug repurposing screen reveals dopamine signaling as a critical pathway underlying potential therapeutics for the rare disease DPAGT1-CDG

Authors:

H. Dalton[1,2], N. J. Young[2], A. R. Berman[2], K. A. Patterson[2], S. J. Peterson[2], H. D. Evans[2], C. Y. Chow[2]; [1]Univ. of Kansas, Lawrence, KS, [2]Univ. of Utah, Salt Lake City, UT

Abstract:

Glycosylation involves essential biological pathways that impact the sugar modifications of proteins. Mutations in glycosylation genes underlie Congenital Disorders of Glycosylation (CDGs) - ultra-rare disorders that cause debilitating, multisystemic disorders. CDGs have few treatment options, and low patient populations make clinical trials challenging. Thus, there is a need for alternative approaches to find new therapies. One such alternative is drug repurposing screens

which use libraries of FDA-approved small molecules to potentially accelerate treatment development. Here I will present a drug repurposing screen to identify new candidate drugs for the rare disease DPAGT1-CDG.
DPAGT1-CDG is a multisystemic disorder with symptoms that include facial dysmorphia, seizures, hypotonia, and early mortality. It is caused by mutations in the gene *DPAGT1*, which encodes for the first essential enzyme in N-linked glycosylation. I use a model of DPAGT1-CDG in *Drosophila* that uses *DPAGT1* RNAi in the fly eye to cause a small, rough eye phenotype. This model can be screened for new treatments by quantitatively measuring improvements to its eye size. To find such treatments, I performed a drug repurposing screen by treating the model with 1,520 small molecules that are 98% FDA/EMA-approved. The top candidate drugs that rescued the flies, resulting in a larger eye, were then validated by dose-response and genetic analyses.
I identified 42 candidate drugs that rescued the DPAGT1-CDG model eye size (Z score ≥ 1.5, 3% hit rate). One of the top rescuing drug categories was dopamine receptor D2 (D2R) antagonists. Of note, gene knockdown of *D2R* via RNAi mimicked the drug effects and strongly improved eye size. Counteracting this, a heterozygous null of the dopamine 1 receptor (*D1R*) worsened eye size and reduced the improvement of *D2R* knockdown. Finally, knockdown of dopamine synthesis and recycling pathways also rescued the *DPAGT1* model. Thus, the loss of dopamine flux, and its subsequent binding to D2R, improved the impairment of *DPAGT1*. Dopamine-affecting drugs could represent a promising new therapy - especially as seizures are common in DPAGT1-CDG patients. I will also describe the validation of acetylcholine-affecting drugs, COX inhibitors (NSAIDs), and ion transporter (NKCC1)-related drugs, as well as the role of a negative hit, an antihistamine. These pathways represent novel biology related to *DPAGT1*, and they may underlie new therapeutic options for DPAGT1-CDG.

Session Title: Genetic Therapies Poster Friday Session

Board 8082F: Adults with acid sphingomyelinase deficiency show continued or sustained improvements in hepatosplenomegaly, respiratory outcomes, and lipid profile after 4 years of olipudase alfa enzyme replacement therapy: Final results of the ASCEND trial

Authors:

R. Gallagher[1], M. Wasserstein[2], A. Barbato[3], R. Giugliani[4], N. Guelbert[5], J. Hennermann[6], C. Hollak[7], T. Ikezoe[8], R. Lachmann[9], O. Lidove[10], P. Mabe[11], E. Mengel[12], M. Scarpa[13], E. Senates[14], M. Tchan[15], J. Villarrubia[16], B. Thurberg[17], N. M. Armstrong[18], D. Peloza[19], U. Wisniewska[20], M. Kumar[19]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Albert Einstein Coll. of Med.,

Bronx, NY, [3]Federico II Univ. Hosp., Naples, Italy, [4]HCPA/UFRGS, Porto Alegre, Brazil, [5]Reina Fabiola Univ. Clinic, Cordoba, Argentina, [6]Univ. Med. Ctr., Mainz, Germany, [7]Amsterdam UMC, Univ. of Amsterdam, Amsterdam, Netherlands, [8]Fukushima Med. Univ., Fukushima, Japan, [9]Natl. Hosp. for Neurology and Neurosurgery, London, United Kingdom, [10]La Croix St Simon Hosp., Paris, France, [11]Clinica Santa Maria, Santiago de Chile, Chile, [12]Clinical Sci. for LSD, SpinCS, Hochheim, Germany, [13]Univ. Hosp. of Udine, Udine, Italy, [14]Istanbul Medeniyet Univ., Istanbul, Turkey, [15]Westmead Hosp., Sydney, Australia, [16]Hosp. Univ.rio Ramón y Cajal, Madrid, Spain, [17]Orphan Sci. Consulting, Newton, MA, [18]Sanofi, Cambridge, MA, [19]Sanofi, Bridgewater, NJ, [20]Sanofi, Warsaw, Poland

Abstract:

Acid sphingomyelinase deficiency (ASMD) is a rare debilitating lysosomal storage disease characterized by pulmonary dysfunction, hepatosplenomegaly, and dyslipidemia. The first disease-specific treatment - olipudase alfa, IV recombinant human ASM - is now approved for non-CNS manifestations of ASMD in >50 countries. During the 1-year primary analysis of the ASCEND clinical trial (NCT02004691, 36 adults with ASMD), olipudase-alfa-treated patients compared to placebo-treated patients (1:1 randomization) had statistically significant increases in %-predicted diffusing capacity for carbon monoxide (DL_{CO}) and decreases in spleen and liver volume. We report long-term outcomes for the 35/36 patients who continued in the open-label extension during which all patients received olipudase alfa (1 placebo patient withdrew during year 1). Overall, 29 patients completed the extension. One withdrawal was due to a related adverse event (rash); all others were unrelated to treatment (other/withdrawal of consent, 3; COVID-19 travel restrictions, 1; pregnancy, 1). All patients underwent olipudase alfa dose escalation (≥14 weeks) to a target dose of 3.0 mg/kg/2 weeks; dose was subsequently adjusted as needed due to adverse events or/and missed infusions. Results (means ± SD) are reported by time on olipudase alfa; baseline values are prior to receiving active treatment. Mean time on olipudase alfa was 4.2 ± 1.0 years; mean compliance was 90% ± 13%. Most patients had at least 1 missed assessment due to travel issues or laboratory closures related to the COVID-19 pandemic. At baseline, mean %-predicted DL_{CO} was 49.6%; spleen volume was 11.5 multiples of normal (MN); liver volume was 1.5 MN; all patients had elevated plasma lyso-sphingomyelin and histologic evidence of sphingomyelin storage in liver. During olipudase alfa treatment, %-predicted DL_{CO} increased in most patients; all patients had reduced spleen and liver volumes, marked reduction in plasma lyso-sphingomyelin and sphingomyelin storage in liver. After 4 years of treatment, mean %-predicted DL_{co} improved by 36% ±28% (n=15); mean spleen volume decreased by 55% ± 9% (n=26); and liver volume decreased by 36% ± 12% (n=26). Early improvements in dyslipidemia and liver function were sustained or amplified with long term treatment. No new safety issues emerged during the trial extension; overall, 98% of treatment-emergent adverse events were mild/moderate, with 1 treatment-related serious adverse event

(extrasystole in patient with medical history of cardiomyopathy). In summary, adults with ASMD showed sustained or continuing improvement in key clinical outcomes after 4 years of olipudase alfa treatment.

Session Title: Genetic Therapies Poster Friday Session

Board 8083F: Advanced Therapeutic Strategies for Fibrodysplasia Ossificans Progressiva: siRNA-Mediated Gene Silencing and sgRNA-Guided Gene Editing Using CRISPR-Cas9

Authors:

V. Sagare; Dublin High Sch., Dublin, CA

Abstract:

Fibrodysplasia Ossificans Progressiva (FOP) is a rare genetic disorder caused by a mutation in the ACVR1 gene, specifically ACVR1-R206H, leading to abnormal bone growth in muscles, tendons, and ligaments. This research explores two therapeutic strategies: gene silencing and gene editing, both using CRISPR-Cas9 technology. Gene silencing aims to reduce the expression of the mutant ALK2 gene using RNA interference (RNAi). Small interfering RNA (siRNA) molecules specifically target the mRNA of the mutated gene, preventing its translation. We will design siRNA sequences, which are delivered to target cells via viral vectors or transfection reagents. This results in the degradation of mutant ALK2 mRNA, reducing the abnormal protein and preventing abnormal bone growth. Key sequences, such as the one at position 4-26 with a GC% of 75.00 and a Tm of 82.34°C, ensure high specificity and efficient binding. Gene editing aims to correct the ACVR1-R206H mutation using CRISPR-Cas9. Single guide RNA (sgRNA) sequences precisely target the mutation site. The Cas9 protein, guided by sgRNA, introduces a double-strand break at the mutation location. The cell's natural repair mechanisms, particularly homology-directed repair (HDR), insert a correct version of the gene. Sequences from gathered data, such as those at positions 102-124 and 106-128, were selected for their high GC content and optimal melting temperatures, minimizing off-target effects and enhancing precision. Several potential sgRNA target sequences were identified for both gene silencing and gene editing. For gene silencing, sequences like the one at position 4-26 with a GC% of 75.00 and a Tm of 82.34°C, and position 93-115 with a GC% of 65.00 and a Tm of 84.61°C, were highlighted. For gene editing, sequences at positions 102-124 and 106-128, with higher GC% and optimal melting temperatures, were selected. FOP presents significant therapeutic challenges due to its genetic basis. Gene silencing and gene editing using CRISPR-Cas9 technology offer promising strategies to address the root cause by reducing the expression of the mutant gene or

correcting the mutation. The specific sgRNA sequences identified provide a foundation for developing targeted therapies to improve the quality of life for FOP patients.

Session Title: Genetic Therapies Poster Friday Session

Board 8084F: An expedited workflow for the development of personalized corrective editing therapies for inborn errors of metabolism ⭐

Authors:

X. Wang[1], A. Quigley[1], S. A. Grandinette[1], M. N. Whittaker[1], **D. L. Brooks**[1], L. Testa[1], E. Kahn[1], W. H. Peranteau[2], M-G. Alameh[2], R. C. Ahrens-Nicklas[2], K. Musunuru[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]The Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Inborn errors of metabolism (IEMs) are rare, devasting disorders arising from pathogenic variants in genes encoding enzymes of key biochemical pathways. The liver plays an important role in the pathogenesis of over 140 IEMs, often failing to metabolize a toxic metabolite that can injure secondary organs, such as brain. Liver transplantation is employed in some IEMs; however, its utility is limited by scarcity of donors and lifelong risk of post-transplant complications. To address the unmet medical need of IEM patients, we aim to develop a master protocol for the rapid development of personalized base-editing and prime-editing therapies for severe, rare hepatic IEMs. To establish an expedited workflow, we initially focused on phenylketonuria (PKU), an IEM caused by pathogenic variants in the phenylalanine hydroxylase (*PAH*) gene. Building on previous work in which we have developed validated *in vivo* base-editing and prime-editing solutions for the recurrent *PAH* R408W and P281L variants—among the most frequent variants identified in PKU patients—we focused on the other most frequent PKU variants. Using a novel platform in which we introduced collinear genomic segments harboring numerous IEM patient variants into HuH-7 hepatocytes, in a matter of weeks we screened and identified base-editing solutions for the *PAH* c.1066-11G>A, R261Q, and R243Q variants—with as good or better efficacy as the R408W and P281L solutions *in vitro*. We subsequently validated the ability of these solutions to correct PKU in humanized mouse models *in vivo* and de-risked the solutions with respect to off-target editing. In parallel, we have applied the same rapid *in vitro* workflow to pathogenic variants from patients with a variety of other liver-centered diseases—including galactosemia, citrullinemia type 1, argininosuccinic acidemia, methylmalonic acidemia, and pseudoxanthoma elasticum—with identification of editing solutions followed by *in vivo* validation in humanized mouse models. To accelerate *in*

vivo testing of the editing solutions, we have begun to use multi-humanized mice with collinear genomic segments harboring human variants inserted into the *Rosa26* safe harbor locus. Our ultimate goal is to develop a platform regulatory framework where IND-enabling studies for a "leader" hepatic IEM therapy open the door for expedited "follower" IND amendments supported by streamlined *in vitro* on-target and off-target analyses, allowing for each new patient with each new variant to be treated with a bespoke corrective editing therapy in a matter of months.

Session Title: Genetic Therapies Poster Friday Session

Board 8086F: From bench to bedside: Lamotrigine therapy demonstrates improvements in attention, anxiety, school performance, and body mass index in a patient with MYT1L-associated neurodevelopmental disorder

Authors:

A. Davis[1], A. Patil[1], E. D. Sperry[2,3], G. Hijazi[2,3], Y. Wu[2,3], R. J. Hopkin[2,3], A. Shillington[2,3]; [1]Coll. of Med., Univ. of Cincinnati, Cincinnati, OH, [2]Div. of Human Genetics, Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [3]Dept. of Pediatrics, Univ. of Cincinnati, Cincinnati, OH

Abstract:

Myelin transcription factor 1-like (MYT1L) is a transcriptional repressor that targets important developmental signaling pathways, including WNT and NOTCH, through the recruitment of epigenetic silencers. *MYT1L* is expressed in nearly all neurons, and its actions result in *in vitro* enhancement of neuronal identity. Pathogenic variants in *MYT1L* contribute to a neurodevelopmental phenotype with features of intellectual disability, schizophrenia, epilepsy, and autism spectrum disorder; with many patients also experiencing hyperphagia, hyperactivity, and speech delay. The genetic syndrome is collectively referred to as *MYT1L*-associated neurodevelopmental disorder.
Mice with haploinsufficient mutations in *MYT1L* also exhibit hyperactivity, social deficits, and obesity. Murine models demonstrate that *MYT1L* knockdown produces neuronal hyperactivity due to upregulation of SCN5A voltage-gated sodium channels. Lamotrigine is a selective inhibitor of voltage-gated sodium channels, stabilizing presynaptic neuronal membranes and inhibiting glutamate and aspartate release. In *MYT1L*-mutant mice, injection of lamotrigine was shown to reverse phenotypic consequences of *MYT1L* deficiency, including anxiety and hyperactivity. Lamotrigine *in vitro* rescued electrophysiological defects in post-mitotic murine

and human neurons. Thus, the use of lamotrigine presents a possible novel therapeutic approach for the management of patients with *MYT1L* pathogenic variants.
This case report details the first clinical application of lamotrigine as a precision therapy for *MYT1L*-associated neurodevelopmental disorder. The patient, a 7-year-old female, was diagnosed with *MYT1L*-associated neurodevelopmental disorder at the age of 2. She displayed a phenotype of non-verbal intellectual disability, attention deficit disorder, and hyperphagia, in addition to subclinical EEG abnormalities. Exome sequencing confirmed a de novo pathogenic variant in *MYT1L*. Following six months of treatment with lamotrigine initiated at the age of 6, teachers, parents, and therapists reported increased efforts at vocalization, increased attention and focus at school, and decreased hyperphagia. Baseline Vanderbilt Assessment Scales were completed prior to lamotrigine initiation, and repeated six months later with demonstrated improvements in attention deficit, anxiety, and cognitive performance. The patient's BMI decreased from the 95th percentile to the 89th percentile 6 months following treatment initiation. This novel case is the first to describe lamotrigine as a possible therapeutic agent for patients with *MYT1L*-associated neurodevelopmental disorder.

Session Title: Genetic Therapies Poster Friday Session

Board 8087F: Generation and characterization of human iPSC derived*NPC1*$^{I1061T/I10161T}$i^{3}Neurons as a model for NPC1 disease.

Authors:

S. Salhotra, I. Kang, C. White, A. V. Prabhu, C. Davidson, C. Wassif, N. Cawley, F. D. Porter; Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, MD

Abstract:

Niemann-Pick disease, type C (NPC) is an autosomal recessive disorder caused by pathological variants in either *NPC1* or *NPC2.* Biochemically NPC is characterized by the accumulation of unesterified cholesterol in endolysosomal compartment of the cell due to impaired function of either NPC1 or NPC2 protein, both of which have a role in cholesterol egress from lysosomes. Approximately 200 pathological variants have been identified in *NPC1.* One of the most common variants is a missense allele, *NPC1*^{I1061T} . NPC1^{I1061T} protein is functional protein although due to misfolding, it stalls in the ER and undergoes intracellular proteasomal degradation. Preliminary cellular-based studies on NPC1 have primarily used patient fibroblasts; however, neuronal cells provide a more suitable model for investigating the neurodegenerative mechanisms of this disease. Thus, the requirement for a cellular system arises to study the neuropathological aspect of the disease and for the discovery of appropriate therapies. Earlier

studies have utilized a human iPSC derived and inducible *NPC1*$^{null/null}$ i^{3}Neuronal model of NPC1. i^3 stands for isogenic, integrated, and inducible where doxycycline inducible expression of neurogenin in iPSC allows for the generation of rapid, synchronized population of homogeneous i^3 neurons in large numbers. This study reports the generation and characterization of a human iPSC derived *NPC1*$^{I1061T/I1061T}$ i^{3}Neuron model system which recapitulates the pathological features of NPC1 including endolysosomal cholesterol storage and alterations in lysosome morphology. Treatment with 2-hydroxypropyl-β-cyclodextrin, a drug that has shown both *in vitro* and *in vivo* efficacy alleviates the NPC1 cellular phenotype.
Thus, *NPC1*$^{I1061T/I1061T}$ i^3 Neuronal cell line provides a suitable model for NPC1 research in conjunction with the existing *NPC1*$^{null/null}$ i^{3}Neuron line and will be utilized in high throughput drug/chemical screens, particularly for identifying potential proteostatic regulators, either to enhance the expression or stabilize the NPC1^{I1061T} protein.

Session Title: Genetic Therapies Poster Friday Session

Board 8088F: Identification of Small Molecule Inhibitors against *Plasmodium falciparum* Bromodomain Protein 1 using Pharmacophore Modeling, Molecular Docking, and Molecular Dynamic Simulation

Authors:

D. Oladejo[1], S. A. Oyegbade[1], D. O. Oladejo[2], T. M. Dokunmu[1]; [1]Covenant Univ., Ota, Nigeria, [2]Babcock Univ., Ilishan Remo, Nigeria

Abstract:

Plasmodium falciparum bromodomain protein 1 (*Pf*BDP1) is a 55 kDa nuclear protein that binds to the acetylated lysine residues of histones and interacts with specific transcription factors to activate the chromatin remodeling and transcriptional machinery. Compounds that bind and inhibit bromodomain proteins are being researched as epigenetic regulation tools and possible malaria treatment approaches. This study, therefore, aims to predict small-molecule inhibitors of *Pf*BDP1. *Pf*BDP1's crystal structure was obtained from the Protein Data Bank. The active sites of the *Pf*BDP1 modeling structure were predicted using CASTp 3.0. Pharmit server was used for pharmacophore modeling of the control ligand and *Pf*BDP1 model structures to generate several compounds for the molecular docking study. AutoDock Vina and Discovery Studio were used for molecular docking and post-docking investigations. The Swiss ADME predictor and OSIRIS Property Explorer were used to predict the toxicity and *in silico* drug-likeness. The

stability of interactions between *Pf*BDP1 and the best-hit chemical was examined using a molecular dynamic simulation run on the Desmond simulation software from Schrödinger LLC. Based on high binding affinity to the active site of *Pf*BDP1, nine (9) top hits with docking scores from -9.8 and -10.7 kcal/mol were selected and subjected to post-screening analysis. Absorption, distribution, metabolism, elimination, and toxicity (ADMET) prediction of the nine (9) top hits showed CHEMBL2216838 had a high drug score of 0.78, drug-likeness of 8.60, and a low-risk toxicity effect. The molecular dynamic simulation of CHEMBL2216838 showed a stable complex conformation in the flexible loop of the *Pf*BDP1 active site. The good, estimated binding energies, drug-likeness, drug score, and stable molecular dynamics simulation interaction observed between CHEMBL2216838 and *Pf*BDP1 active site suggests CHEMBL2216838 as a lead candidate for *Pf*BDP1 inhibition. Pre-clinical studies should be carried out to validate the compound's potential to be regarded as a therapeutic drug.

Session Title: Genetic Therapies Poster Friday Session

Board 8089F: Neuron-targeted lipid nanoparticles for miR-137-3p therapeutic delivery to Alzheimer's disease

Authors:

S. Xu[1,2], S. Zhou[1], S. Huangpu[1], H. Yang[1], X. Zheng[1], D. Wang[1], J. Bi[1]; [1]Second Hosp. of Shandong Univ., Jinan, China, [2]Dept. of Human Genetics, Emory Univ. Med. Sch., Atlanta, GA

Abstract:

Alzheimer’s disease (AD) is a progressive neurodegenerative disease characterized by accumulation of protein beta-amyloid (plaques) outside neurons and twisted strands of the protein tau (tangles) inside neurons in the brain. This leads to significant cognitive impairment and is the most common cause of dementia. Exploring an effective treatment for AD remains a significant challenge, as there is currently no cure for the condition. MiRNA-based treatments have attracted significant attention due to their high therapeutic potential in treating neurodegenerative diseases like AD. In current study, we investigated the mechanisms of miR-137-3p in AD pathogenesis and developed a drug delivery system of lipid nanoparticles with rabies virus glycoprotein 29 (RVG29) and carrying miR-137-3p (RVG29/miR-137-3p@LNPs). We found the miR-137-3p was significantly decreased in AD model and it could inhibit GSK3β expression. We also demonstrated that miR-137-3p ameliorated GSK3β-induced tau hyperphosphorylation and synaptic damage in SH-SY5Y cells, improved Aβ1-42 oligomer-induced damage as well. Our results showed that RVG29/miR-137-3p@LNPs improved cognitive function and moderated tau hyperphosphorylation and neuroprotection in APP/PS1 mice. The

delivery system successfully penetrated the blood-brain barrier (BBB) and targeted neurons, leading to a significant increase in miR-137-3p levels in the hippocampus of APP/PS1 mice. In conclusion, this study verified a new mechanism of action of miR-137-3p in AD pathogenesis and developed an miR-137-3p brain delivery system based on lipid nanoparticles for the AD treatment.

Session Title: Genetic Therapies Poster Friday Session

Board 8090F: Personalized antisense oligonucleotides modulate splicing to restore SDCCAG8 expression in patient-derived fibroblasts with biallelic intronic mutations

Authors:

K. McEntee[1,2], B. L. McCurdy[3], A. Larson[4], E. A. McCourt[5], A. E. Campbell[1], C. G. Pearson[3], J. R. Hesselberth[1,2], S. Jagannathan[1,2]; [1]Dept. of Biochemistry and Molecular Genetics, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]RNA BioSci. Initiative, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Dept. of Cell and Dev.al Biology, Univ. of Colorado Sch. of Med., Aurora, CO, [4]Dept. of Pediatrics, Section of Genetics, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Dept. of Ophthalmology, Univ. of Colorado Sch. of Med., Children's Hosp. of Colorado, Aurora, CO

Abstract:

Bardet Biedl Syndrome (BBS) is a ciliopathy that often results in progressive vision loss, obesity, polydactyly, and other symptoms. A patient presenting with BBS was discovered to have two mutations within 55bp of each other in intron 7 of SDCCAG8 (BBS16). One of the biallelic mutations, c.740+356C>T, is known to cause inclusion of cryptic exon(s) while c.740+301G>A has not been characterized. Personalized antisense oligonucleotide (ASO) therapies, like Milasen, have been used to successfully modulate splicing in the past. We hypothesized that ASOs may be able to modulate the splicing of SDCCAG8 between exons 7 and 8 to prevent the inclusion of cryptic exons and to restore SDCCAG8 expression in this patient.We performed an ASO walk with 20bp ASOs placed every 2bp across each mutation. The ASOs were then screened in patient-derived fibroblasts using RT-PCR assays to assess exon 7 and 8 splicing. After an initial screen, one ASO for each mutation was chosen based on its ability to restore the wildtype splicing pattern. These hits were further investigated through RT-qPCR, RNAseq, and western blotting to confirm restoration of wild-type transcript and protein. Through the ongoing collaboration between the Colorado Children’s Hospital and CU Anschutz, we will further

explore clinical application of the ASO hits as a potential treatment option for this patient and others that share these mutations.

Session Title: Genetic Therapies Poster Friday Session

Board 8091F: Preliminary results from the ongoing APHENITY extension study: Sepiapterin reduces blood Phe with improved dietary Phe tolerance in participants with phenylketonuria

Authors:

N. Smith[1], S. Sacharow[2], D. Bratkovic[3], A. Muntau[4], L. Margvelashvili[5], D. Agladze[5], L. Guilder[6], I. D. Schwartz[7], M. Minami[8], E. Kiykim[9], I. Guerra[10], R. Zori[11], J. Campistol[12], A. Oliveira[13], P. Janeiro[13], F. Rutsch[14], A. Lund[15], H. Mungan[16], S. Vijay[17], A. MacDonald[18], A. Chakrapani[19], T. Opladen[20], M. Tchan[21], K. Ingalls[1], C. Hughes[1], Z. Zhao[1], L. Alexandra[1], K. Somera-Molina[1], **R. Miller**[1], M. Lah[22]; [1]PTC Therapeutics, Inc., Warren, NJ, [2]Boston Children's Hosp., Boston, MA, [3]PARC Clinical Res., Adelaide, Australia, [4]Univ. Med. Ctr. Eppendorf, Hamburg, Germany, [5]Pediatric Surgery Ctr., Tbilisi, Georgia, [6]The Hosp. for Sick Children, Toronto, ON, Canada, [7]Hosp. de Clinicas de Porto Alegre, Porto Alegre, Brazil, [8]Hosp. Das Clinicas Da Faculdade De Med. De Ribeirao Preto, Ribeiraro Preto, Brazil, [9]Istanbul Univ.-Cerrahpasa, Cerrahpasa Med. Faculty, Istanbul, Turkey, [10]PanAmerican Clinical Res., Guadalajara, Mexico, [11]Dept. of Pediatrics, Univ. of Florida Coll. of Med., Gainesville, FL, [12]Neuropaediatrics Dept., Hosp. Sant Joan de Déu, Univ.t de Barcelona, Barcelona, Spain, [13]Centro Hosp.ar Univ.rio Lisboa Norte, Lisbon, Portugal, [14]Dept. of Gen. Pediatrics, Muenster Univ. Children's Hosp., Muenster, Germany, [15]Ctr. for Inherited Metabolic Diseases, Dept.s of Pediatrics and Clinical Genetics, Copenhagen Univ. Hosp., Copenhagen, Denmark, [16]Cukurova Univ.si Tıp Fakultesi Balcalı Hastanesi, Adana, Turkey, [17]Paediatric Inherited Metabolic Disease, Birmingham Women's and Children's, Birmingham, United Kingdom, [18]Dept. of Dietetics, Birmingham Children's Hosp. NHS Fndn. Trust, Birmingham, United Kingdom, [19]Great Ormond Street Hosp. for Children NHS Fndn. Trust, London, United Kingdom, [20]Heidelberg Univ., Med. Faculty Heidelberg, Ctr. for Pediatric and Adolescent Med., Div. of Pediatric Neurology and Metabolic Med., Heidelberg, Germany, [21]Dept. of Genetic Med., Westmead Hosp., Sydney, Australia, [22]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med. Indianapolis, Indianapolis, IN

Abstract:

Objectives: The Phase 3 APHENITY trial (NCT05099640) was a global, two-part, registration-directed study evaluating the efficacy and safety of sepiapterin in a broad phenylketonuria (PKU)

population. Upon completion, participants could enroll into the APHENITY open-label extension study (NCT05166161). Here, we describe preliminary results from the ongoing extension study, which assesses the safety of sepiapterin and its effect on dietary Phe tolerance.
Methods: Participants receive sepiapterin once daily for ≥12 months. Those with mean blood Phe <360µmol/L from Month 1 undergo a concurrent assessment of dietary Phe tolerance. For a 26-week period, mean blood Phe is assessed along with dietary Phe consumption; changes in prescribed Phe are allowed. Participants with mean blood Phe ≥360µmol/L continue treatment but without Phe tolerance assessment.
Results: As of September 22, 2023, 104 participants (median age, 14.0 years [min,max: 2,54]) received sepiapterin treatment. Dietary Phe tolerance assessments were performed in 73 participants. The least-square mean change (95% CI) in daily dietary Phe consumption (baseline to Week 26) was 54.6 mg/kg/day (44.2,64.9). An ~3.3-fold increase from baseline (24.4 mg/kg/day) in mean daily Phe consumption was achieved at Week 26 (79.3 mg/kg/day). Overall, sepiapterin showed a favorable safety profile and was well tolerated in the study.
Conclusions: Preliminary data from the ongoing APHENITY extension study demonstrate continued clinical benefit and a consistent safety profile with sepiapterin. The gradual increase in dietary Phe consumption while maintaining blood Phe within target levels suggests sepiapterin may allow a more liberalized diet in patients with PKU.

Session Title: Genetic Therapies Poster Friday Session

Board 8092F: Rescuing *KANSL1* Haploinsufficiency as a Therapeutic Strategy for Koolen-de Vries Syndrome

Authors:

D. Murphy[1,2], E. Fathi[1], H. Mefford[1]; [1]St. Jude Children's Res. Hosp., Memphis, TN, [2]Rhodes Coll., Memphis, TN

Abstract:

Background: Koolen-de Vries syndrome (KdVS) is a rare autosomal dominant neurodevelopmental disorder largely caused by a de novo heterozygous deletion of the *KANSL1* gene resulting in haploinsufficient protein expression. *KANSL1* encodes the primary subunit of the KAT8 regulatory NSL complex, a system that is involved in chromatin remodeling and gene regulation across the genome. KdVS is characterized by intellectual and developmental delays, hypotonia, and congenital anomalies in cardiac, brain, and renal systems. We hypothesize that increasing the amount of KANSL1 protein by targeting the remaining functional copy of *KANSL1* is a potential therapeutic approach for KdVS. **Methods:** We utilized a

catalytically inactive CRISPR-dCas9 system to activate *KANSL1* transcription and subsequently increase protein production in a heterozygous *KANSL1* knockout SF126 (human glioblastoma) cell line. We tested two different transcriptional activators, VP64 and VPR, bound to dCas9, for *KANSL1* upregulation. We designed five guide-RNAs (gRNA) targeting different locations of the *KANSL1* transcriptional start site (TSS) and tested the efficiency of each gRNA in the *KANSL1* haploinsufficient SF126 cell line stably expressing the dCas9-VP64/VPR system. **Results:** Using RT-qPCR and western blotting, we demonstrate increased *KANSL1* expression in this cell line approximately two-fold. Currently, we are testing the efficacy of selected gRNAs in patient-derived iPSCs stably expressing dCas9-VPR system. We will investigate the effect of *KANSL1* upregulation on downstream pathways in patient-derived iPSCs using RNA-sequencing. Furthermore, we will verify the efficacy of selected gRNAs in patient-derived 2D and 3D neurons and measure changes in synaptic activity using multielectrode array. **Conclusions:** These findings suggest that CRISPRa-mediated upregulation of *KANSL1* may be a potential therapeutic strategy for KdVS. Further studies are ongoing to validate these results in patient-derived cells and explore the effects on downstream pathways.

Session Title: Genetic Therapies Poster Friday Session

Board 8093F: Rescuing lethality in the Homozygote *Vcp*$^{R155H/R155H}$mouse model provides a robust disease model for testing AAV gene therapy.

Authors:

A. Shmara[1], L. Weiss[1], V. Boock[1], P. Pal[1], C. Cheng[2], V. Kimonis[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]UC Irvine, Irvine, CA

Abstract:

Pathogenic variants in Valosin Containing Protein (*VCP*) gene cause a unique autosomal dominant multisystem proteinopathy (MSP1) characterized by inclusion body myopathy, Paget disease of bone and frontotemporal dementia, and amyotropic lateral sclerosis. The mechanism has been shown to be gain of function of the *VCP* missense variant. The knock-in VCP mouse model carrying the common *R155H* mutation includes several of the features typical of the human disease. The VCP $^{R155H/R155H}$ homozygous mice exhibit progressive weakness and accelerated pathology prior to their early demise, typically before 3 weeks of age. While homozygous VCP $^{R155H/R155H}$ mice die before the weaning age when kept on the normal chow diet, lipid-enriched chow can rescue lethality by 50% and over 50% of mice can survive into young adulthood. We have also been able to rescue lethality of the homozygous mice by removing healthy litter mate pups and delayed weaning of the VCP$^{R155H/R155H}$ homozygous pups.

This strategy results in improved survival from 3 weeks to over 20 weeks of age. Using this robust mouse model, we tested the muscle specific AAV variant, AAVMYO as a vector for delivery of gene payloads to muscle tissue. Treatment with AAVMYO vector carrying micro-RNA targeting VCP, with or without wild-type VCP replacement, demonstrates improvement in the muscle pathology in VCP $^{R155H/R155H}$ mice. Conclusion: Our strategy of high fat diet and delayed pup weaning and removing littermates is helpful in providing a useful disease model. Downregulating the VCP allele with AAVMYO vector carrying anti VCP micro-RNA is associated with improvement in the muscle pathology in the VCP$^{R155H/R155H}$ mice. Success in these preclinical AAV studies holds potential for promising therapeutic benefits in patients with VCP disease and related diseases.

Session Title: Genetic Therapies Poster Friday Session

Board 8094F: Restoring CHAMP1 expression and action in modeling studies using AAV9-mediated gene therapy

Authors:

M. Matos[1], B. Strenkowski[1], M. Bakay[1], H. Hakonarson[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med. Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Chromosome alignment maintaining phosphoprotein 1 (CHAMP1) is crucial for maintaining kinetochore-microtubule attachment during cell division and regulating chromosomal segregation, an essential process for brain development. Mutations in the CHAMP1 gene are associated with global developmental delay, intellectual disability, hypotonia, and dysmorphic features. Given that CHAMP1 deficiency results from a single gene alteration, it represents a promising target for gene therapy. Here, we explored a gene therapy strategy to restore CHAMP1 expression using a novel mouse model carrying a 35bp deletion in CHAMP1 exon 2, created via CRISPR/Cas9 gene editing. We successfully characterized the model based on pronounced phenotypes, including motor development delays, impairments in memory and learning abilities, providing a robust platform for investigating CHAMP1 and testing therapeutic interventions.To initially evaluate the feasibility of AAV CHAMP1 gene therapy, we used the SH-SY5Y cell line. CHAMP1 was effectively knocked down using CRISPR-Cas9 system, and the knockdown cells were treated with AAV9 (Adeno-Associated Virus) vectors carrying a wild-type copy of human CHAMP1. Gene and protein expression levels were assessed at 24, 48, and 72 hours post-treatment. The treatment successfully restored CHAMP1 expression at both gene and protein levels that was maintained up to 72 hours post-treatment without adverse events.

Treatment of neonatal Champ1 mutant mice with the CHAMP1 AAV9 vector is underway. The therapeutic efficacy of the gene therapy is being evaluated through a series of behavioral tests on mice aged 8-10 weeks, divided into two batteries: motor and reactivity, and cognitive. The first battery will assess motor delay, muscular strength, and overall health using open field, rotarod, and catwalk gait tests. The second battery will evaluate learning ability, memory, and social behavior using novel object recognition, acoustic startle response (ASR), and social choice tests. Our study highlights the potential for an effective gene therapy for CHAMP1 disease. The successful restoration of CHAMP1 expression in cellular models coupled with in vivo studies pave the way for developing targeted treatments for individuals affected by CHAMP1 disease.

Session Title: Genetic Therapies Poster Friday Session

Board 8096F: The natural history of NGLY1 deficiency, a rare neurodevelopmental disorder, and preliminary safety of GS-100, an AAV9 gene therapy for its treatment ⭐

Authors:

H. Landy[1], B. Suter[2], J. Thompson[2], R. Deck[1], M. Wilsey[3]; [1]Grace Sci., LLC, Menlo Park, CA, [2]Baylor Coll. of Med., Houston, TX, [3]Grace Sci. Fndn., Menlo Park, CA

Abstract:

Background: N-glycanase 1 (NGLY1) is a key enzyme in the degradation of misfolded glycosylated cytosolic proteins. Variants in the *NGLY1* gene cause a rare neurodevelopmental disorder with a core phenotype comprising severe global developmental delay/intellectual disability, hyperkinetic movement disorder, transient transaminase elevation, (hypo)alacrima and progressive sensorimotor polyneuropathy. Shortened life span is also described. In the absence of functional NGLY1, there is accumulation of aspartylglucosamine (GlcNAc-Asn, or GNA). As only 130 patients have been identified worldwide, natural history data are sparse. GS-100 is a recombinant, single-stranded AAV9 vector that encodes a codon-optimized full-length version of human NGLY1 under the control of the CAG promoter. AAV9 was chosen based on its tropism for central and peripheral nervous system tissues. **Objectives:** We present data from a natural history study (NHS) and from a recently initiated Phase 1/2 interventional gene therapy study (GS-100-A301; NCT0619953) in NGLY1-deficient patients, conducted in parallel, to determine optimal clinical outcomes and biomarkers for Phase 3. **Methods:** The NHS is a single center, longitudinal observational study with assessments at Baseline and Months 6, and 12. Each visit includes neurological and physical examinations, ophthalmological assessment (including Schirmer test), plasma GNA levels and developmental assessments (BSID-4, Vineland-

3, and WPPSI); additional assessments include GMFCS, GMFM-88, EEG, and nerve conduction studies. Study GS-100-A301 is a multicenter, Phase 1/2, ascending dose, treatment study to assess the safety and preliminary efficacy of GS-100 administered by intracerebroventricular (ICV) infusion in the treatment of NGLY1-deficient patients. **Results:** Fifteen patients (9 females) aged 4-13 years from the USA, Canada, Mexico, England, and Brazil are enrolled in the NHS. Biallelic variants encountered comprise missense, nonsense, frameshift and splice site mutations. The most common variant is c.1201A-T, seen in 8 patients. Data for all patients through Month 12 will be presented. Additionally, safety data on 2 patients who have received ICV GS-100 at a dose of 4e14 vg/individual will be presented.Conclusions: Natural history and Phase 1/2 studies of rare genetic disorders provide information that assists with disease diagnosis, understanding disease course and phenotypic spectrum, and endpoint selection for future interventional studies. With these endeavors, we hope to gain a thorough understanding of NGLY1-related disease to further inform the design of a Phase 3 gene therapy trial.

Session Title: Genetic Therapies Poster Friday Session

Board 8097F: Treatment paradigms for studying functional & molecular characterization of the Phospholamban L39 Stop mutation (PLN L39X) & its relevance to cardiomyopathy using patient-derived pluripotent stem cell models (hiPSCs)

Authors:

A. Devendran[1], S. Ng[2], C. Liu[3], R. Dariolli[4], N. Dubois[5], K. D. Costa[1], M. G. Trivieri[1]; [1]Cardiovascular Res. Inst., Icahn Sch. of Med. at Mount Sinai, New York city, NY, [2]Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York city, NY, [3]Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York city, NY, [4]Bonds Bio systems Corp., Massachusetts, MA, [5]Mindich Child and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York city, NY

Abstract:

Background: Heart failure (HF) is a complex clinical condition associated with substantial morbidity and mortality worldwide. The contractile dysfunction and arrhythmogenesis related to HF has been linked to the remodeling of calcium (Ca^{++}) handling. Phospholamban (PLN) has emerged as a key regulator of intracellular Ca^{++}concentration. Of the PLN mutations, L39X is intriguing as it is not been properly characterized. This mutation is believed to be functionally equivalent to PLN null (KO) but contrary to PLN KO mice, L39X carriers develop a lethal

cardiomyopathy (CMP). This study aims to establish patient-derived human induced pluripotent stem cells derived cardiomyocytes (hiPSC-CMs) from homozygote L39X carriers (human subjects) as model system, to explain the functional and molecular characterization of the L39X in human pathophysiology related to cardiomyopathy. **Hypothesis:** The hiPSC-CMs harboring the mutations (L39X carriers) pertaining to the L39 experimental (n =2), L39-KO (n =2), and the WT-KO (n = 2), lines display a specific abnormal cellular distribution, along with irregular transcriptional, translation expression and functional activity compared to the healthy controls (Wild type, WT, n = 2). **Results:** Compared to the healthy controls (WT-hiPSC-CMs), the PLN, along with SERCA2a, and other calcium handling markers had an abnormal cytoplasmic distribution and formed intracellular aggregates, with the loss of perinuclear localization. Similarly, the functional assays pertaining to the calcium handling (action potential and calcium transient) displayed significant arrythmogenicity. Besides, the PLN-dependent modulation of human cardio myocyte differentiation from days 0, 10, 20, 30 and 40, via the mesoderm or cardiac specification markers like p-Akt, p-GSK3β, MESP1, and WNT3 showed significant differences among the L39X carriers and WT, both transcriptionally and translationally. On the other hand, the PLN-dependent mitochondrial synthase activity for the above time course showed significant differences among different group along with higher mitochondrial membrane potential and oxidative stress. **Conclusion:** This data confirms an abnormal intracellular distribution of PLN, with the loss of perinuclear accumulation and mis-localization, suggestive of ineffective targeting to or retention of L39X.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1042F: A lay-translation of the human phenotype ontology aids diagnostics

Authors:

M. Haendel[1], K. Shefchek[2], J. McMurry[1], V. Nicole[3], **P. Robinson**[4], E. Riggs[5], C. Brownstein[6], S. Ziniel[7], M. Might[8], I. Holm[9]; [1]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [2]Oregon State Univ., Corvalls, CA, [3]Critical-Path, Tucson, OR, [4]Charite, Berlin, Germany, [5]Geisinger, Huntsville, AL, [6]Boston Children`s Hosp., Boston, MA, [7]Univ. of Colorado Anschutz, Aurora, CO, [8]Univ. of Alabama, Birmingham, AL, [9]Boston Children s Hosp, Boston, MA

Abstract:

Deep phenotyping is essential for improving diagnostics and research on rare diseases. The Human Phenotype Ontology (HPO) is a global standard used to capture disease phenotypes.

HPO patient profiles are compared against those curated by the Monarch Initiative in a process that has been proven to increase diagnostic efficacy in variant prioritization pipelines and tools. Patients are an under-utilized source of information; therefore, to facilitate self-phenotyping, we "translated" HPO into plain language ("lay-HPO"), representing approximately 1/3 of the full clinical HPO. Standard care tends to leverage patient survey instruments, such as the GenomeConnect survey, which asks patient-friendly questions that map to approximately 200 HPO terms. In this study, we compared self-phenotyping using lay-HPO to clinician phenotyping using the full HPO and the GenomeConnect survey, aiming to empower patients to participate in their diagnostics and complement clinician-based phenotyping, which often misses patient-reported symptoms such as "snoring" or "lack of tears" in infants.
We created theoretical maximum-accuracy phenotype profiles for each disease using lay-HPO terms and GenomeConnect-mapped terms. Both sets of synthetic profiles performed well in ranking diseases in semantic similarity analyses to the Monarch gold standard disease profiles (area under the curve 0.991 and 0.954, respectively). We randomized rare disease participants with diagnosed genetic diseases to complete the GenomeConnect survey, lay-HPO, or both. For each diagnosed disease, we compared the synthetic disease profiles to patient-completed profiles for the lay-HPO versus GenomeConnect. Profiles from GenomeConnect survey participant responses were more accurate than lay-HPO, but lay-HPO had a tighter distribution of scores for respondents who did both, indicating greater precision. We evaluated the ability of each known Mendelian disease HPO phenotype profile to retrieve the corresponding disease. Participants generally preferred the GenomeConnect multiple-choice format over using the lay-HPO directly. Our results demonstrate that individuals can provide rich HPO phenotype data, suggesting self-phenotyping could improve diagnostics by supplementing clinician-created profiles.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1043F: A reference panel for linkage disequilibrium and genotype imputation using whole-genome sequencing data from 2680 LASI-DAD participants across India

Authors:

W. Zhao[1,2], Z. Li[3], X. Zhou[3], P. Moorjani[4,5], S. Dey[6], J. Lee[7], J. A. Smith[2,1], A. B. Dey[8], S. L. R. Kardia[2]; [1]Survey Res. Ctr., Inst. for Social Res., Univ. of Michigan, Ann Arbor, MI, [2]Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [3]Dept. of Biostatistics, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [4]Dept. of Molecular and Cell Biology, Univ. of

California, Berkeley, Berkeley, CA, [5]Dept. of Computational Biology, Univ. of California, Berkely, Berkeley, CA, [6]Dept. of Biophysics, All India Inst. of Med. Sci., New Delhi, India, [7]Dept. of Economics, Univ. of Southern California, Los Angeles, CA, [8]Venu Geriatric Inst., New Delhi, India

Abstract:

With a population of over 1.4 billion, India is the most populous country in the world. In contrast to its vast population, genetic studies involving Indian individuals remain limited. The Indian population is composed of many founder groups and has a mixed genetic ancestry, including an ancestral component not observed anywhere outside of India. This presents a unique opportunity to uncover novel disease variants and develop more tailored medical interventions or treatments. To facilitate genetic research in India, a crucial first step is to create a foundational resource characterizing the unique genetic population structure of this population. To achieve this goal, we have constructed the largest and most nationally representative linkage disequilibrium (LD) and genotype imputation reference panel in India to date, using whole genome sequence (WGS) data from 2,680 Indian participants from the Longitudinal Aging Study in India-Diagnostic Assessment of Dementia (LASI-DAD). As an LD reference panel, LASI-DAD covers significantly more variants than the South Asian panel in TOP-LD (69 million versus 22 million). Among all pairs of genetic variants evaluated within 1 Mb of each other, the proportion of pairs with $r^2 \geq 0.2$ is 0.047%, where r^2 represents the squared Pearson correlation coefficient between the phased haplotypes. On average, each variant is in LD with 26 variants in LASI-DAD based on an r^2 threshold of 0.2. We further characterized the LD patterns in LASI-DAD by identifying LD blocks and evaluating regional LD differences between LASI-DAD and five super-populations from the 1000 Genomes Project (1KGP). Despite the overall smallest difference in LD patterns between LASI-DAD and the South Asian population in 1KGP, we still observed a significant number of regions with substantial differences in LD. In addition, as a genotype imputation reference panel, LASI-DAD improves the imputation accuracy Rsq (correlation between imputed genotype and true genotype) by an average of 15% compared to the Genome Asia Pilot (GAsP) panel and 34% compared to Trans-Omics for Precision Medicine WGS (TOPMed) panel. More importantly, the reduced variation in imputation accuracy for samples from different Indian states, compared to the other two panels, highlights the significance of LASI-DAD as a nationally representative panel for the Indian population. Currently, we are working to make the LASI-DAD reference panel available as a public resource to benefit future studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1045F: All models are wrong, but which ones are useful? Characterizing broad and recurrent transcriptional profiles in neuronal cell models

Authors:

S. Reed[1], D. Tai[1,2,3,4], C. De Esch[1,2,3,4], K. O'Keefe[1], N. Burt[1], R. Bhavsar[1], C. Vaine[2,5], X. Nuttle[1,2,3,4], B. Currall[1,2,3], C. Bragg[2,5], S. Erdin[1,3], J. Gusella[1,2,6,7,3], R. Yadav[1,2,3,4], M. Talkowski[1,2,3,4], D. Gao[1,2,3,4]; [1]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [2]Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [5]The Collaborative Ctr. for X-linked Dystonia-Parkinsonism, Massachusetts Gen. Hosp., Charlestown, MA, [6]Dept. of Genetics, Blavatnik Inst., Harvard Med. Sch., Boston, MA, [7]Harvard Stem Cell Inst., Harvard Univ., Cambridge, MA

Abstract:

One of the most pervasive challenges in neurological research is the need to use cellular models to proxy brain function and their limitations for drawing neurobiological conclusions. New advances have led to the development of organoids, models designed to emulate macro-level features of cellular organization not present in more conventional “2D” cell models. We sought here to quantify the transcriptional and compositional differences between 2D models, organoids, and both fetal and adult human brains across laboratories, protocols, and analytic methods. We have performed analysis on 60 single-cell RNA (scRNA) libraries from iPSC derived neuronal lineages, including neural stem cells (NSCs), neural progenitor cells (NPCs), NGN2-induced neurons (iNs) and cerebral brain organoids (COs).After correcting for technical confounding factors (batch, lab, protocol, etc.) we found that COs showed the largest overall variance in gene expression among all lineages, owing to their more varied cell composition. Using a multiple correspondence analysis (MCA) method, we identified that COs had 25% of cells classified as excitatory neurons, 21% astrocytes, 19% inhibitory neurons, and 7% oligodendrocytes. On the other hand, most NSCs (70%) resemble dividing cells and iNs that do not transcriptionally mimic any specific neurons in the central nervous system. We also used the curated set of cell type marker genes provided by the Human Commons Cell Atlas (HCCA) and applied their robust cell type annotation method to verify that these results are not specific to one method or marker gene annotation. We then compared cell type abundances from our cell models against public scRNASeq datasets of human brains. In ongoing analyses, we sought to capitalize on a unique dataset from BrainSpan that contains bulk RNASeq for 600 samples across 25 brain regions ranging from 8 weeks post-fetal to 40 years. In these analyses, we applied the

HCCA cell type annotation to map the age of individual cells using age-specific marker gene sets derived from BrainSpan to characterize the degree to which neural cell models approximate transcriptional and compositional features of developing brains. Our analyses will ultimately help guide cellular-model studies and therapeutic targeting in neurological research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1046F: Assessing the value of the human pangenome reference for trait mapping

Authors:

S. Lu[1], W-w. Liao[1], M. D. Gorter[2], P. C. Goddard[2], HPRC Pangenome working group, S. B. Montgomery[2], I. M. Hall[1]; [1]Dept. of Genetics, Yale Univ. Sch. of Med., New Haven, CT, [2]Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA

Abstract:

The human pangenome project is generating hundreds of high quality reference genomes from diverse populations, and describing the variation within them in graph data structures that are amenable to bioinformatic analyses such as read alignment and variant calling. A key goal of the project is to make routine analysis of human genome sequencing data more sensitive, more accurate, and less biased. In our study, we examine the extent to which using the human pangenome reference for variant detection will improve human trait association studies, using molecular traits such as gene expression and chromatin accessibility as a model for complex trait genetics.

We conducted eQTL analysis using RNA-seq data of 430 individuals from 5 African subpopulations. These samples underwent deep whole genome sequencing as part of the 1000 Genomes Project (1KG), with RNA-seq and ATAC-seq data generated by the African Functional Genomics Resource. We performed eQTL analyses using variant calls derived from several different methods including: (1) a standard small variant callset generated using GATK and the GRCh38 reference genome; (2) a structural variant (SV) callset generated using multiple traditional SV detection methods and GRCh38; (3) a combined SNP/indel/SV callset generated using Pangenie, a graph-based variant genotyping method run by the Human Pangenome Reference Consortium (HPRC) as part of the Freeze 1 publication (including 94 haplotypes); and (4) a novel method that we developed to genotype all types of variation based on alignment of short read WGS data to pangenome reference graphs, followed by analysis of normalized read-depth at edges and nodes.

We performed a joint eQTL analysis including all variants genotyped by our novel method as well as the other three methods, and identified 10453 eQTLs. At 51.1% of these eQTLs, the lead marker was derived from our pangenome graph-based method, and preliminary analyses indicate that ~17% of these were not found at genome-wide significance by any of the other three methods, meaning the association would otherwise not have been discovered. In contrast, 29.7% of lead markers were from GATK, 0.2% from the 1KG SV callset and 18.9% from Pangenie. These results indicate that implementation of pangenome-based variant analysis methods will increase the power of molecular trait association studies, and by extension complex trait genetics projects as well. We will further discuss various additional details of the variant calling and eQTL mapping experiments, and also present data on the added value of the pangenome resource for ATAC-seq peak calling and chromatin accessibility-QTL mapping using this same dataset.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1047F: Automated large-scale quantification of sex bias in biomedical studies

Authors:

L. Valtadoros[1], K. Johnson[1], H. Yuan[2], M. Ahmadian[1], A. Krishnan[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Michigan State Univ., East Lansing, MI

Abstract:

Historical exclusion of female study subjects in human, and especially, animal and cell culture research has led to considerable sex bias in biomedical studies. Quantifying this bias is critical for identifying areas for future compensatory and balanced investigations, but doing so on a large scale remains a major challenge. Manual reviews of publications have helped estimate the sex bias in a small subset of research areas, but there is an immediate need for automated methods that can scale to a variety of areas. Here, we present a computational approach to obtain an estimate of sex bias in hundreds of biomedical areas by combining data of ~230,000 human transcriptomics samples with their associated ~3,500 publications. As most samples in public databases such as GEO and SRA lack information about sex in their metadata, we began with inferring sex labels for all unlabeled samples based on a robust sex-classification gene signature derived from ~30,000 expert curated primary human samples. Using the predicted sample sex, we then calculated a “sex ratio” (fraction of samples labeled as female) for each GEO and SRA study. Finally, using natural language processing techniques, we estimated the

linear and non-linear correlation of the normalized occurrence frequency of each of the ~14,500 biological terms in the titles and abstracts of studies with the sex ratio values of those studies. This analysis identified terms that are biased in the data but are not biologically sex-specific. Examples include diseases (such as “myocardial infarction” and “hepatocellular carcinoma”), genes (“epidermal growth factor receptor” and “KIF20A”), and species (“murine” and “P. gingivalis”). These and several other terms identified in our study are critical biomedical areas where there is an urgent need for both new studies focused on females to correct our sex-biased knowledge and others with a balanced representation of sex to understand sex differences.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1048F: AVITI Sequencing for high-quality, low-cost genomic analysis at the NWGC

Authors:

E. Ryke, C. Frazar, S. Mcgee, J. Smith, C-L. Wei; Univ. of Washington, Seattle, WA

Abstract:

The rapid advance of next generation sequencing has been a keystone in modern biology, delivering unprecedented insights over the last two decades, yet there remain challenges with accuracy, cost effectiveness, and options. Recently, there has been a significant breakthrough in sequencing technologies that promise cost-effective, highly accurate, and high-output short-read data. Among them, Avidite sequencing from Element Biosciences leverages rolling circle amplification and avidity-base chemistry to perform separate base detection and extension steps, which combined with conventional sequencing-by-synthesis chemistry, offer sequence data with higher accuracy and lower cost compared to the current gold standard, Illumina SBS and X-LEAP chemistry. The launch of Cloudbreak FreeStyle flow cell technology allows for the direct loading of third-party libraries without the need for costly, time-consuming library conversion. Here, we evaluate Avidite technology through multi-omic sequencing applications including whole genome, RNA-Seq, epigenomic, 3D chromatin, and single-cell sequencing. We compare the data quality to data generated by the Illumina NovaSeq platform. We find that mean quality scores from AVITI sequencing typically exceed Q40, produce lower error rates, and maintain consistency in base quality throughout long sequencing cycles. We describe the performance metrics (mapping, SNP and indel calling, passing filter, and error rate, etc) for the different data types, the performance by genomic context (GC, homopolymers, etc), complexity

and consistency, and quantitative expression analysis. Finally, we define how we implement the AVITI workflows at the Northwest Genomics Center (NWGC) and the limitations and caveats of sequencing with the Element Biosciences AVITI system.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1049F: Building phenotype trajectories for rare disease: A pilot focusing on Noonan Syndrome

Authors:

T. Groza[1], D. Gration[2], C. Derrick[1], W. Lim[3], J. Goh[4], J. Lim[5], E. Tan[6], A. Lai[7], B. Goh[8], E-S. Tan[5], S. S. Jamuar[8], G. Baynam[1]; [1]Rare Care Ctr., Perth Children's Hosp., Perth, Australia, [2]Western Australian Register of Dev.al Anomalies, King Edward Mem. Hosp., Perth, Australia, [3]Duke-NUS Precision Med. Inst., Singapore, Singapore, [4]Nursing Div., KK Women's and Children's Hosp., Singapore, Singapore, [5]Genetics Service, KK Women's and Children's Hosp., Singapore, Singapore, [6]Res. Ctr., KK Women's and Children's Hosp., Singapore, Singapore, [7]Singapore Hlth.Services, Singapore, Singapore, [8]KK Women's and Children's Hosp., Singapore, Singapore

Abstract:

The absence of established treatment protocols for over 90% of rare diseases necessitates appropriate management of their phenotypic presentations. Therefore, it is crucial to develop care pathways for both diagnosed and undiagnosed rare disease patients, addressing prognosis, surveillance and coordinated clinical care. This approach is essential for providing emotional relief and ensuring appropriate access to resources.
Currently, the curated body of knowledge and computational methods in this field cannot document the temporal progression of phenotypes and their associated interventions, whether or not a known disease context (diagnosis) exists. Existing knowledge bases, such as OMIM, the Monarch Initiative, and Orphanet, focus on encoding disease-phenotype-genotype associations, which are vital for diagnostics but lack the granularity needed to map onset and progression. Hence, we have initiated an international effort to model, capture, and externalize the relationships governing phenotypes across the lifespan. Our goal is to produce and openly publish phenotype trajectories that provide a more accurate view of prognosis and functional consequences.
To start, we focused on a single rare disease - Noonan syndrome - and studied the types of relationships that emerge from EHR data, authoritative disease reviews, and the comparison between manual curation and large language model (LLM)-based automatic extraction. The EHR

data encompassed a cohort of 7 patients from the genetic services at KK Hospital in Singapore, tracked longitudinally over 2 to 8 years, with over 400 encounters across 10+ specialties. By applying our best-in-class phenotype concept recognition algorithm on the clinical notes, we built longitudinal patient phenotype profiles aligned with the Human Phenotype Ontology.
In parallel, we manually curated the GeneReviews description of Noonan syndrome, resulting in over 100 phenotypes across the entire lifespan, 20+ interventions associated with system-specific phenotypes, and 10+ surveillance methods. Additionally, we used advanced prompt engineering techniques with GPT-4 to explore the feasibility of automating the curation process. Coding phenotype trajectories and understanding the specific time points that trigger management events enable clinical teams to proactively intervene and deliver personalized care plans for both diagnosed and undiagnosed patients. We discuss the challenges encountered in aligning domain knowledge with data, the utility of using LLMs to curate phenotype trajectories, and the applicability of the resulting phenotype trajectories to existing and new patients.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1050F: Challenges and lessons learned in developing ClinGen initial modified ACMG/AMP criteria for variant curation in X-linked inherited retinal disease genes RPGR, RS1, CHM and RP2

Authors:

K. Worley[1], M. Mero[1], W. C. Hankey, IV[2], K. Lee[2], L. S. Sullivan[3], R. Ayyagari[4], M. Benson[5], V. Chong[6], K. Goetz[7], L. Haer-Wigman[8], R. Hufnagel[9], S. Hull[10], J. H. Hussain[1], K. Kämpjärvi[11], J. Sallum[12], P. A. Sieving[13], M. Wang[1], G. Wang[14], C. Zeitz[15], R. A. Lewis[1,16], R. Chen[1], on behalf of the ClinGen X-linked Inherited Retinal Disease Variant Curation Expert Panel; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univsity of Texas Hlth.Sci. Ctr. Houston, Houston, TX, [4]Univ California San Diego, San Diego, CA, [5]Univ. of Alberta, Edmonton, AB, [6]Univ. of Utah, Salt Lake City, UT, [7]Natl. Eye Inst., NIH, Bethesda, MD, [8]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [9]Ctr. for Integrated Hlth.care Res., Kaiser Permanente, Honolulu, HI, [10]Dept. of Ophthalmology, Univ. of Aukland, Aukland, New Zealand, [11]Blueprint Genetics OY, Espoo, Finland, [12]Inst. de Genética Ocular, Federal Univ. of Sao Paulo, Sao Paulo, Brazil, [13]Ophthalmology & Vision Sci., UC Davis, Davis, CA, [14]Invitae Corp., San Francisco, CA, [15]Sorbonne Université, INSERM, CNRS, Inst. de la Vision, Paris, France, [16]Dept. of Ophthalmolology, Baylor Coll. of Med., Houston, TX

Abstract:

Purpose: The X-linked Inherited Retinal Disease Variant Curation Expert Panel (IRD, VCEP) addresses consistency in variant interpretation for genes which account for ~15% of IRD, to improve specificity and accuracy of molecular diagnoses of patients and to guide for eligibility for IRD genetic therapies. **Methods:** We adapt and implement the ClinGen variant curation practices to assess individual variants in seven X-linked genes (*CACNA1F*, *CHM*, *NDP, OFD1*, *RPGR*, *RP2* and *RS1*). **Results:** We developed and refined with pilot curation exercises gene-specific rules for *RPGR* (cause of 9% of rod-cone and cone-rod dystrophies); developed initial specifications, planned, and are executing pilot curation exercises for *RS1* and *CHM;* and are defining *RP2* initial specifications. **Conclusions:** We identified issues specific to IRD phenotypes and X-linked inheritance that were not adjudicated previously by ClinGen. Female carriers of some X-linked IRD gene variants have no phenotypes, and carrier phenotype reporting is variable. Thus, we limit the PP4 proband designation in *RPGR* and *RP2* to males and allow female carriers with an affected and phenotyped male relative and a pedigree showing their relationship to score PS4. Population frequency codes must use hemizygotes in gnomAD, as “normal” females may carry disease-associated alleles without a phenotype. This reduces the number of alleles available for benign variants. Over 100 genes are implicated in IRDs, so we increased specificity by requiring diagnostic decreased *electroretinogram* or fundus autofluorescence measures. Such measurements are clinically relevant only while the patient’s retina still functions, and young children do not tolerate the test, reducing the number of PS4-eligible probands to males between ~8 and 25 years. Gene features also require adjustments. *RPGR* C-terminal variants are associated with later age of onset and cone predominant disease phenotype - we defined regions to assess these phenotypes and adjustments for age of onset and scores for PP4 and PS4. *RS1* has specific amino-acid binding interactions required for structural stability of monomer and functional paired-octamer structures, allowing use of PS3 and some common deletion variants designated for PM1. With highly specific IRD phenotype associated with *CHM* variants, we enhanced the PP4 strength and the value of *de novo* cases under PS2/PM6. Variant curation with these FDA-recognized guidelines will improve the specificity and accuracy of molecular diagnoses, increasing the value of genetic testing as a diagnostic tool and as a guide for patient eligibility for IRD genetic therapies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1051F: Clinical application of whole exome sequencing in the diagnosis of Pediatric neurological disorders: A single center study from Nepal

Authors:

R. Mishra[1], A. Poudyal[2]; [1]Civil Service Hosp., Kathmandu, Nepal, [2]Univ. of New Haven, New Haven, CT

Abstract:

Introduction: Pediatric neurological disorders encompass various conditions impacting children's brain, spinal cord, and nerves, affecting their development and well-being. Genetic variants significantly contribute to these disorders' onset and severity. The study analyzes pediatric neurological disorders using DNA tests, focusing on variant distribution, pathogenicity, inheritance patterns, and the impact on disease likelihood.
Methodology: This study utilized a retrospective approach, analyzing Whole Exome Sequencing (WES) test reports from pediatric patients diagnosed with neurological disorders. Reports were reviewed at Civil Services Hospital Kathmandu, Nepal. Genetic variants were annotated for pathogenicity and inheritance. Statistical analyses, including Fisher's Exact Test and logistic regression modeling, were conducted to assess associations and estimate disease likelihood while controlling for variables.
Results: Analysis of DNA test reports from 7 pediatric neurological disorder patients revealed diverse genetic variants associated with conditions like cerebellar atrophy, seizure, and developmental delay. The variant of Uncertain significance (VUS) accounted for 50% of the cases, indicating diagnostic complexities. Fisher's Exact Test demonstrated a significant association ($p < 0.05$) between inheritance patterns and pathogenicity classifications, with autosomal variants more uncertain than X-linked. Logistic regression identified variants significantly impacting disease likelihood, guiding personalized treatment. Positive coefficients indicated increased disease likelihood, while negative coefficients suggested protective effects. Variant c.19C>T in the WDR45 gene showed a significant association ($p < 0.001$) with disease likelihood, highlighting its clinical relevance in pediatric neurological disorders.
Conclusion: These findings highlight the complexity of genetic diagnostics in pediatric neurological disorders and the importance of genetic testing in guiding treatment. Further research is needed for validation and to advance precision pediatric neurology.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1052F: Comparison of variant callers using60,532 diverse genome sequences

Authors:

H. Zhou; Harvard Univ., Boston, MA

Abstract:

Understanding the human genome entails analyzing functional variants and their influence on physiological characteristics. However, these variants are often infrequent and pose challenges for thorough investigation. To address this challenge, large-scale whole-genome sequencing (WGS) has become the method of choice to identify functional variants and their physiological effects. A comprehensive identification of variants necessitates sequencing numerous samples with dependable variant calling methods. Variant calling from WGS is a complex process that involves mapping, genotyping, annotating, and rigorous quality control (QC). The dependability of variant calling tools is of paramount importance. VT and GATK, commonly utilized in large-scale sequencing projects, offer distinct quality control metrics. Rigorous quality control procedures were applied to align the outputs of variant callers, ensuring consistency in the results after quality checks. The National Human Genome Research Institute's (NHGRI) Genome Sequencing Project (GSP) has produced an extensive amount of WGS DNA with aims to identify genes and variants related to hereditary diseases of individuals with diverse backgrounds. Within the GSP, the Centers for Common Disease Genomics (CCDG) Freeze 2 WGS data revealed more than 450 million variants across 60,532 samples. The QC mitigated batch effects, enhancing uniformity across various sequencing centers. It provides valuable insights into the role of rare variants in phenotypic variation. This QC protocol uses Hail for analyzing WGS. A systematic comparison between VT and GATK was conducted within the same extensive WGS project to assess their agreement. Both VT and GATK exhibited high concordance, with the choice depending on the user's needs and resources.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1053F: Comprehensive identification and characterization of genetic variants within a pangenome reference

Authors:

P. Salehi Nowbandegani; Harvard Med. Sch., Dept. of BioMed. Informatics, Boston, MA

Abstract:

Graph genomes provide a vital perspective on genetic diversity, capturing variation that is missed when aligning to a linear reference. A graphical, pangenome reference captures differences between every pair of individuals, including those whose ancestry differs from that of any single reference genome. However, it is surprisingly challenging to define genetic variants in a pangenome. Variants are often represented as "bubbles," with internally disjoint paths between two vertices, but the number of bubbles can grow exponentially. An additional challenge is to define the position of a variant with respect to a linear reference.

We introduce a new representation of genetic variation and a method to identify all variants in a pangenome. Our method utilizes the fact that the cycle space of a graph has a basis consisting of the edges that do not belong to a spanning tree. Our method relies on a Depth-First Search (DFS) tree, incorporating the linear reference hg38 as a lineage, which we call the reference tree. We represent genetic variants as edges outside the DFS tree. The variants of a sample uniquely determine that sample's entire genome, confirming the appropriateness of our approach. Our method is computationally efficient, and it automatically classifies variants as either inversions, duplications, deletions, or replacements (including SNPs).

Because our reference tree incorporates a linear reference (e.g., hg38) as its first branch, we are able to define the position of variants not on the linear reference, as a certain interval on the linear reference. This strategy ensures compatibility with the linear reference framework. We have proven that the set of variants in a sample within an interval is sufficient to reconstruct the genome within this interval, allowing users to work within loci. An advantage of our approach compared with one based on contig naming is that it maintains the approximately linear nature of the genome where appropriate.

Our algorithm has been rigorously tested on two human graph genomes, including the c4a gene graph and the minigraph-cactus pangenome of chromosome 20 from Hickey et al. 2024, effectively identifying 10 and 555,314 variants respectively. It processes the hg38-chr20 minigraph-cactus in about 20 minutes. It also accurately classified 463,516 variants as SNPs, closely matching the 459,908 SNPs identified using a standard VCF approach. We plan to release open-source software as well as a comprehensive list of genetic variants found in the human pangenome reference.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1054F: Comprehensive network modeling approaches unravel dynamic enhancer-promoter interactions across neural differentiation

Authors:

A. Kreimer[1], W. DeGroat[1], F. Inoue[2], T. Ashuach[3], N. Yosef[3], N. Ahituv[2]; [1]Rutgers Univ., Piscataway, NJ, [2]UCSF, San Francisco, CA, [3]Berkeley Univ., Berkeley, CA

Abstract:

Background: Increasing evidence suggests that a substantial proportion of disease-associated mutations occur in enhancers, regions of non-coding DNA essential to gene regulation. Understanding the structures and mechanisms of regulatory programs this variation affects can shed light on the apparatuses of human diseases. Results: We collected epigenetic and gene expression datasets from seven early time points during neural differentiation. Focusing on this model system, we constructed networks of enhancer-promoter interactions, each at an individual stage of neural induction. These networks served as the base for a rich series of analyses, through which we demonstrated their temporal dynamics and enrichment for various disease-associated variants. We applied the Girvan-Newman clustering algorithm to these networks to reveal biologically relevant substructures of regulation. Additionally, we demonstrated methods to validate predicted enhancer-promoter interactions using transcription factor overexpression and massively parallel reporter assays. Conclusions: Our findings suggest a generalizable framework for exploring gene regulatory programs and their dynamics across developmental processes. This includes a comprehensive approach to studying the effects of disease-associated variation on transcriptional networks. The techniques applied to our networks have been published alongside our findings as a computational tool, E-P-INAnalyzer. Our procedure can be utilized across different cellular contexts and disorders.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1055F: Connecting with the Impact of Genomic Variation on Function (IGVF) Consortium

Authors:

A. Asare, NHGRI IGVF Team; Natl. Human Genome Res. Inst., Bethesda, MD

Abstract:

The Impact of Genomic Variation on Function Consortium (IGVF) was launched in September 2021 and seeks to achieve two objectives: first, enhancing the scientific community's understanding of the effects of genomic variation on genome function and phenotypes, and second, creating a resource to gather data, models, and predictions on how variation in the genome's protein coding and non-coding regions affect cellular function and molecular phenotype. The IGVF Consortium is made up of five components: Functional Characterization Centers, Mapping Centers, Predictive Modeling Projects, Regulatory Network Projects, and a Data and Administrative Coordination Center. Across these five components, IGVF researchers are utilizing a variety of high-throughput experimental and computational methods to examine connections between regulatory elements, variants, and phenotypes. To date, over 100 preprints have been submitted to the IGVF bioRxiv channel, showing how these methods are being put into practice to help advance the field. The resource developed by IGVF will consist of a data portal and an interactive data catalog. It will be populated with experimental data, integrative analyses, computational predictions, as well as standards and pipelines produced by consortium members. To create the catalog of genomic variation, IGVF researchers are employing a map-perturb-predict framework - which consists of three approaches: single-cell mapping, systematic genomic perturbations, and predictive modeling - with the goal of connecting genome variation to genome function and phenotypes. To date, the data portal contains more than 300 experimental datasets, and an initial set of links between genes, variants, and phenotypes can be found via the Catalog API. The completed resource is expected to aid both the scientific and clinical communities in answering biological questions regarding the implications of protein coding and non-coding genomic variation for human health and disease. This poster provides an overview of the IGVF consortium and the resource progress to date, highlighting the types of data included and demonstrating ways the scientific community can utilize the IGVF resource.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1056F: Creating a virtual education & thought leadership webinar program to engage the UK Biobank Research Analysis Platform community

Authors:

B. Busby[1], B. Pyle[1], A. Sedlakova[1], A. Lee[1], A. Fungtammasan[2], T. Laderas[3]; [1]DNAnexus, Mountain View, CA, [2]Independent Res.er, Calgary, AB, Canada, [3]Fred Hutch, Seattle, WA

Abstract:

In 2021 DNAnexus & UK Biobank launched the UK Biobank Research Analysis Platform (UKB-RAP), a cloud platform allowing UK Biobank researchers to access & analyze the world's largest biomedical dataset. UK Biobank researchers had to understand how to migrate their existing research processes onto the platform which required our team to engage & educate the new community. We also had to accommodate a global community, split across varied regions & timezones. To solve this we created a wide ranging webinar education & engagement plan that taught the UKB-RAP community practical ways to work on the platform. In addition, we hosted roundtables on specific, emerging topics in an effort to encourage experts to further push their data science capabilities.

Our goal with creating our session plan was to build a research community from the ground up & foster that community through engagement. Over the course of two years we ran 38 webinars, covering a wide range of topics ranging from introductory courses to advanced topics (image analysis, variant annotation). We also identified wider bioinformatics topics & partnered with HDRUK to build tutorials for cross biobank analysis with docker & other container technologies, as well as a session on ontological considerations when crossing between these datasets.

These were built with the intention of our community in mind and we also worked to foster engagement by keeping them apprised of the latest data releases & scientific research. We partnered with UK Biobank & their research consortium to present on the new data releases & strategies for analyzing them. We also partnered with experts in the scientific community to hold roundtable discussions on multimodal research, oncology, & more.

User communities cannot be sustained solely through events, so DNAnexus created a community forum (which has since evolved to the UK Biobank Community Forum, available for all UK Biobank researchers) for users to ask questions & engage with fellow researchers. In the context of our webinars, the forum served as a place to find announcements on upcoming sessions & as a repository for webinar related resources (recordings, slides, Q&A documents). Attendees could also continue to ask questions from the sessions to the speakers & panelists from the event. We'll speak to the input we gathered from our audience to ensure our sessions continue to be valuable.

The purpose of this session is to inform the genomics education community about lessons learned in educating thousands of users, such that we can more effectively train people to do better science as more data enclaves like the UKB-RAP come online.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1057F: Cross-population replication analysis of STR-trait associations in the All of Us dataset

Authors:

N. Ma[1], J. Margoliash[2], T. Mirmira[2], S. Javazdazeh[2], H. Z. Jam[2], A. Massarat[3], W. G. Rivera[3,4], Y. Li[5], K. A. Frazer[6], M. D'Antonio[4], M. Gymrek[5,2]; [1]Dept. of Med., Univ. of California San Diego,, La Jolla, CA, [2]Dept. of Computer Sci. and Engineering, Univ. of California San Diego, La Jolla, CA, [3]Bioinformatics and Systems Biology Graduate Program, Univ. of California San Diego, La Jolla, CA, [4]Dept. of BioMed. Informatics, Univ. of California San Diego, La Jolla, CA, [5]Dept. of Med., Univ. of California San Diego, La Jolla, CA, [6]Dept. of Pediatrics, Univ. of California San Diego, La Jolla, CA

Abstract:

Recent studies have demonstrated a widespread role of short tandem repeats (STRs) in complex traits, but these have been mostly limited to individuals of European descent. The All of Us (AoU) cohort has the potential to enable detailed study of the influence of genetics and other factors on complex traits across diverse populations. To evaluate the feasibility of performing similar studies in AoU, we performed a replication analysis of 119 STR-trait association signals identified in the UKBiobank (UKB) for blood traits. We first developed reproducible WDL pipelines for preprocessing phenotypes, STR genotyping, and association testing, which can be deployed on multiple cloud platforms. We applied our pipelines to genotype 59 unique STRs and performed association testing with 19 phenotypes with at least 24,000 samples available in AoU to test a total of 71 STR-trait pairs overlapping those tested in UKB. Overall, 93% of STR-trait pairs tested showed the same direction of effect between European AoU and UKB individuals, with strong correlation between association -log10 p-values in the two cohorts (Pearson r=0.87, p=1.21e-22). As expected, this correlation decreased but remained strong when considering individuals of African ancestry (r=0.61, p=4.95e-08). Our results include some of the strongest blood trait associations identified in UKB cohort. For example, we replicated an association between repeat length of a CGG repeat in the promoter of CBL with platelet count (p=5.12e-19; n=92099). Finally, we evaluated the ability to impute STRs into available SNP data on AoU. We used our published SNP-STR reference panel, which contains 1.09 million genome-wide repeats, to impute STRs on Chromosome 11. At the CBL repeat, we achieved 79.8% concordance between imputed repeat lengths vs. those obtained by direct genotyping with HipSTR, comparable to the 71% concordance achieved at this locus in UKB. In future work, we are expanding our analysis to include genome-wide tandem repeats and perform association testing against all available traits in AoU with sufficient sample size.
Overall, our results demonstrate the feasibility of performing STR association testing in AoU and show the majority of associations identified in UKB replicate in the diverse AoU population.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1058F: Custom genome assembly of the reference IPSC line KOLF2.1J improves annotation and mapping for genomic data

Authors:

P. Alvarez Jerez[1], A. Rhie[1], J. Kim[1], B. Walenz[1], D. Antipov[1], E. Lara[1], A. Beilina[1], D. Patel[1], F. Grenn[1], K. Billingsley[1], R. Genner[2], E. Gustavsson[3], M. Ryten[4], B. Paten[5], M. Ward[1], M. R. Cookson[1], M. Jain[6], A. Phillippy[7], C. Blauwendraat[1]; [1]NIH, Bethesda, MD, [2]Johns Hopkins Univ. / NIH, Bethesda, MD, [3]Univ. Coll. London, London, United Kingdom, [4]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom, [5]UCSC, Santa Cruz, CA, [6]Northeastern Univ., Boston, MA, [7]NIH/NHGRI, Bethesda, MD

Abstract:

Induced pluripotent stem cells (IPSCs) are widely used to study how genetic variants associated with neurodegenerative diseases (NDDs) affect cellular function. However, the heterogeneity of IPSCss may limit our ability to compare mutation effects across cell lines. For example, cell differentiation protocols may behave differently across different donor backgrounds. The iPSC Neurodegenerative Disease Initiative (iNDI) has promoted the use of the KOLF2.1J cell line as a reference line for studying neurological diseases due to favorable properties for growth and editing as well as having a neutral genetic risk for Alzheimer's disease (AD) and related disorders. This line, and its derivatives expressing >100 variants pathogenic for AD and other dementias, are publicly available. Here, we generated data on the differentiated cell types including mixed forebrain neurons, excitatory neurons, microglia, astrocytes and oligodendrocytes from the KOLF2.1J background. For each cell type we generated high depth long-read RNA sequencing and CAGE sequencing data in order to generate a reference transcriptome, which will be made publicly available. Additionally, through long-read DNA sequencing, we explored methylation differences across KOLF2.1J cell types. Finally, we present a telomere-to-telomere draft of a custom genome assembly for KOLF2.1J and demonstrate that this higher resolution and accurate genome allows for robust mapping of other genomic modalities for this reference line.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1059F: Deciphering Genetic Associations in Autoimmune Thyroid Diseases for Drug Repurposing

Authors:

A. Porika[1], K. Yadav[1], S. Khader[2]; [1]Engineering Med. Program at Texas A&M Sch. of Med., Houston, TX, [2]Imperial Coll. London Sch. of Publ. Hlth., London, United Kingdom

Abstract:

This study compares the genetic mechanisms of the two major autoimmune thyroid diseases, Hashimoto's thyroiditis and Graves' disease. Utilizing gene lists procured via DisGeNET: 185 genes for Hashimoto's, 334 for Graves', and 109 common to both, enrichment analysis was performed using Enrichr. The Descartes Cell Types and Tissue 2021, and the Bio Planet 2019 databases were used to evaluate cell types and pathways, respectively. OpenTargets identified gene drug targets from the EMBL-EBI ChEMBL database for both diseases. Among the findings, four cell types emerged as common across all gene lists, while Hashimoto's uniquely exhibits two distinct cell types. Notably, no singular pathways are exclusive to either Hashimoto's or Graves', although 12 pathways are uniquely associated with genes common to both conditions. Despite the contrasting clinical manifestations of hypothyroidism in Hashimoto's and hyperthyroidism in Graves', the enrichment analysis revealed a shared underlying etiology/pathophysiology that links both diseases. Moreover, it suggests a lack of unique cell types or pathways specifically attributable to Graves' disease. This insight may help elucidate why Graves' patients are more prone to later developing Hashimoto's while the opposite is rare. Additionally, OpenTargets identified one potential drug target for Hashimoto's and 32 for Graves' disease. Of the 32 gene drug targets identified for Graves' disease, nine of them can also be found in the DisGENET gene list for Hashimoto's thyroiditis, indicating potential opportunities for drug repurposing.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1060F: Divergent genetic influences on body mass index and body fat percentage: Implications for musculoskeletal and cardiometabolic health

Authors:

M. Romero Lado[1], I. Khrystoforova[2], G. Carrasquilla[3], T. Kilpeläinen[3]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res. - Univ. of Copenhagen, Copenhagen, Denmark, [2]The Novo Nordisk Fndn. Ctr. For Basic Metabolic Res., Univ. of Copenhagen, København, Denmark, [3]Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

Body mass index (BMI) and body fat percentage (BF%) are key metrics for assessing adiposity. While BMI is widely adopted for its simplicity, BF% provides a more nuanced view by distinguishing fat mass (FM) from fat-free mass (FFM). Despite their widespread use, the genetic differences underlying BMI and BF% and their implications for body composition and cardiometabolic health are not well-defined.

To shed light on genetic differences between BMI and BF%, we conducted genome-wide association studies comprising 417,959 UK Biobank participants. We identified 814 and 697 genome-wide significant ($p < 5x10-8$), independent loci ($r2< 0.01$ within ±1Mb) associated with BMI and BF%, respectively. Despite a high genetic correlation (0.85 in women, 0.81 in men) and many shared loci, tissue enrichment analyses revealed distinct biology, with genetic influences on BMI predominantly enriched in the brain whereas BF% was enriched in adipose tissue.

Next, we delved into differences in the effects of BMI and BF% loci on body composition and fat distribution by constructing weighted polygenic scores. Both scores were associated with increased FM; however, the BMI score uniquely associated with increased fat-free mass (FFM), bone mineral density, and hand-grip strength. In contrast, the BF% score was associated with a preferential deposition of subcutaneous fat over visceral fat, a feature not observed with the BMI score.

We then classified BMI and BF% loci into distinct body composition clusters based on their effects on FM and FFM. We identified two shared clusters between BMI and BF%, and one unique cluster to BF%. The shared clusters were associated with either increased FM alone or together with increased FFM. The unique BF% cluster showed association with increased FM but decreased FFM, which was tied to lower bone mineral density and handgrip strength. This cluster favored subcutaneous over visceral fat deposition and showed a protective effect against type 2 diabetes. Our findings underscore the distinct genetic architectures of BMI and BF%, offering new insights for advancing obesity research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1061F: DNA Methylation levels within human genome variable number of tandem repeats (VNTRs) ⭐

Authors:

D. Roshandel[1], **S. Mastromatteo**[1], S. Chirmade[1], B. Thiruv[1], S. Allana[1], K. Keenan[1], F. Lin[1], L. Strug[2,1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada

Abstract:

Variable number of tandem repeats (VNTRs) are a significant source of genetic variation. They have been understudied as their investigation is challenging with short-read sequencing. Little is also known about their DNA methylation levels. Pacific Biosciences (PacBio) HiFi long-read sequencing technology can simultaneously sequence ~20kb reads and estimate DNA methylation levels. Here, we aimed to use this technology to investigate VNTRs across autosomal chromosomes, their DNA methylation levels, and the factors that may affect their methylation. PacBio HiFi sequencing (30X) was performed on blood-derived DNA from individuals with Cystic Fibrosis (CF) recruited into the Canadian CF Gene Modifier Study (CGMS). We have previously shown that the genome-wide genotype distribution in individuals with CF is similar to the European population with the exception of the *CFTR* locus. DNA methylation levels were estimated at each CpG site in HiFi reads using Primrose and site methylation probabilities were generated using pb-CpG-tools. Tandem Repeat Genotyping Tool (TRGT) and its repeat catalog were used to genotype VNTRs and estimate their methylation levels. CpGs were annotated according to protein coding genes using MANE, regulatory features using BioMart, and CpG islands using UCSC Table Browser. Seventy-six individuals (43 males & 34 females) with a mean (SD) age of 24 (13) years were included in the current analysis. There were 29,062,275 CpGs on Chr1-22 of which 47% were within coding genes. They were mostly methylated across the genome with the exception of CpG islands, promoter regions, 5' upstream of genes, and 5' UTRs where they were typically unmethylated. There were 218,881 VNTRs on Chr1-22 in the TRGT catalog. Of these, 40,916 (19%) had CpGs within their motif. The majority of CpG containing VNTRs (57%) were methylated in all 118 haplotypes, 42% were methylated in some and unmethylated in others, and only 1% were unmethylated in all haplotypes. There were 388 VNTRs whose lengths were correlated with their methylation levels accounting for multiple hypothesis testing (Spearman correlation coefficient $p < 1.2E-6$ (0.05/40,916)). Of these, lengths of 224 VNTRs (58%) were positively correlated with their methylation levels whereas lengths of 164 VNTRs (42%) were negatively correlated with their methylation levels. VNTRs contribute to disease susceptibility and the mechanism could be methylation. We have been generating a comprehensive resource on VNTRs including their

alleles, allele frequencies and DNA methylation levels in blood which investigators will be able to use to understand the contribution of VNTRs to diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1062F: Empower Users with Programmatic Access to dbGaP Data Using FHIR

Authors:

M. Ward, E. Moyer, J. Bloom, I. Fore, M. Feolo, V. Lyoshin, A. Russette, B. Kattman; NIH/NLM/NCBI, Bethesda, MD

Abstract:

Signed into law in 2016, the 21st Century Cures Act states that Electronic Health Records (EHR) be exchanged using FHIR (Fast Healthcare Interoperability Resources) format. FHIR offers benefits such as open-source server implementations and a standard API interface, facilitating faster server development and easier integration for application developers. While FHIR for EHR is widespread, the application of FHIR in genomic research data is still emerging.Comprising over 4 million study participants with genomic and or transcriptomic data, more than 450,000 demographic, clinical and exposure “phenotype” variables in ~2,800 studies the NCBI’s dbGaP provides free download or cloud access to an abundance of NIH- (and non-) funded research data. Since 2007, over 8,000 manuscripts have cited the use of dbGaP data. However, obtaining access to these data involves request and approval as well as a technical expertise. Accessing dbGaP datasets via FHIR API both simplifies data delivery and more importantly provides interoperability. In this presentation, we will highlight some key benefits of the new dbGaP FHIR API and provide links to GitHub sample scripts demonstrating the benefits of the dbGaP FHIR API. The open-access API serves study level public information, for researchers to discover datasets most relevant to their needs before applying for access. The dbGaP FHIR API provides programmatic “dataset discovery” offering flexibility and interoperability beyond the existing Advanced Search web UI. For instance, with minimal Python script, users can easily create specialized alerts for new studies of interest or develop custom web interfaces listing dbGaP public information.In support of the new NIH federated data cloud computing ecosystem, the controlled-access API aims to deliver over 1.1 billion individual phenotypic values with access to file locations to the rich molecular and sequence data files using GA4GH DRS API. Historically, dbGaP is built to archive submitted data as collected, lacking harmonization of the phenotypic data. We will show how the FHIR API, with the HL7 CodeSystem resource, represents a

significant step toward metadata harmonization across studies. Harmonization facilitates searching across studies and “cohort building” for participants that match specific phenotypic criteria (e.g., African American women over 40 y/o with bp>130). This method of cohort building increases research power and saves time.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1063F: Enhancing LRRK2 Detection in Parkinson's Disease through Direct-to-Patient Genetic Testing and Community Engagement Strategies

Authors:

H. O'Brien, L. Wahlstrom-Edwards, S. Shah, L. Conroy, P. Short; Sano Genetics, London, United Kingdom

Abstract:

Background: LRRK2:G2019S variants occur in approximately 1% of sporadic Parkinson’s Disease (PD) cases and 4% of familial PD cases. This variant has been identified in 3% of participants undergoing genetic testing in PD research programs and is the focus of 40+ current clinical trials. In 2021, we initiated a genetic testing program to identify individuals with the G2019S variant using research-informed community engagement strategies and at-home testing.
Methodology: We developed a multi-channel recruitment strategy targeting regions in the United States with high populations of Ashkenazi Jewish and North African descent, ethnicities with a higher prevalence of genetic PD. Our outreach efforts included collaboration with healthcare providers, community organizations, digital platforms, patient-facing and advocacy organizations, and online publications. We conducted Whole Exome Sequencing and data were analyzed following GATK best practices to identify pathogenic and likely pathogenic variants.
Results: Over a 9-month period, 1,504 individuals were screened, 1,322 were eligible for testing, and 835 completing genetic testing. Participants who registered via direct-to-consumer outreach were split between digital marketing (54%) and partner channels (46%). Participants were enrolled in a virtual waiting room to receive personalized communications while their results were processed. The average time to return results was 5 weeks. Of those tested, 5.2% (n=43) individuals had a positive genetic finding, including 37 with the G2019S variant, 2 with R1441C, and 4 novel likely pathogenic variants. 3.1% of participants (n=26) had a variant of uncertain significance (VUS). Notably, those who registered through the research-informed direct-to-patient (DTP) strategy had a 5.8% LRRK2+ finding rate, nearly double the industry

norm, while HCP-ordered tests reflected average rates with 1.8% LRRK2+ findings.
Conclusion: Research-informed DTP recruitment campaigns coupled with at-home genetic testing and ongoing participant engagement achieved nearly double the industry-standard rate of variant identification in a genetic PD study. This approach can be applied to other studies to reduce the time and cost of recruitment for precision medicine research programs.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1064F: Expression landscape of TE-derived transcripts in human adult tissues

Authors:

B. Zhang, M. Benpeng; Washington Univ. in St. Louis, St. Louis, MO

Abstract:

Transposable elements (TEs) are significant components of eukaryotic genomes and play essential roles in genome evolution. Although most TEs were silenced in the mammalian genome, mounting evidence indicated that some specific TEs actively involved in gene regulation in the early developmental stage. Our previous study revealed that TEs could derive tissue-specific expression of protein-coding and lincRNA genes in the mouse tissues development. Correspondingly, how human TEs derive genes' transcription in distinct tis- sues largely remains unclear. In this study, we systematically analyzed the 17,329 human transcriptome data generated from the GTEx project, to explore how TEs drive the tran- scription of human genes in 47 male and female adult tissues. Our study described how TEs control gene expression in human tissues, and comprehensively evaluated TE's contri- bution to gene regulatory networks.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1065F: GENE TARGET 2.0: A refined tool for evaluating monogenic neurodevelopmental disorders for therapeutic development

Authors:

R. Belanger Deloge[1], A. Sveden[1], Z. Frazier[1], H. Osika[1], C. Byrne[2], K. Dies[1], S. J. Brewster[1], B. Zhang[1], L. Prock[1], J. Hubbs[1], E. D. Buttermore[1], W. K. Chung[1], M. Sahin[1], M. Chopra[1]; [1]Boston Children's Hosp., Boston, MA, [2]MGH Inst. of Hlth.Professions, Boston, MA

Abstract:

Recently, a framework to evaluate the translational building blocks for gene-based therapeutic development for monogenic neurodevelopmental disorders (NDDs) was proposed (Chopra et al, 2022). This framework, called GENE TARGET was designed to identify gaps in the translational pipeline for a given disorder to inform prioritization of future research efforts for both academic translational research centers and patient advocacy groups (PAGs). Considerations falling into the domains of genetic mechanism, preclinical modeling, clinical trial, and ethical factors were scored. The composite score, Gene Target Suitability (GTS Score), with a maximum possible score of 40, quantified therapeutic readiness for a given monogenic NDD. Given the preliminary nature of this framework, it has not been subjected to validation or systemic use in the translational research setting, and inter-user variability in scoring is unknown. Here, we used GENE TARGET to evaluate 50 NDD gene-disease pairs nominated by collaborators including Simons Searchlight, researchers, and PAGs. Through this exercise, we identified strengths and limitations of the GENE TARGET framework which informed iterative refinements of the tool for real-world applications, culminating in GENE TARGET 2.0. Additional numerical score categories were added to 3 criteria to capture more disease-gene pair nuances. For 2 criteria, score descriptions were clarified, but the numerical scoring categories remained the same. The remaining 5 criteria did not change. A resource guide was developed to aid users in their search for information when generating GENE TARGET 2.0 scores. Currently, we are evaluating the inter-user variability of GENE TARGET 2.0 using a subset of 20 NDD genes and 5 raters. In preliminary assessments, the ethics domain demonstrates the greatest variability amongst users. Although the refined framework remains focused on gene-based therapeutic suitability for scoring purposes, points to consider for medicinal chemistry approaches to treatment will be outlined. GENE TARGET 2.0 will serve as a more refined, validated tool for the evaluation of monogenic NDDs for therapeutic readiness.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1066F: Genetic Determinants of Type 1 Diabetes: A Comprehensive GWAS and Local Biobank Initiative in Ukraine

Authors:

T. Oleksyk[1], W. Wolfsberger[1], K. Shchubelka[1], L. Turova[2], O. Kurutsa[3], O. Dubiv[3], S. Patskun[4], O. T. Oleksyk[5], Y. Hasynets[4], T1D Ukraine Consortium; [1]Oakland Univ. - Biology, Rochester, MI, [2]Dept. of Clinical Immunology and Allergology with Med. Genetics Section, Natl. Bogomolets Med. Univ., Kyiv, Ukraine, [3]Uzhhorod Natl. Univ., Dept. of Computer Sci., Uzhhorod, Ukraine, [4]Uzhhorod Natl. Univ., Dept. of Biology, Uzhhorod, Ukraine, [5]A. Novak Transcarpathian Regional Clinical Hosp., Uzhhorod, Ukraine

Abstract:

There is a need to discover and evaluate rare and local genetic determinants of Type 1 Diabetes (T1D) that are often overlooked in broader studies from other populations. Through a comprehensive genome-wide association study (GWAS), we are identifying novel genetic factors linked to T1D specific to Ukraine. This project has led to the development of a T1D Collaborative Research Network, a local biobank, all contributing to an open-access genetic database with tens of thousands of exomes, whole genome genotypes, and phenotypes that fosters further regional collaboration. As part of this project, we have built the necessary infrastructure and cultivated local expertise in genomics locally. We describe the progress of this work in the context of the current state of genome data availability from Eastern Europe, the progress of our research from building a map of genetic diversity in Ukraine and bordering countries, and explore ethical and logistic considerations of collecting DNA, as well as the challenges of big data analysis and sharing. This initiative not only enhances our understanding of T1D but also aims to create a more equitable landscape in genomic research, ultimately improving disease prevention and treatment strategies. To date, we have collected and sequenced exomes of 3,666 T1D patients and 5,094 matching controls in Ukraine and conducted a genome-wide association study to discover new genetic factors that affect the development and pathology of T1D. Preliminary analyses highlight the significant role of HLA haplotypes and other genes, reinforcing the utility of our cohort for further research. These discoveries could help predict disease, identify therapies to delay progression and improve the lives of those living with T1D in the region.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1067F: Genetic diversity and structural complexity of the killer-cell immunoglobulin-like receptor gene complex: A comprehensive analysis using human pangenome assemblies

Authors:

T-K. Hung[1], W-C. Liu[1], S-K. Lai[2,1], H-W. Chuang[1], Y-C. Lee[1], H-Y. Lin[1], C-L. Hsu[2,1], C-Y. Chen[1], Y-C. Yang[1,2], **J. Hsu**[1], P-L. Chen[1,2]; [1]Natl. Taiwan Univ., Taipei, Taiwan, [2]Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Abstract:

Background: The killer-cell immunoglobulin-like receptor (*KIR*) gene complex, a highly polymorphic region of the human genome that encodes proteins involved in immune responses, poses strong challenges for genotyping due to its high genetic diversity and structural intricacy. Accurate analysis of *KIR* alleles, including their structural variations, is crucial for understanding their roles in various immune responses. Leveraging the high-quality genome assemblies from the Human Pangenome Reference Consortium (HPRC), we present a novel bioinformatic tool, the Structural *KIR* annoTator (SKIRT), to investigate KIR gene diversity and facilitate precise *KIR* allele analysis.
Results: We applied SKIRT on 47 HPRC-phased assemblies and identified a recurrent novel *KIR2DS4/3DL1* fusion gene in the paternal haplotype of HG02630 and maternal haplotype of NA19240. Additionally, SKIRT accurately identifies eight structural variants and 15 novel nonsynonymous alleles, which were independently validated using short-read data or quantitative polymerase chain reaction (qPCR). In total, our study discovered 570 novel alleles, including eight haplotypes that harbored at least one *KIR* gene duplication, six haplotypes that lost at least one *KIR* framework gene, and 75 out of the 94 haplotypes (79.8%) that carried at least five novel alleles, thus confirming KIR genetic diversity.
Conclusions: High-resolution personal genome assemblies provide an unprecedented opportunity to dissect polymorphic genetic regions that short-read sequencing cannot. Our findings provide valuable insights into *KIR* gene diversity and a solid foundation for the functional consequences of *KIR* structural variations. The SKIRT pipeline demonstrates remarkable efficiency in detecting the complete spectrum of *KIR* alleles in human genome assemblies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1068F: Haplotype-resolved characterization of repeat expansions and patterns of methylation from 1000 Genomes ONT Consortium data

Authors:

S. Gibson[1], J. Gustafson[1], N. Damaraju[1], M. P. Zalusky[1], K. Hoekzema[1], J. Goffena[1], S. H. Storz[1], S. Ward[1], J. Knuth[1], M. Sinha[1], Z. Anderson[1], The 1000 Genomes ONT Sequencing Consortium, H.

Dashnow[2], F. Sedlazeck[3], M. Loose[4], M. Jain[5], E. Eichler[1,6], L. Starita[1], D. Miller[1]; [1]Univ. of Washington, Seattle, WA, [2]Univ. of Utah, Salt Lake City, UT, [3]Baylor Coll. Med., Houston, TX, [4]Univ. of Nottingham, Nottingham, United Kingdom, [5]Northeastern Univ., Boston, MA, [6]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA

Abstract:

Expansions of short tandem repeats (STRs) and changes in DNA methylation patterns are both causes of rare genetic disorders. It has been previously difficult to establish normal patterns of both from databases such as the 1000 Genomes Project (1KGP) due to the limitations of short read sequencing. However, long reads are able to simultaneously span repeat expansion loci and provide epigenetic modification information bringing new potential to the data available from this diverse cohort. We have re-sequenced more than 250 1KGP individuals on the ONT platform to an average coverage of 30x and read N50 of 40 kbp. For the first 100 samples, analysis of known disease-causing repeat expansions, including characterization of motif sequence alleles and haplotype-resolved repeat lengths, has revealed expected levels of variation, but has also included the identification of some premutation alleles in this presumably healthy control set, which would not have been attainable with short read sequencing. These data will allow us to filter and prioritize variants in individuals who are unsolved after standard genetic testing and to compare expansion sizes to benchmarks when assessing the ability for the expansion to be tolerated. The 1KGP collection assemblies also have the potential to identify novel repeat expansion sites. Separately, LRS data enables concurrent analysis of haplotype-resolved methylation patterns, enhancing our understanding of the impact of CG-rich repeat expansions on methylation and gene expression. We have globally quantified CpG methylation, confirming known differential methylation regions (DMRs) on autosomes including the loci associated with Prader-Willi and Beckwith-Wiedemann syndromes. Additionally, we assessed methylation variation in X chromosome haplotypes of 46,XX individuals, developing computational methods for determining X chromosome inactivation status. Our established database for common DMRs in 1KGP samples will facilitate rapid identification of methylation differences related to imprinting or X-linked disorders using automated analysis pipelines.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1069F: How Open Targets resources use Genetics data to facilitate target identification and prioritisation

Authors:

A. Buniello[1,2], I. Lopez[1,2], D. Suveges[1,2], C. Cruz[1,2], M. Bernal Llinares[1,2], P. Uniyal[1,2], J. Xiangyu[3,2], D. Considine[3,2], K. Tsukanov[1,2], C. Mehta[1,2], R. Martinez-Osorio[1,2], J. Hayhurst[1,2], J. Ferrer[1,2], G. McNeill[1,2], S. Szyszkowski[3,2], H. Cornu[1,2], Y. Tsepilov[3,2], D. Ochoa[1,2], D. Hulcoop[1,2], E. M. McDonah[1,2]; [1]EMBL-EBI, Cambridge, UK, United Kingdom, [2]Open Targets, Cambridge, UK, United Kingdom, [3]Wellcome Sanger Inst., Cambridge, UK, United Kingdom

Abstract:

Open Targets is an industry-academia partnership that uses human genetics and genomics data to facilitate systematic drug target identification and prioritisation. Its two flagship open source resources are the Open Targets Platform (https://platform.opentargets.org/), integrating public domain data to enable target identification and prioritisation and Open Targets Genetics (https://genetics.opentargets.org/), identifying targets based on GWAS and functional genomics.Target-disease associations with supporting genetic evidence are more likely to succeed; however, the challenge arises with connecting evidence from GWAS loci to the likely disease-causing genes in a systematic way. Open Targets Genetics (OTG) was established five years ago as a resource to address this challenge and provide post-GWAS analysis for the scientific community, bringing together data from human GWAS with functional genomics datasets (eQTLs, pQTLs and sQTLs). These large scale genetics and functional genomics datasets feed into a machine learning model, developed to provide a predicted locus-to-gene (L2G) score to help prioritise the most likely disease-causing genes from GWAS. We have recently released Gentropy (https://opentargets.github.io/gentropy/), a modular toolkit developed for the community to facilitate analysis and interpretation of GWAS and functional genomics studies.The L2G scores are a key source of genetic evidence that is integrated with >20 different data sources feeding into the Open Targets Platform to provide target-disease associations. As well as its L2G predictions for common traits, rare disease and somatic variation, evidence from clinical resources such as ClinVar, PanelApp and ClinGen are integrated, scored and combined into the Platform to provide extensive genetic evidence for target-disease associations. Gene Burden test results that aggregate rare and ultra-rare variants at the gene level to provide gene to phenotype associations. Earlier this year, we introduced a new Pharmacogenetics widget in the Platform, integrating data from PharmGKB. This new dataset examines the link between genetics and drug response, helping in the prioritisation of drug targets that minimise the risk of adverse effects. The Platform users can now also use our new “Associations on the Fly'' tool to dynamically change the weighting of contributing evidence from each source, altering the prioritisation of targets. This new feature also introduces navigation through target-specific attributes in a disease agnostic way.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1070F: Impact and evolution of a philanthropic global network providing clinical genomic testing

Authors:

S. Terry[1], E. Thorpe Venti[1], M. Bennett[2], S. Urbaniak[2], V. Nelakuditi[2], A. Ramakrishnan[2], A. Chawla[2], A. Coffey[2], B. Thomas[2], D. Perry[2], A. Kesari[2], R. Rajkumar[2], J. Lowry[2], M. Jones[3], M. del Campo[3], D. Basel[4], J. McCarrier[4], L. Schmidt[5], E. Lotsof[6], M. Dueñas-Roque[7], N. Urraca[8], C. Brown[8], F. Sirchia[9], E. Giorgio[9], H. Peña Salguero[10], A. Lumaka[11], G. Mubungu[11], P. Lukusa[12], A. Vanderver[13], A. Pizzino[14], O. Sherbini[14], J. Bazalar-Montoya[15], M. Cornejo-Olivas[16], K. Milla Neyra[16], M. Shinawi[17], P. Magoulas[18], S. Wiafe[19], T. Arsov[20], D. Masser-Frye[21], P. Jayakar[22], A. Serize[23], A. Taylor[24], S. Shenbagam[24], T. Nyakambangwe[25], R. Gjergja Juraski[26], C. Marshall[27], K. Boycott[28], V. Jobanputra[29], S. David[30], **R. Taft**[31]; [1]Genetic Alliance, Damascus, MD, [2]Illumina Inc, San Diego, CA, [3]Rady Children's Hosp., San Diego, CA, [4]Children's Wisconsin, Med. Coll. of Wisconsin, Milwaukee, WI, [5]Sanford, Fargo, ND, [6]Rare Genomics Inst., Los Angelas, CA, [7]Hosp. Natl. Edgardo Rebagliati Martins, EsSalud, Lima, Peru, [8]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [9]Univ. of Pavia, Cuneo, Italy, [10]Padrino Fndn., Todos los Santos, Mexico, [11]Univ. de Kinshasa, Kinshasa, Congo, Democratic Republic of the, [12]Univ. de Kinshasa, Kinshasha, Congo, Democratic Republic of the, [13]Children's Hosp. of Philadelphia, Philadelphia, PA, [14]The Children's Hosp. of Philadelphia, Philadelphia, PA, [15]Inst. Natl. de Salud Nino-San Borja, Lima, Peru, [16]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [17]Washington Univ. Sch. of Med., Saint Louis, MO, [18]Texas Children's Hosp., Houston, TX, [19]Rare Disease Ghana Initiative, Accra, Ghana, [20]Goce Delcev Univ., Stip, North Macedonia, The Republic of, [21]Rady Children's Hosp., San Diego, CA, [22]Nicklaus Children's Hosp., Miami, FL, [23]South Miami Hosp., Miami, FL, [24]Al Jalila Children's Specialty Hosp., Dubai, United Arab Emirates, [25]Child and Youth Care Zimbabwe, Harare, Zimbabwe, [26]Children's Hosp. Srebrnjak, Zagreb, Croatia, [27]The Hosp. for Sick Children, Toronto, ON, Canada, [28]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [29]New York Genome Ctr., Paramus, NJ, [30]Mendelics, Sao Paulo, Brazil, [31]Genetic Alliance, San Diego, CA

Abstract:

Rare genetic diseases (RGD) impact at least 250 million individuals worldwide, the majority of which reside in low- and middle-income countries (LMIC) and experience prolonged diagnostic odysseys due to barriers accessing specialist clinicians and molecular testing. The iHope program is a philanthropic program that provides no-cost clinical genome sequencing (cGS) to underserved children with suspected RGD. We have previously reported data on 1,004 patients supported by the iHope program, and here add an additional 1,028 probands with suspected

RGD referred from a total global network of 31 clinical sites. Forty-six percent of probands (943/2032) were referred from one of ten sites in LMIC including Peru (3), Mexico (2), Croatia (1), Democratic Republic of the Congo (1), Ghana (1), North Macedonia (1), and Zimbabwe (1). Consistent with prior analysis, the diagnostic yield was 40.4% (820/2032) overall, with probands from LMIC more likely to receive a diagnosis (50.9%, 480/943) compared to probands from HIC (31.3%, 341/1090). These data indicate that the iHope model is robust, with impacts that are durable and reproducible over time, which supports the extension and democratization of the program to serve an increased volume of patients.
To support this expansion, iHope was transferred to Genetic Alliance 501(c)3 in January 2024, with the goal of developing a network of testing laboratories and partner clinical sites. Participating laboratories receive subsidized materials to provide no-cost cGS or whole exome sequencing (WES) to patients referred from an iHope clinical site. A public request for applications inviting laboratories to apply was released in February and May 2024, with an aim to engage three laboratories in 2024, with at least one residing in a LMIC. A total of 29 laboratories applied and were evaluated based on defined criteria, which considered certification and accreditation, a clinical cGS or WES test offering, and the corresponding test capabilities and volume commitments. The awardees were Sick Kids and Children's Hospital of Eastern Ontario in Canada, the New York Genome Center in the United States, and Mendelics in Brazil. More than 500 patients are expected to be tested through the program by early 2025. Current awards support a sequencing-by-synthesis workflow, but the model is technology-agnostic and can support multiple testing platforms and modalities. Over the next five years, iHope Genetic Health aims to scale to include 10 laboratories and 50 clinical sites, with a focus on growth in LMIC and enablement of in-region or in-country testing.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1071F: Impact of Bias in Variance Analysis: From SNPs to Pathways

Authors:

B. Queme, T. Mushayahama, A. Mushayahama, M. Ye, H. Mi; Univ. of Southern California, Los Angeles, CA

Abstract:

Single Nucleotide Polymorphisms (SNPs), the most prevalent genetic variations, are pivotal in understanding human genetics and disease susceptibility. Accurate SNP annotation is essential

for elucidating their functional implications, impacts on gene expression, and associations with phenotypic traits, crucial for advancing personalized medicine and genetic disorder research. Previous studies, using subsets of SNPs, have shown that different annotation tools lead to varying gene mappings. However, a comprehensive, genome-wide study comparing these tools has not been conducted until now. In this study, we perform a genome-wide analysis to evaluate the consistency of SNP-to-gene annotations using SnpEff, ANNOVAR, and VEP. Statistically significant differences in the gene mappings produced by these tools (p-value < 0.001) were identified. These discrepancies highlight a potential source of bias that can significantly impact downstream analyses. Specifically, pathway overrepresentation analyses yield varying results depending on the annotation tool used, leading to inconsistent and potentially misleading conclusions Our findings underscore the importance of considering tool-specific biases in SNP-to-gene annotation processes. We illustrate the gaps between these tools and their downstream consequences in pathways overrepresentation analysis. Addressing these biases is critical for ensuring genomic research's accuracy, reliability, and reproducibility, enhancing the interpretation of genetic data and its applications in personalized medicine and genetic disorder studies. To mitigate these biases, we have developed a pipeline that allows users to choose a comprehensive approach integrating multiple tools or select a specific tool tailored to their needs; providing researchers with a more holistic view of their results, promoting more robust and reliable genomic analyses.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1072F: Impact of COVID-19 vaccination status and strain-specific infection on risk of developing dementia in the National COVID Cohort Collaborative (N3C) Data Enclave

Authors:

C. Rose, W. S. Bush, D. C. Crawford; Case Western Reserve Univ., Cleveland, OH

Abstract:

Late-onset Alzheimer’s Disease (LOAD), the most common cause of dementia and a leading cause of US deaths, affects 1 in 9 people ≥65 years of age. While several risk factors have been identified, including advanced age, female sex, and host genetics, they neither individually nor collectively fully explain LOAD risk. Exposure to specific infectious agents has long been suspected to play a role, but their study has been difficult given the uncertainty in timing of exposure relative to subsequent development of LOAD. The emergence of the novel SAR-CoV-

2/COVID-19 in early 2020 provides a unique opportunity to assess a specific infectious agent as a risk factor for LOAD development and severity. We accessed the National COVID Cohort Collaborative (N3C), one of the largest centralized de-identified collections of individual-level clinical data available for COVID-19 research, to identify older dementia patients with and without clinically documented COVID-19 infection. At the time of data access (v175), 84 US institutions had contributed clinical data representing 22.7 million (M) patients with 5.2 billion (B) medical records. We identified patients ≥65 years of age with and without COVID-19 according to N3C criteria. Of the 5.58M patients ≥65 years, the majority were white (73.8%) or Black (11.4%) race and non-Hispanic ethnicity (86.7%). A higher proportion of white patients had a COVID-19+ test or diagnosis (38.2%) compared with Black patients (33.4%; $p<2.2x10^{-16}$), and a greater proportion of COVID-19+ patients was female (white: 53.4%; Black: 59%) compared with male (both $p<2.2x10^{-16}$). Less than half (35.1%) of patients ≥65 years of age had evidence of at least one dose of COVID-19 vaccination in the EHR, with similar proportions between white (35.44%) and Black (35.57%) populations. N3C identified 400,480 patients classified as having dementia (1.25%), and consistent with the literature, dementia cases were majority female (57.7%). Interestingly, compared with the general ≥65 years of age group, dementia cases had lower rates of COVID-19 vaccination (24.23%) yet similar rates of COVID-19 infection (37.5% versus 37.6%). Viral strain data obtained from privacy preserving record linkage with sequencing centers are available for 6,440 patients ≥65 years. Individual-level host genetic data are not available. Work supported by 5T32AG071474 and 1UM1TR004528 is ongoing to execute a data use request to access a de-identified dataset for further modeling of pre-pandemic clinical data to further describe possible associations between vaccination status and strain-specific infection with development of dementia among older patients.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1073F: Impact of Gene-Disease Validity on Variant of Unknown Significance rates in hereditary cancer panels

Authors:

E. Ryan, J. Gryzbowski, **D. Thrush**, D. Kostecki, C. Horton; Ambry, Aliso Viejo, CA

Abstract:

Gene-Disease Validity scores (GDV) measure the strength of evidence that exists to support the association of a particular gene with a particular condition. GDVs are utilized to make decisions about the context in which each gene should be offered, as well as how variants in the gene

should be classified and reported. Here we review 3 iterations of an expanded-phenotype hereditary cancer panel (EX-P) to evaluate how rates of Variants of Unknown Significance (VUS) have changed with new knowledge, updated GDVs, and changes to panel content.
From September 2018 to June 2024 there have been 3 versions of EX-P (V1, V2, and V3). Seven genes found on V1 were once thought to have an association with heritable cancer but were eventually removed due to disputed evidence of pathogenicity and subsequently weakened GDV. These genes are *BLM*, *FANCC*, *GALNT12*, *MRE11A*, *NBN*, *RAD50*, and *XRCC2*.
Similarly, *RECQL* was added to EX-P V2 and removed with the update to V3 based on evolving evidence and GDV, and so is included in our list of 8 eventually removed genes (RGs).
It is worth noting that changes beyond gene removal are made to panel contents with each update. Genes are also added for both previously indicated and new phenotypes. For example, the addition of *EGFR* for a new lung cancer indication on V2, as well as *MSH3* for the existing colorectal cancer indication. To account for these additions, as well as any laboratory changes, we compared each panel to a theoretical version of itself with RGs removed rather than comparing the VUS rate of V1 to V2 to V3.
V1 of EX-P included 7 of the 8 RGs and the VUS rate was 34.7%. Had RGs not been present, the VUS rate would have been 28.9%. *MRE11A* and *RAD50* were removed during the update to V2 while *RECQL* was added, and the VUS rate was 37.1%. This rate would have been 33.4% had RGs not been present. The latest update, V3, lacks all RGs and shows a VUS rate of 32.2%. The lower VUS rate of V3 when compared to V1 and V2 is particularly noteworthy as the population of patients being tested has shifted since 2018 with more clinically and ancestrally diverse patients undergoing hereditary cancer testing, including those who are unaffected. This change to the testing population is expected to contribute to an increase in VUS rates.
These perhaps seemingly small percentage differences equate to hundreds of patients each year who are not given a VUS result. VUS results are notoriously complicated for both patient understanding and provider counseling, and there is general agreement in the field that efforts should be made to reduce the volume of VUS results released. Utilizing GDV in the curation of panels is one way to help accomplish this and to maximize the clinical utility of panels.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1075F: Insights from Comprehensive Fertility Biobank with Deep Phenotyping and Genomic Data in East Asian Population ⭐

Authors:

Y. Shi[1], L. Majara[2], B. Cai[3], D. Yuwei[3], T. Cao[3], Y. Xu[3], A. Martin[4]; [1]Broad Inst., Cambridge, MA, [2]Univ. of Cape Town, Cape Town, South Africa, [3]Sun yat-sen Univ., Guangzhou, China, [4]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Background:While biobanks have grown rapidly in popularity, fertility and reproductive health-focused biobanks are still rare. Fertility centers have uniquely valuable biospecimens, hormonal assay test results, ultrasound images, in vitro cultures, and longitudinal clinical data, which can be used to study fertility, pharmacogenomics, and embryonic development.
Methods:This study included infertile women undergoing in vitro fertilization (IVF) at the fertility center of the First Affiliated Hospital of Sun Yat-sen University, China. We performed deep exome sequencing and developed a multiplex PCR assay for screening genes with large effects on fertility. Comprehensive clinical information, including basic characteristics, hormone levels, oocytes sizes via ultrasound during ovulation stimulation, embryo culture data, and pregnancy outcomes were recorded for further analysis.
Results:To date, we have enrolled 1,183 infertile women in this study, with phenotypic data from over 2K ovarian stimulation cycles including growth information from ~ 15K in vitro oocytes, ~11K embryo culture cycles, and over 1K pregnancy cycles recorded. The genomic database contains allele frequencies of 2.02 million variations (1.907M single-nucleotide polymorphisms and 370K insertions/deletions). A genome-wide association analysis (GWAS) of ovulation stimulation response identified novel signals in or near the FSHB (rs6169, P = 2.71*10-24) and LHB genes (rs201674286, P = 4.67*10-8). To develop a personalized ovulation stimulation dosage model, we incorporated these associations into a drug response prediction model and improved performance by 41.5%. Utilizing embryo culture data, we identified an association between blastocyst formation rates and variants in the PLK4 gene (Pmin = 6.15*10-7). Rare variants analysis showed that bi-allelic GDF9 variants (His209GlnfsTer6/S428T, Q321X/S428T) contributed to the defect in antral follicle development. Additionally, two TUBB8 mutations (R320H, A54V) are associated with developmental arrest in human oocytes and early embryos. We further validated its causal role in mouse experiments.
Conclusion:Reproductive centers possess rich phenotypic resources and genomic data that can be used to decipher genetic contributions to infertility mechanisms, tailor treatment plans, design drug targets, and uncover mechanisms of embryonic development. Establishing fertility biobanks is essential for enhancing personalized care for patients and providing valuable-insights into reproductive health management.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1076F: Introducing Human Induced Pluripotent Stem Cells (iPSCs) to the NHGRI Sample Repository for Human Genetic Research

Authors:

L. Scheinfeldt, S. Sampson, T. McCalla, L. Willis, E. Horton, M. Mitchell, T. Pozner; Coriell Inst. for Med. Res., Camden, NJ

Abstract:

The NHGRI Sample Repository for Human Genetic Research (NHGRI Repository) established by NHGRI in 2006 at the Coriell Institute for Medical Research facilitates studies of human genetic and genomic variation by establishing, characterizing and distributing a large (over 3,700) and diverse publicly available collection of renewable biospecimens from thousands of people living around the world. The NHGRI Repository includes lymphoblastoid cell lines and DNA samples associated with the International HapMap Project, 1000 Genomes Project and the Human Pangenome Reference Consortium. Participants that have generously donated to the NHGRI Repository consented to their samples and associated data being used for a wide range of general research uses and to public data sharing of largescale genomic data. Through the 1000 Genomes Project, the majority of this collection (N>3200) has been characterized with publicly available whole genome sequencing data and a subset of samples have been characterized with RNA-Seq data (N>460). In addition, web-based search tools of single variant genotypes, multiple variant genotypes, pharmacogenetic and HLA annotations, and lymphoblastoid cell line gene expression profiles can be accessed at https://www.coriell.org/1/Browse/Genomic-Data-Search. In late 2023, the NHGRI Repository incorporated human induced pluripotent stem cells (iPSCs) into the collection for the first time. To date, four iPSCs have been added to the collection and are available to the research community. The iPSCs were reprogrammed using the Sendai Virus method and subjected to thorough quality control evaluation. Their inclusion in the NHGRI Repository represents a significant addition, providing a valuable resource for future research. Additional information about the NHGRI Repository can be found at https://catalog.coriell.org/NHGRI.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1077F: Investigating methylation patterns associated with obesity in *Danio rerio*

Authors:

A. Walz, K. Dobrinski; Univ. of Tampa, Tampa, FL

Abstract:

DNA methylation is an epigenetic pattern that maintains and changes the integrity and expression of genes. Mammalian methylation commonly occurs at CpG locations where cytosine is adjacent to guanine nucleotides. Patterns of genomic methylation can be quantified with Reduced-representation bisulfite sequencing (RRBS). Bisulfite converts unmethylated cytosines to uracils but is ineffectual on methylated cytosines allowing for differential analysis. This study's purpose is to use zebrafish (Danio rerio) to generate a model for human obesity in extracted livers to compare methylation patterns between obese and control groups. This study used RRBS to observe methylation changes and how they affect subsequent gene expression in zebrafish with induced obesity. Methylation patterns have been shown to change in response to physiological stress and this study measures how obesity can stimulate subsequent changes in methylation patterns. The RRBS data adapter sequences were trimmed with Trim Galore and aligned to a reference genome using Bismark. Following alignment methylation status analysis using SeqMonk and the R package: Methylkit was conducted. GO Analysis was used to analyze the differential expression of specific genes and how that manifests changes in biological function in zebrafish and correspondent human orthologs. This study has revealed distinctly different methylation profiles comparing obese and controlled zebrafish livers inferring a connection between obesity and methylation regulation. Loss of typical methylation maintenance mechanisms may contribute to obesity associated diseases such as non-alcoholic fatty liver disease, type 2 diabetes, hypertension, as well as other diseases including various types of cancer.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1078F: Investigation of febrile seizure susceptibility and motor function in *Magel2* mutant mice as a model for Schaaf-Yang syndrome

Authors:

T. Iwaki, D. Ieda, M. Fujimoto, E. Sato, Y. Nakamura, Y. Negishi, A. Hattori, S. Saitoh; Nagoya city Univ., Graduate Sch. of Med. Sci., Nagoya, Japan

Abstract:

Background: The *MAGEL2* gene is located in the Prader-Willi syndrome critical region at 15q11-q13 and is the causative gene for Schaaf-Yang syndrome (SYS), a neurodevelopmental disorder

caused by paternally inherited heterozygous truncating variants. SYS is characterized by hypotonia in the neonatal period, feeding difficulties in infancy, developmental delay, and joint contractures. We have previously reported multiple cases of SYS patients presenting with encephalopathy-like symptoms with neurological sequelae following febrile seizures. Additionally, to elucidate the pathophysiology, we generated a SYS model mouse with a frameshift variant (c.1690_1924del; p.(Glu564Serfs*130)) introduced in the paternally inherited allele of the *Magel2* gene. The *Magel2* mutant mouse resembles SYS because most pathogenic variants detected in SYS are truncating variants, while previously established SYS model mice carry deletions of the entire *Magel2* gene. **Objective:** To determine whether the *Magel2* mutant mouse can replicate the encephalopathy phenotype observed in human SYS patients. **Methods:** Mice carrying the paternally inherited *Magel2* mutation as well as littermate controls were uses. At 10 days of age, these mice underwent a febrile seizure model experiment using a high-temperature chamber. Subsequently, at 8 weeks of age or older, the rotarod test was conducted to compare the impact of febrile seizures on post-seizure motor function between control male mice and *Magel2* mutant male mice. **Results:** In the febrile seizure model experiment, *Magel2* mutant mice tended to experience seizures earlier than normal mice, yet no significant difference was observed in the rotarod test. **Conclusion:** This experiment did not clearly demonstrate a relationship between febrile seizures and motor function, nor did it replicate encephalopathy-like phenotypes. Nevertheless, the increased susceptibility to febrile seizures in *Magel2* mutant mice suggests the need for further research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1079F: Leveraging the *All of Us* platform to enable rapid genomic analysis of primary autoimmune hypothyroidism in diverse populations

Authors:

A. Babbar[1], H. Mo[1], A. Williams[1], D. Schlueter[1], C. Zeng[1], J. Keaton[1], S. Goleva[1], T. Cassini[1], T. Tran[2], J. Denny[1]; [1]NIH, Bethesda, MD, [2]NIH, Bethesda, MD

Abstract:

The *All of Us* Research Program is a longitudinal cohort collecting a diversity of data including genomic, electronic health record (EHR), survey, and digital health technology data from diverse participants across the United States. With the goal of increasing representation of underrepresented groups in biomedical research, *All of Us* emphasizes accessibility and democratization of analysis tools. The *All of Us* Researcher Workbench uses graphical interfaces

for common research processes such as cohort building and dataset generation and grants users access to high-powered compute clusters as well as access to R, Python, SQL and other analysis tools.
To demonstrate the utility of *All of Us* in accelerating genomic science and incorporate diverse participants in genomic analyses, we set out to replicate a genome-wide association study (GWAS) and phenome-wide association study (PheWAS) of primary autoimmune hypothyroidism (PH), whose phenotype was derived via an EHR algorithm. Two case-control studies were built; the first used only the *All of Us* Cohort Builder tool, which offers a user-friendly GUI to query *All of Us* datasets (the "CB study"); and the second used SQL in the Python notebook interface (the "SQL study"). We compared the case-control groups built by both of these approaches. There was 86% case overlap between the two final approaches; however, the SQL case algorithm required significantly more time to implement and refine compared to the CB algorithm. We performed sex and ancestry-stratified GWAS using Hail, a pre-installed cloud-compatible analytic tool for genomic analyses. GWA studies included 9,575 cases and 56,585 controls, of which 1,815 cases and 22,969 controls were of non-European ancestry. We replicated all 4 significant SNPs at genome-wide significance near *FOXE1* that were identified in the original study in trans-ancestry and European ancestry GWAS and identified >9500 SNPs in other PH loci. PheWAS confirmed that thyroid conditions were associated with rs965513 from the prior study. Most peak SNPs identified in the trans-ancestry GWAS shared direction of effect across sex and ancestry-specific GWAS. Leveraging the accessibility-oriented tools provided by *All of Us*, we implemented this study design in *All of Us* in a significantly smaller timeframe with fewer researchers compared to earlier studies. This study demonstrates the utility of *All of Us* in accelerating genomic studies in diverse populations for novice and experienced researchers alike through the use of accessibility-focused tools, sensible data organization using the OMOP schema, and access to large compute clusters.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1080F: Long-read based detection of large copy number variants with potential functional significance using the ContextSV structural variant caller

Authors:

J. Perdomo[1,2], J. Akoto[1,3], N. Akizu[1,3], K. Wang[1,3]; [1]CHOP, Philadelphia, PA, [2]Drexel Univ. Sch. of BioMed. Engineering, Philadelphia, PA, [3]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

Abstract:

Long read sequencing technologies such as the Oxford Nanopore (ONT) and Pacific Biosciences (PacBio) platforms provide the resolution required to resolve larger and more complex SVs. Widely used structural variant callers developed for long read sequencing data, such as Sniffles2, cuteSV and PBSV, have limitations in the size and complexity of the structural variants (SVs) they can identify with high confidence, largely due to a focus on alignment information. We show that SVs identified with these tools are generally less than 50kb (most under 10kb) in length, and therefore large, potentially disease causal SVs may be overlooked. To overcome these limitations, we developed ContextSV, a novel structural variant calling method which combines long read alignment information and copy number predictions using a PennCNV Hidden Markov Model (HMM) adapted to long read sequencing data. This HMM uses parameters empirically estimated previously from known copy number variant (CNV) regions to predict a copy number state based on changes in coverage and expected SNP allele frequencies. We leverage the ethnicity-specific allele frequencies available in the gnomAD database to improve the accuracy of these predictions. SVs are then clustered using DBSCAN and merged using a ranking system based on breakpoint support from alignments and read depth. We recently published a study using this approach to successfully validate the presence of 5 large, previously unidentified CNVs with potential functional consequences in the KOLF2.1J induced human pluripotent stem cell line (iPSC), a reference line used for modeling neurological diseases. Thus, characterizing the potential functional significance of large CNVs in reference iPSCs is crucial for guiding the interpretation of data derived from these models, and for identifying additional suitable reference lines. Here we show that ContextSV can be used for this purpose: We expand on the KOLF2.1J analysis by running whole-genome CNV detection on ONT PromethION R10.4.1 30x long read sequencing data, followed by annotation of exonic CNVs with potential functional significance using ANNOVAR, and manual validation in IGV. From comparison of CNV callsets with Sniffles2, PBSV, cuteSV, and DELLY (ONT mode), it is clear that many >10kb CNVs missed by these callers are accurately identified using contextSV via our combined-evidence approach.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1081F: Mining for gold standard gene sets: maximizing the accuracy and utility of effector gene predictions from GWAS results

Authors:

M. Costanzo[1], L. W. Harris[2], A. McMahon[2], J. Flannick[3], N. P. Burtt[1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]European Molecular Biology Lab., European Bioinformatics Inst., Hinxton, United Kingdom, [3]Boston Children's Hosp., Boston, MA

Abstract:

Although genome-wide association studies (GWAS) have now been performed for hundreds of common diseases, progress in the use of these results towards understanding disease mechanisms has been slow, because it is challenging to link genetically associated variants to the genes that they affect. GWAS researchers often approach this challenge by integrating multiple types of genetic and genomic evidence to prioritize genes and predict the causal, or "effector," gene at each locus. The resulting lists of predicted effector genes (PEGs) have the potential to greatly increase the utility of GWAS results, helping researchers to formulate hypotheses for functional testing and serving as "gold standard" training sets for bioinformatic methods.

We reviewed GWAS literature from the past decade and identified 128 papers that included PEG lists across 127 traits. The evidence types supporting the predictions can be grouped into 2 major categories: variant-centric (e.g., eQTLs) and gene-centric (e.g., mutant phenotypes of model organism orthologs). However, within these general categories, the types of supporting evidence vary widely between lists. Furthermore, where multiple prioritization efforts have been performed for the same trait, there is little concordance between the PEGs identified. Far from providing gold standards, this proliferation of methods and evidence sources has the potential to make PEG lists potentially more confusing than illuminating. Furthermore, the lists are difficult to find, access, and interpret, and are often not machine-readable.

Based on these findings, we see two urgent needs for common disease researchers: 1) an open-access resource for PEG lists and 2) standards for content, format, and interoperability of PEG lists. To address the first need, we have created the Predicted Effector Genes Knowledge Portal (PEGKP; pegkp.org), which includes PEG lists for 115 unique traits from 50 publications, as well as 5 unpublished PEG lists contributed by the research community. PEGKP represents the lists in an interactive, downloadable table format that includes the evidence supporting each prediction. To address the need for standards, with our collaborators at the GWAS Catalog we have conducted a community survey and are jointly planning an open workshop in September 2024 to gather input from PEG list creators and consumers. This presentation will include a summary of the results from the survey and workshop.

PEGKP was developed as part of our work on the NHGRI community resource, the Association to Function Knowledge Portal (A2FKP; a2fkp.org) and was supported by NHGRI U24HG011453.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1083F: Pipeline palooza: Streamlining multi-omic discoveries across biobanks

Authors:

Z. Rodriguez[1], L. Guare[1], L. Caruth[2], K. Cardone[1], A. Ikuzwe[1], M. Condiff[1], C. Carson[1], A. Garofalo[1], X. Li[1], K. Keat[1], R. Kumar[1], T. Odia[1], C. Morse[1], S. Verma[3], A. Verma[1], T. Drivas[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Perelman Sch. of Med. Inst. of BioMed. Informatics, Philadelphia, PA, [3]Univ. of Pennsylvania, Cherry Hill, NJ

Abstract:

The field of genomics is undergoing a transformative era, driven by the rapid expansion of biobanks like the UK Biobank, NIH's All of Us Research Program, the Million Veteran Program, and Penn Medicine BioBank. These large-scale repositories, which pair genetic data with deep phenotyping from electronic health records and questionnaires, are accelerating genomic discovery at an unprecedented pace and enabling systematic studies of genetic and environmental influences on diseases and traits globally. Despite the move towards cloud computing to share data, biobanks face significant technical hurdles that slow their potential. Data is often siloed and researchers have to navigate a technical maze to use different platforms and tools. This hinders the speed of research, leads to unnecessary replication of data analysis, and results in mixed, unscalable practices that cast doubt on the findings of genomic studies. Here, we demonstrate the approaches taken at the PMBB to address these critical issues within bioinformatics. We have developed a series of scalable, portable, and reproducible pipelines for conducting common genomic analyses like GWAS, ExWAS, and PheWAS, publicly available on GitHub. Using Nextflow workflow language, we tested, containerized, and deployed our platform-agnostic pipelines for seamless data analysis across diverse computing environments, including the NIH's All of Us, DNAnexus, and several high-performance computing environments.

Our work showcases the successful integration and deployment of these pipelines, emphasizing their flexibility and robustness across different platforms. By ensuring that our pipelines are seamlessly operable in these varied environments, we aim to streamline data analysis, reduce redundancy, and enhance the reproducibility of genomic research. This effort not only accelerates the pace of discovery but also fosters a more collaborative and efficient research community. Our contribution stands as a testament to the potential of innovative bioinformatics solutions in overcoming the technical barriers that have historically hampered genomic studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1085F: Polygenic risk scores and risk of drug-induced liver injury following initiation of antiretroviral therapy in people living with HIV

Authors:

Z. Cindi[1], K. M. Cardone[1], Y. Bradford[1], H. Kim[2], VA Million Veteran Program, E. S. Daar[3], R. Gulick[4], S. A. Riddler[5], M. Vujkovic[6,1], K-M. Chang[6,1], P. S. Tsao[7,8], G. Maartens[9], P. Sinxadi[9], D. W. Haas[2,10], M. D. Ritchie[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Lundquist Inst. at Harbor-UCLA Med. Ctr., Torrance, CA, [4]Weill Cornell Med., New York, NY, [5]Univ. of Pittsburgh, Pittsburgh, PA, [6]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [7]VA Palo Alto Hlth.Care System, Palo Alto, CA, [8]Stanford Univ. Sch. of Med., Stanford, CA, [9]Univ. of Cape Town, Cape Town, South Africa, [10]Meharry Med. Coll., Nashville, TN

Abstract:

Background: Drug-induced liver injury affects some people living with HIV (PWH) following initiation of antiretroviral therapy (ART). The contribution of genetic susceptibility to liver injury in this setting is not known. We examined whether polygenic risk score (PRS) models could help predict the likelihood of liver injury within 48 weeks following initiation of ART among PWH.**Methods:** As a reference population, we derived PRS_{ALT} using summary statistics from the Million Veteran Program, regardless of HIV status. We filtered variants based on ALT associations with $P \leq 0.05$. With the remaining variants, we separately constructed PRS for individuals of European (n=68,725) and African (n=13,387) ancestry using PRS_{csx} (continuous shrinkage priors across populations). We applied these models to treatment-naïve participants in prospective, randomized ART studies of the AIDS Clinical Trials Group (ACTG, n=5114). ALT elevations are graded as follows; grade 1: <2.5X upper limit of normal (ULN), grade 2: >2.5-5X ULN, grade 3: >5X ULN. Participants with baseline ALT grade >2 were excluded from analyses. Cases with liver injury were defined as having ALT grade >3 within 48 weeks following initiation of ART. Controls had ALT grade ≤1 from baseline to week 48. We assessed for associations in participants of European (31 cases, 1975 controls) and African (23 cases, 1795 controls) ancestry. Multivariable logistic models applied separately within each group included PRS, sex, age, body mass index, and 5 genetic principal components.**Results and Conclusion:** In multivariable models, PRS was not associated with grade >3 ALT elevation following initiation of ART in PWH. This finding suggests that inherent genetic susceptibility may not contribute greatly to drug-induced liver injury in this setting.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1086F: PopCNVBrowser: a web-based Platform for CNV analysis and visualization using 1000 Genomes Project Data

Authors:

D. Kim[1], J. Oh[1], N. Park[2], Y. Ko[1], J. Kim[2]; [1]Hankuk Univ. of Foreign Studies, Seoul, Korea, Republic of, [2]Konkuk Univ., Seoul, Korea, Republic of

Abstract:

A copy number variant (CNV) is a type of genetic variation caused by the duplication or deletion of specific genome regions. CNV plays a crucial role in genetic diversity and disease susceptibility. Given the importance of CNV, numerous computational CNV calling methods have been developed. However, predicted CNV regions often yield inconsistent results, leading to the necessity for a platform capable of systematically comparing and integrating diverse CNV calls generated from different methods. The 1000 Genomes Phase 3 project includes 3202 samples from 22 populations, providing crucial data for understanding human genetic diversity. Despite the significance of this dataset, there is a lack of web-based visualization tools specifically designed for analyzing CNVs within this extensive data. To address this need, we propose PopCNVBrowser, a web-based platform designed for the CNV visualization and analysis from the 1000 Genomes Project Phase 3 data. PopCNVBrowser provides CNV call sets for all samples from four state-of-the-art CNV calling methods: ReadDepth, cn.mops, CNVKIT, and FREEC. These results are visualized through the web-based platform, allowing researchers to compare and integrate predicted CNV data. Using the extensive CNV dataset from the 1000 Genomes Project, PopCNVBrowser also provides population-specific CNVs and various disease susceptibility associated with these CNV regions. Additionally, our platform provides the comparison with known CNV regions using DGV database, gene annotation information, and pathway analysis. Beyond comparing results from different CNV calling programs, PopCNVBrowser provides an integrated and comprehensive resource of population-level CNV data and extended analyses. Although, our current version only supports the 1000 genome data, it can easily extend to other large-scale genomic data such as UK biobank. Our platform allows researchers to gain insights into the functional implications of CNVs across diverse populations, facilitating cross-ethnic comparisons and understanding the genetic basis of diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1087F: Predicting GWAS causal variants from individually assayed regulatory effects

Authors:

K. Dura, K. Strouse, K. Siklenka, S. Morrow, T. Reddy; Duke Univ., Durham, NC

Abstract:

Experimental studies of the effects of non-coding genetic variation on gene regulatory activity have been instrumental to fine mapping mechanisms contributing the traits and disease. To help make those mechanistic studies more routine, we recently developed high-throughput approaches to measure the effects of non-coding genetic variation on gene regulatory element genome-wide. Here, we report initial results from a population-wide analysis of non-coding genetic variation using high-throughput reporter assays. Our study experimentally assayed over 27 million genetic variants across 100 individuals from the Thousand Genomes Project. We have focused our initial analysis on a subset of 5 of those individuals, with almost 14 million total variants, and ~360,000 variants in ATAC-seq peaks. Those variants include 42,417 insertions and deletions and 9,256 multi-allelic sites. Combinations of individual regulatory effects of variants can be used to explain the effect sizes seen in whole-genome GWAS studies, both when assessing caQTLs (r2=0.61) and eQTLs (r2=0.21). These individual variant effects find smaller genomic regions of significance than those defined by QTLs, indicating the ability to narrow down smaller sets of variants that might be causal to the trait associated with the QTL. Using this analysis, only 1-3 variants in a single open chromatin region can be considered potentially causal for any one region, whereas many variants could be correlated in a GWAS study. We further assessed potential mechanisms for these regulatory effects by examining their relationships to transcription factor binding. Of 147 transcription factors, 5 have correlations between the regulatory effect size of the variant and the change in the PWM score as caused by the variant, which was used as a proxy for binding affinity. Further, to support the analyses reported here, we have made algorithmic advances, including new methods for assessing individual regulatory effects that were assayed over multiple experiments and a novel alignment method. Taken together, we see these population-scale allele-specific reporter assay as a valuable resource for fine-mapping molecular mechanisms contributing to many complex traits and disease.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1088F: Quadrupia: Derivation of G-quadruplexes for organismal genomes across the tree of life

Authors:

A. Nayak[1], N. Chantzi[2], F. A. Baltoumas[3], E. Aplakidou[3], S. Liew[4], J. Galuh[4], M. Patsakis[2], A. Montgomery[2], C. Moeckel[2], I. Mouratidis[2], S. Sazed[2], W. Guiblet[5], P. Karmiris-Obratański[6], G. Wang[7], A. Zaravinos[8], K. M. Vasquez[7], C. Kwok[4], G. A. Pavlopoulos[3], I. Georgakopoulos-Soares[2]; [1]Pennsylvania State Univ., State College, PA, [2]Pennsylvania State Univ. Coll. of Med., Hershey, PA, [3]Inst. for Fundamental BioMed. Res., BSRC "Alexander Fleming", Vari, Greece, [4]City Univ. of Hong Kong, Kowloon Tong, Hong Kong, [5]Frederick Natl. Lab. for Cancer Res., Frederick, MD, [6]AGH Univ. of Krakow, Krakow, Poland, [7]Univ. of Texas at Austin, Austin, TX, [8]European Univ. Cyprus, Nicosia, Cyprus

Abstract:

G-quadruplex DNA conformations play crucial roles in key biological functions, including transcription, replication, telomere maintenance, and genomic instability, and have been an essential component of organismal evolution. However, no database provides G-quadruplex maps for organismal genomes across the tree of life, creating a gap in evolutionary research on G-quadruplexes. Here, we bridge that gap by developing Quadrupia, the largest and most comprehensive database for G-quadruplexes to date. In this database, we provide all the G-quadruplexes identified in the genomes of 108,449 organisms using a simple motif-based algorithm and a more sophisticated G4Hunter algorithm, resulting in 30,306,672 and 109,874,605 G-quadruplexes respectively. Further, we validate the most frequently observed G-quadruplexes using CD-spectroscopy, UV melting, and fluorescence-based methods. We observe a large variability in the distribution and density of G-quadruplexes across different taxa. Examination of the derived G-quadruplexes across functional genomic elements in different taxa indicates preferential orientation and positioning, with significant differences between individual taxonomic groups, suggesting their intrinsic regulatory roles in various molecular functions and biological mechanisms. We also perform clustering analysis of G-quadruplexes based on their sequence similarity and examine the frequency of the clusters across taxonomic groups and functional genomic elements in each genome. Quadrupia integrates motif-based and BLAST search methods, enabling efficient searching of G-quadruplexes at organismal and taxonomic levels and facilitating evolutionary research. It can also enable large-scale studies of G-quadruplexes to better understand the mechanisms underlying their role in essential biological processes like DNA replication, telomere maintenance, and genome stability, potentially contributing to our understanding of diseases such as cancer.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1089F: Reproducible single cell annotation of programs underlying T-cell subsets, activation states, and functions

Authors:

M. Curtis[1], D. Kotliar[2], R. Agnew[1], K. Weinand[1], A. Nathan[3], Y. Baglaenko[4], Y. Zhao[5], P. Sabeti[6], D. Rao[1], S. Raychaudhuri[7]; [1]Brigham and Women's Hosp., Boston, MA, [2]Brigham and Women's Hosp., Cambridge, MA, [3]Harvard Univ., Boston, MA, [4]Cincinnati Children's Hosp., Cincinnati, OH, [5]Harvard Med. Sch., Boston, MA, [6]Harvard Univ, Cambridge, MA, [7]Brigham and Women s Hosp., Brookline, MA

Abstract:

Cells concurrently express multiple gene expression programs (GEPs) corresponding to lineage, activation states, lifecycle processes, and responses to external stimuli. Contrary to this, single-cell RNA sequencing (scRNA-seq) data is typically annotated by discrete clustering, or transfer of discrete labels which are unable to capture the multifaceted usage of GEPs by a cell. Here, we address these limitations by introducing CellAnnoTator (*CAT), a pipeline which discovers a reproducible catalog of GEPs and simultaneously quantifies their activity in novel datasets. We demonstrate this approach using T-cells, which are known to simultaneously express multiple activation, polarization, and other functional GEPs. From 1,700,000 T-cells from 700 individuals across 38 tissues and five diverse disease contexts, we produce the most comprehensive T-cell GEP catalog to date, encompassing 46 reproducible GEPs. Our catalog reflects the known core functions of T-cells including cytotoxicity, exhaustion, antigen-specific activation, and T helper effector states. We experimentally characterize several novel activation programs and apply TCAT to describe T-cell activation and exhaustion in Covid-19, cancer, and rheumatoid arthritis. We identify associations of GEPs with disease states and highlight the enrichment of T peripheral helper (Tph) cells in tumors. Our approach and software provides a starting point for a global catalog of GEPs for single-cell annotation that span cellular functions across cell types, tissues, and diseases. We demonstrate how these resulting single-cell annotations can be used in practice for disease studies and single cell molecular QTL analyses.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1090F: Systematic Evaluation of Polygenic Score Performance Across Diverse Population Characteristics in the All of Us Research Program

Authors:

A. Hendricks, **J. Staples**, S. White, M. Lin, M. Fisher, C. Gignoux, N. Pozdeyev; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Polygenic Scores (PGS) are useful for predicting diseases and refining diagnoses. Despite potential benefits, use of PGS have risks including high uncertainty for individuals and reduced accuracy when the prediction sample varies from the training sample. Poor generalizability is well documented for differences in genetic ancestry and there is growing appreciation that PGS predictive performance also varies by other sample characteristics (e.g., sex, age). Evaluations thus far have been narrowly focused in conditions and sample characteristics and have been applied in homogeneous samples. Here, we evaluate predictive performance of existing PGS catalog scores across a variety of sample characteristics in the diverse All of Us (AoU) cohort.
We extracted variants from v7 AoU short read whole genome sequencing data using Hail and calculated PGS for a variety of continuous traits (e.g., body mass index, education) and categorical conditions (e.g., rheumatoid arthritis, type 1 diabetes T1D). We compared PGS by attributes such as size and training sample diversity (i.e., one or multi-ancestry) and evaluated performance across sample characteristics (i.e., sex, age, gender, race, ethnicity, genetic ancestry, income, education, geographic region) both marginally and by two-way intersections (e.g., sex by race) using incremental R^2 for continuous traits and incremental area under the curve (AUC) for categorical conditions within each ancestry group and controlling for age, sex, and principal components.
We find PGS performance often differs substantially by score attribute and sample characteristics. Some PGS have lower performance in AoU cohort compared to the initial report such as a T1D PGS with an AoU AUC of 0.64 compared to previous reports of 0.78 to 0.87. We find the reduced performance may be due to differences in the distribution of sample characteristics within AoU. While some trends are expected, such as the T1D AUC decreasing as age quartile increases (0.69 Q1, 0.61 Q2, 0.53 Q3, and 0.50 Q4), others are surprising such as the highest AUC for inferred European ancestry (0.53 African, 0.65 European) and for self-identified White individuals (0.63 White, 0.54 Black) despite the PGS being trained in an African-like sample. Many of these surprising trends can be explained by the intersection of sample characteristics.
Here, we provide insights into PGS performance for a variety of traits and conditions across a

spectrum of demographic and biological variables in the AoU cohort. Our work will inform the development of more accurate and equitable genetic risk assessments, enhancing the potential of personalized medicine across diverse populations.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1091F: The Helmholtz Munich imputation server: an EU based, GDPR compliant imputation server ⭐

Authors:

N. Rayner[1], Y-C. Park[1], C. Fuchsberger[2,3], A. Barysenka[1], E. Zeggini[1,4]; [1]Helmholtz Zentrum München, Munich, Germany, [2]Eurac Res., Bolzano, Italy, [3]Med. Univ. of Innsbruck, Innsbruck, Austria, [4]Technical Univ. of Munich and Klinikum Rechts der Isar, Munich, Germany

Abstract:

Genomics has the potential to transform healthcare through personalised disease management, including precision prevention. Genome-wide association studies (GWAS) have been instrumental in generating new biological insights into common complex human diseases. The power of GWAS can be increased by increasing sample size through meta-analysis, which requires the imputation and analysis of genotypes at untyped variants across studies. Imputation relies on the availability of phased haplotype reference panels of whole genome sequenced individuals. Due to data privacy and security reasons, as well as size, these are not usually available to be used locally by researchers. Imputation servers have, therefore, been developed to provide a solution: researchers upload their genotyped dataset to the imputation server hosting the reference panels and imputation machinery, where the data are imputed, and downloaded back to the researchers' individual local compute. Currently, three imputation servers serve the majority of the global community of researchers: two based in the US (University of Michigan and TOPMed) and one based in the UK (Wellcome Sanger Institute). The need for an EU-based imputation server arises from restrictions imposed by the General Data Protection Regulation (GDPR), a comprehensive data privacy law in the EU. Genetic data are considered a special category of personal data under the GDPR and are, hence, subject to strict data sharing rules and safeguards. Uploading genotype data to non-EU imputation servers constitutes a breach of GDPR, unless explicitly covered in individual study informed consent forms or specific legal agreements such as standard contractual clauses. This has generated an urgent need for an EU-based imputation server. Here, we introduce the Helmholtz Munich Imputation Server (https://imputationserver.helmholtz-munich.de/), designed to provide a cost-

free genotype imputation service in a GDPR compliant manner for EU-based researchers.
The server utilises the publicly available Cloudgene software, as utilised in the Michigan and TOPMed servers to provide a familiar user experience as well as efficient processing of the imputation jobs. Currently the server provides the Haplotype Reference Consortium (HRC) and the 1000 genomes GRCh37 and GRCh38 reference panels, and is able to impute over 20,000 genomes per day on the HRC panel.
Going forward, a strong focus will be to increase the available haplotype reference panels both in terms of sample size and scale, and in terms of diversity of global populations represented.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1092F: The NeuroBioBank Whole-Genome Catalog : sequencing from human brain donorswith nervous system disorders

Authors:

D. Hupalo[1], J. L. McCauley[2], M. Wilkerson[1], C. L. Dalgard[1], W. K. Scott[2], NIH NeuroBioBank; [1]Uniformed Services Univ., Bethesda, MD, [2]Univ. of Miami, Miami, FL

Abstract:

Nervous system disease is a prevailing cause of morbidity and death worldwide. This set of neurological and psychological diseases have diverse causes aided by unknown genetic risks. Here we generated genome-wide genetic data on a large cohort of brain tissue donors with well annotated medical histories, allowing for broad investigations into the heritable risk of developing these neurological, neurodevelopmental, and psychological conditions. This resource consists of 9,663 donors with whole-genome genotyping and 9,543 donors with whole-genome sequencing (WGS) completed. The clinical diagnoses of these donors include 148 nervous system diseases clustered in 15 broad categories by ICD-10 coding. These donors were collected by six repositories comprising the NIH NeuroBioBank, with an average participant age of 60 years old. While primarily older individuals of European descent, the cohort also contains younger donors and individuals from non-European backgrounds. Full demographic and phenotypic descriptions for all donors are available via the NeuroBioBank Portal (neurobiobank.nih.gov). Variants detected by WGS were jointly called genome-wide and annotated to describe their functional impact; this identified 1,047,815 non-silent mutations. To illustrate applications of this resource, we replicated the strong association observed in previous studies between pathogenic CAG repeat expansions in the HTT gene with the clinical diagnosis of Huntington's disease.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1093F: The repertoire of short tandem repeats across the tree of life

Authors:

I. Georgakopoulos-Soares, N. Chantzi; PSU, Hershey, PA

Abstract:

Short tandem repeats (STRs) are widespread, dynamic repetitive elements with a number of biological functions and relevance to human diseases. However, their prevalence across taxa remains poorly characterized. Here we examined the impact of STRs in the genomes of 117,253 organisms spanning the tree of life. We find that there are large differences in the frequencies of STRs between organismal genomes and these differences are largely driven by the taxonomic group an organism belongs to. We found an extremely high variance between domains, with eukaryotic organisms having the highest STR density and archaea the lowest. Interestingly, similar variation in STR density was identified upon partitioning the viruses with respect to the host domain. Viruses of eukaryotic hosts had a disproportionately higher STR density than the corresponding viruses of bacterial and archaeal hosts. *O*n the kingdom level, Protista had the highest average STR density followed by Animalia. Upon further partitioning on the phylum level, Euglenozoa had the highest average STR density followed by several other eukaryotic phyla. Using simulated genomes, we find that on average there is no enrichment of STRs in bacterial and archaeal genomes, suggesting that these genomes are not particularly repetitive. In contrast, we find that eukaryotic genomes are orders of magnitude more repetitive than expected. STRs are preferentially located at functional loci at specific taxa. Finally, we utilize the recently completed Telomere-to-Telomere genomes of human and other great apes and find that STRs are highly abundant and variable between primate species, particularly in peri/centromeric regions. We conclude that STRs have expanded in eukaryotic and viral lineages and not in archaea or bacteria, resulting in large discrepancies in genomic composition.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1094F: Unraveling Cardiovascular Disorders: Integrative Analysis of GWAS-Phewas and Single-cell Data

Authors:

D. Sharma[1], P. S. Tsao[1], S. Clarke[2], T. Assimes[3]; [1]Stanford Univ., Palo Alto, CA, [2]Stanford Univ, Stanford, CA, [3]Stanford Univ Sch Med., Palo Alto, CA

Abstract:

The integration of genome-wide association studies (GWAS) with single-cell RNA sequencing (scRNA-seq) marks a significant advancement in delineating the complex genetic landscapes of diseases, particularly emphasizing the inclusion of underrepresented non-European populations. This combined approach facilitates the precise identification of gene expression within key tissues and cells, clarifying the influence of genetic risk variants at a cellular level. The innovative single-cell Disease Relevance Score (scDRS) method is pivotal in this analysis, as it quantifies polygenic disease signal enrichment across individual cells, pinpointing those with increased expression of disease-linked genes. This approach is supported by insights from a population-stratified phenome-wide GWAS conducted on 635,969 participants from the diverse U.S. Veteran population through the Million Veteran Program (MVP). This study identified 38,270 unique genetic variants associated with a variety of traits, many of which are novel in non-European populations, underscoring the importance of diversity in genetic research. Our specific investigation utilized scRNA data from the coronary arteries of six patients to compute scDRS for over 150 circulatory system-related phenotypes, revealing critical associations, such as fibromyocytes in ischemic heart disease and coronary atherosclerosis. We also found fibroblasts and plasma cells in thrombophlebitis. Future directions include constructing a variant-trait-gene network to deepen the understanding of circulatory system phenotypes and integrating single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) with GWAS and PheWAS. This will establish a comprehensive, cell-type-specific framework aimed at enhancing disease mechanism comprehension and fostering targeted, equitable therapeutic advancements. This holistic approach not only augments our grasp of complex genetic disorders but also ensures that research findings are applicable across diverse populations, thereby promoting equitable health outcomes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1095F: Unveiling the shared genetic architecture and causality between kidney dysfunction and bladder cancer: A genome-wide cross-trait analysis

Authors:

L. Wang[1,2,3], N. Xiang[1,2], Y. Yang[1], S. Huang[4], T. Fu[1,2,3], S. Su[1,2], G. Liao[5], L. Du[6], Y. Lin[1,7], J. Huang[1]; [1]Dept. of Urology, Innovation Inst. for Integration of Med. and Engineering, West China Hosp., Sichuan Univ., Chengdu, China, [2]Frontiers Sci. Ctr. for Disease-related Molecular Network, West China Hosp., Chengdu, China, [3]Key Lab. of Bio-Resource and Eco-Environment of Ministry of Ed., Coll. of Life Sci., Sichuan Univ., Chengdu, China, [4]West China Sch. of Publ. Hlth., West China Hosp., Sichuan Univ., Chengdu, China, [5]State Key Lab. of Oral Diseases, Natl. Clinical Res. Ctr. for Oral Diseases, West China Hosp. of Stomatology, Sichuan Univ., Chengdu, China, [6]Chinese Evidence-Based Med. Ctr., West China Hosp., Sichuan Univ., Chengdu, China, [7]Program in Genetic Epidemiology and Statistical Genetics, Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Background Clinical and epidemiological evidence suggests a potential association between kidney dysfunction and bladder cancer (BC). One hypothesis for this comorbidity is the presence of a common genetic etiology. However, little is known about the shared genetics and causality of this association. Therefore, we aimed to investigate the shared genetic architecture and causal link between kidney dysfunction and BC. **Methods** We retrieved summary statistics from GWAS studies. Kidney function traits included eGFRcrea (N=1,004,040), eGFRcys (N=460,826), BUN (N=852,678), UACR (N=288,649), and urate (N=547,361). BC summary statistics were from the Finngen consortium ($N_{cases/controls}$: 1115/175121) and a GWAS meta-analysis ($N_{cases/controls}$: 2242/412592). Genetic correlation analysis used LDSC with the 1000 Genomes Project data. We performed cross-trait meta-analysis using MTAG, functional annotation with FAVOR, colocalization analysis with "coloc," and TWAS with FUSION. Biological insights were gained through DAVID and FUMA tools. Bidirectional two-sample Mendelian Randomization (MR) analysis explored genetic causal relationships. **Results** Genetic correlation analysis found significant correlations between eGFRcys (Rg=-0.140, P=0.046), BUN (Rg=0.464, P=0.003), and urate (Rg=0.283, P=0.015) with BC. Cross-trait meta-analysis identified 157 loci associated with both kidney function traits and BC. For eGFRcrea and BC, 72 loci were identified, with significant loci near *CERS2*, *STC1*, and *MRTFA* genes. For eGFRcys and BC, 42 loci were identified, including *CST9L* and *STC1*. For BUN and BC, 17 loci were identified, including *LRIG1*, *PPP2R3A*, and *MEX3A*. For urate and BC, 19 loci were identified, including *WDR1*, *SLC22A12*, and *MIR548I2*. For UACR and BC, 7 loci were identified, including *FOXD2-AS1* and *MIR22HG*.

Colocalization analysis revealed four SNPs causally associated with kidney function traits and BC, including signals in the *SHH* and *AC021218.1* regions. TWAS identified significant gene-tissue pairs shared between kidney function biomarkers and BC. Notably, *ZFHX3* and *TTC33* genes were identified across all analyses. Bidirectional MR analysis found significant causal correlations between BUN and UACR with BC. **Conclusions** Our comprehensive cross-trait analysis demonstrates a shared genetic basis between kidney function and BC, providing novel insights into the biological functions and molecular mechanisms underlying these complex traits. This study highlights the importance of considering genetic overlap in understanding the comorbidity of kidney dysfunction and BC, offering potential avenues for targeted therapies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1096F: Upcoming version of Integrated Genome Browser will better support visualizing and exploring personal genome sequence data

Authors:

P. Kulzer, K. Gopu, N. Freese, A. Loraine; Univ. of North Carolina at Charlotte, Charlotte, NC

Abstract:

Direct-to-consumer genomic sequencing services are relatively inexpensive now, making them accessible for families and individuals to explore the health and ancestry implications of their own personal genomes. To help personal genome sequencing clients better understand their own data, we added new capabilities to the Integrated Genome Browser (IGB), a freely available open-source desktop genome browser that supports fast navigation through vast data scenes. IGB supports layering multiple data sources onto an annotated reference genome. It includes customizable “LinkOuts” to external Web sites, enabling users to easily navigate to external data sources with information about genes and polymorphisms. As of upcoming IGB version 10.1.0, IGB will be able to display genome sequence data imported from Compressed Reference-oriented Alignment Map (CRAM) files, used by genome sequencing providers to distribute data. Because IGB runs locally, CRAM files can reside on a user’s personal computer, without requiring prior upload to a server or website. We also developed an IGB App for formatting of SNP array data for display. By comparing whole-genome sequencing to SNP array data, users can better understand and sanity-check results.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1097F: Visual analysis of RNA-Seq data in the Integrated Genome Browser; why every researcher needs to view their data

Authors:

N. Freese, J. Sirigineedi, P. Kulzer, A. Loraine; Univ. of North Carolina at Charlotte, Charlotte, NC

Abstract:

Next-generation RNA sequencing (RNA-Seq) is a widely used and powerful method for detecting gene expression and understanding gene function. RNA-Seq data processing pipelines involve many steps, with many parameter choices along the way. Producing and processing RNA-Seq data requires diligent quality control to catch or avoid mistakes. One of the best ways to avoid mistakes is to visually examine the RNA-Seq alignments in a genome browser. To make visual inspection of RNA-Seq alignments more effective and convenient, we added several new features to Integrated Genome Browser (IGB), an open-source, highly interactive and customizable computational tool users can download and install from the BioViz project website. Here, we demonstrate how to use these IGB features to identify problems in RNA-Seq data arising from experimental or data processing errors. Examples to be presented include: visualizing soft-clipped reads in an RNA-Seq experiment where the reads failed to split across introns, layering data from multiple replicates to identify point mutations that indicate sample switching at a sequencing facility, and viewing RNA-Seq data where the intron length parameter was not set correctly, leading to chromosome-spanning introns. By documenting instances where visualization in a genome browser helps expose and ultimately correct mistakes, we hope to convince researchers that genomic visualization is essential to a successful RNA-Seq experiment.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1098F: WebGWAS: A web server for instant GWAS on arbitrary phenotypes

Authors:

M. Zietz[1], K. Karczewski[2], N. Tatonetti[3]; [1]Columbia Univ., New York, NY, [2]Massachusetts Gen. Hosp., Medford, MA, [3]Cedars-Sinai Med. Ctr., West Hollywood, CA

Abstract:

Complex disease genetics is a key area of research for reducing disease and improving human health. Genomewide association studies (GWAS) are an important method for identifying the regions of the genome that contribute to complex disease risk. However, GWAS are computationally intensive and require access to individual-level genetic and health information, which presents concerns about privacy and imposes costs on researchers seeking to study complex diseases. In previous work, we developed tools for approximating GWAS for arbitrary phenotype definitions using only summary statistics. Here, we present WebGWAS a new tool that allows researchers to produce GWAS summary statistics for a phenotype of interest without needing access to individual-level genetic and phenotypic data. Our tool can be used by researchers to define custom phenotype definitions, including inclusion and exclusion criteria, and to produce approximate GWAS summary statistics for that phenotype. Our server computes GWAS summary statistics very quickly (<5 seconds), and it does not store private health information. Overall, our tool provides a faster approach to GWAS for researchers interested in complex disease, allowing them to obtain approximate summary statistics in short order, without needing to collect, process, and produce GWAS results themselves. This will speed up disease research into complex diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Friday Session

Board 1099F: Investigation of Immune Related Gene Transcription Differences Associated with Obesity in Danio rerio (Zebrafish)

Authors:

A. DiPietro, K. Dobrinksi; The Univ. of Tampa, Tampa, FL

Abstract:

The United States faces an elevated prevalence of one of the most acknowledged chronic diseases, obesity. Obesity, in various eukaryotic organisms, is known to impair the immune defense system prompting a heightened susceptibility to illness and disease progression. Additionally, obesity is correlated to the development of several cancers. The Danio rerio, also known as zebrafish, is an exceptional model organism for humans. Within the zebrafish genome, there is 70% of the human genome identified, and 84% of human disease genes are equivalent to zebrafish. For this study, one fish group was fed a regular diet (0.006 g of artemia/day), and another fish group was fed a heavier diet (0.06 g artmenia/ day) over eight weeks. During the experiment, body mass index calculations were calculated on the regular diet fish ensuring an

appropriate control. The diet difference resulted in normal and obese fish respectively. The fish livers were dissected, and transcriptome sequencing was carried out. A heatmap was generated using R software, which displayed the obese genes that were the most significantly different from the normal genes. Many overexpressed and underexpressed genes in obese fish compared to normal fish were observed. The human ortholog, Interferon Induced 44 protein like (IFI44L), was underexpressed in all obese fish. IFI44L acts in various immune responses and is a tumor suppressor. When there is a suppression of IFI44L, there is an increase in chemoresistance, tumor size, and phosphorylation of Met. Met is a prominent component of several oncogenic cell signaling pathways such as the Met/Src, and Ras/MAPK pathways. Conversely, C-type lectin domains and hemoglobin genes were observed to be overexpressed in obese fish. C-type lectin domains are an innate immune response and may act as a compensational line of immunity. Hemoglobin is associated with oxidative stress in erythrocytes. Hemoglobin is a catalyst for free radical reactions for reactive oxygen species which can damage proteins that aid in proper cell regulation. Deciphering the genetic signatures linked to obesity will reveal potential triggers in obese individuals that influence disease susceptibility and extremity.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2002F: A comprehensive analysis of structural variations in 2,733 Orofacial Clefts (OFC) trios reveals novel genetic drivers ⭐

Authors:

N. Kurtas[1,2], A. Sanchis-Juan[1,2], S. Curtis[3], V. Jalili[4], B. Weisburd[5], A. Lee[1,2], K. Robinson[3], A. Alade[6], X. Zhao[1,2], J. Fu[1,2], K. Perez[3], L. Gowans[7], M. Eshete[8], W. Adeyemo[9], C. Buxó[10], C. Padilla[11], F. Poletta[12], G. Wehby[6], J. Hecht[13], J. Shaffer[14], L. Moreno Uribe[6], N. Mukhopadhyay[14], S. Weinberg[14], J. Murray[6], T. Beaty[15], A. Butali[6], M. Talkowski[1,2], M. Marazita[14], E. Leslie[3], H. Brand[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of Harvard and MIT, Cambridge, MA, [3]Emory Univ., Atlanta, GA, [4]Data Sci. Platform, Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Univ. of Iowa, Iowa City, IA, [7]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [8]Addis Ababa Univ., Addis Ababa, Ethiopia, [9]Univ. of Lagos, Lagos, Nigeria, [10]Dental and Craniofacial Genomics Core, Univ. of Puerto Rico Sch. of Dental Med., San Juan, PR, [11]Univ. of the Philippines Manila, Manila, Philippines, [12]Laboratorio de Epidemiología Genética, Buenos Aires, Argentina, [13]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [14]Univ. of Pittsburgh, Pittsburgh, PA, [15]John Hopkins Univ., Baltimore, MD

Abstract:

Orofacial Clefting (OFC) is characterized by the abnormal formation of the lip and/or palate and is among the most prevalent birth defects (1 in 700 births globally). Heritability studies indicate a strong genetic component, yet the etiology of most cases remains elusive. OFC is a common comorbidity in numerous copy number variant driven genomic disorders (GD, e.g 22q11.2Del), yet the impact of structural variants (SVs) outside of these GDs is much less understood. In this study, we aggregate and examine SVs from OFC sequencing data in 2,733 case-parent trios. Data were derived from several large scale sequencing efforts including the Gabriella Miller Kids First program (GMKF) and the Deciphering Developmental Disorder (DDD) study to encompass 1,790 genome and 943 exome based trios. All SVs were detected with the genome analysis toolkit (GATK) SV pipelines, and we also explored 2.1 million short tandem repeats (STRs) on a subset of 437 trios using the Expansion Hunter tool. We initially focused on analysis on *de novo* mutations, which represent the strongest drivers of pathogenicity in rare disease. Overall, we discovered 626 high-confidence *de novo* SVs in 459 affected samples in line with expectations based on previous studies. Among *de novo* SVs, 26 were detected in the following GD regions: 22q11.2, 15q11.2, 1q21, 2p16.3, 15q26, 16p11.2 and 8p23.1. We observed an additional 237 de novo loss-of-function (LoF) SVs with 16 disrupting established OFC genes; including non-canonical events like an ALU insertion disrupting *WDR35* and a 2 Mb inversion disrupting *SATB2*. We found several putative novel OFC genes with recurrent disruptions including 2 LoF mutations in *NBEA*, an established neurodevelopmental disorder gene that has been previously implicated by common variants in an OFC GWAS. Beyond LoF SVs, we observed 13 *de novo* SVs in OFC genes including; 2 copy number gains, 1 intronic deletion, 1 intronic duplication, 7 exonic duplications and 2 partial duplications. Functional follow up is necessary to fully understand how these mutations may contribute to OFC risk. We resolved 5 *de novo* reciprocal translocations and 18 complex SVs with the most complex event involving an insertion on chromosome 4 that had a 129 kb deletion at the breakpoint and two large translocated pieces from ~60 MB away of 174 kb and 464 kb. Finally, we observed 30 *de novo* STR expansions per proband, which showed an enrichment of *de novo* expansions located within overall genic space when compared to matched controls from the 1000 genomes ($p = 8.27 \times 10^{-4}$, OR = 1.09). Overall, our study demonstrates the importance of including SVs in genomic analyses and their relevance to the genetic underpinnings of OFC.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2003F: A *de novo WDR45* gene deletion in a pediatric patient with atypical beta-propeller protein-associated neurodegeneration (BPAN) features

Authors:

M. Slife[1,2], N. Semaska[1,2], B. Koch[2,3], A. Shillington[2,3], E. Sperry[3], G. Hijazi[3], R. Hopkin[2,3], K. Kremer[2,3], D. Brightman[3], Y. Wu[2,3]; [1]Univ. of Cincinnati, Cincinnati, OH, [2]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [3]Cincinnati Children's Hosp., Cincinnati, OH

Abstract:

Heterozygous or hemizygous pathogenic variants in the *WDR45* gene are associated with beta-propeller protein-associated neurodegeneration (BPAN), a type of X-linked neurodegeneration with brain iron accumulation. BPAN is a progressive neurological disorder characterized by global developmental delay, intellectual disability, epilepsy, and behavioral issues in infancy or early childhood, and adulthood neurological deterioration, dystonia, parkinsonism, and other movement impairments. Individuals with BPAN often present with abnormal brain imaging, which typically shows signs of iron deposition in the globus pallidus and substantia nigra, hypomyelination, and global atrophy. BPAN is caused by pathogenic *WDR45* variants, mostly *de novo* but can be inherited as well. The majority of BPAN individuals are females. The current proband is of an 11-year-old European American female with autism, intellectual disability, speech apraxia, and behavioral issues at the time of genetic consultation. Some subtle MRI findings suggest the possibility of brain iron accumulation. Previous genetic testing, including SNP microarray, *Fragile X* testing, and *MECP2* sequencing, were negative. By a comprehensive in-house autism and intellectual disability panel, this patient was shown to have a *de novo* likely pathogenic *WDR45* variant, c.414_419del; p.Glu138_Phe139del. This in-frame deletion was reported in a Japanese female patient previously; she had infantile febrile convulsions, developmental delay, intellectual disability, and young-onset parkinsonism: features typical of BPAN. While the current proband's autism/ID panel results were pending, the proband had spells indicative of seizures. Follow-up brain imaging showed symmetric hypointense magnetic susceptibility in the bilateral globus pallidus and the bilateral substantia nigra, suggestive of neuronal degeneration with brain iron accumulation. Thus, the diagnosis of BPAN is established based on her clinical, imaging, and molecular genetic findings. A previous study has shown that the diagnosis of *WDR45*-related disease is usually delayed due to its heterogeneous clinical presentations and complex disease history. The clinical presentation and progression of our patient are atypical of BPAN individuals, with the absence of febrile convulsions in infancy and seizure manifestation delayed until age 11, which is consistent with the lack of genotype-

phenotype correlation in *WDR45*-related disease. This case highlights the benefit of comprehensive genetic testing to minimize delayed diagnosis.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2004F: A drug repurposing screen reveals dopamine signaling as a candidate modifier pathway in *PIGA*-CDG ⭐

Authors:

M. Aziz, J. Wilson, C. Y. Chow; Univ. of Utah, Salt Lake City, UT

Abstract:

PIGA-CDG is a Congenital Disorder of Glycosylation (CDG) caused by loss-of-function variants in the *Phosphatidylinositol glycan class A (PIGA)* gene. *PIGA*-CDG typically presents with seizures, developmental delays, intellectual disability, and hypotonia. *PIGA* encodes an essential catalytic component for the first step of glycosylphosphatidylinositol (GPI) anchor biosynthesis. GPI-anchor biosynthesis is necessary for proper protein folding, trafficking, and membrane dynamics. GPI-anchored proteins (GPI-APs) localize to the cell surface and have diverse functions, including roles in signaling, adhesion, immunity, and more. There are over 150 different GPI-APs, making it difficult to determine how the loss of certain GPI-APs or pathways contributes to disease phenotypes. To investigate the pathophysiology of *PIGA*-CDG, we performed a drug repurposing screen on a *Drosophila* eye model of *PIGA*-CDG using a small molecule library consisting of 1,520 FDA/EMA approved drugs. Flies were raised on food containing 5 µM drug, and eye size was measured at adulthood. This screen yielded 90 candidate drugs that increased the eye size of the *PIGA*-CDG model ($Z \geq 1.5$). Dopamine targeting, tubulin targeting, adrenergic receptor targeting, anti-inflammatory, and steroid drug classes were enriched, collectively making up ~30% of the suppressor hits. Because patients have primarily neurological symptoms, we prioritized pathways with neurological relevance, such as dopamine signaling, for validation. Prior to this screen, dopamine signaling had not been linked to *PIGA* pathophysiology. I am using pharmacological and genetic tools to investigate dopamine signaling as a modifier pathway of *PIGA*-CDG. I will present data that demonstrates modifying dopamine receptor signaling, but not dopamine synthesis, alters the outcome of the *PIGA* model. Inhibition of the dopamine receptor D2 (D2R) using pharmacological and genetic tools improved the outcome of the *PIGA* model, while inhibition of the dopamine receptor D1 (D1R) worsened the outcome. Genetically inhibiting dopamine synthesis did not affect the model. This suggests that D2R signaling is overactive in *PIGA*-CDG and may be a promising target for therapeutic development. I will also present data on how

inhibition of D2R and D1R signaling affects *PIGA*-CDG neurological models that recapitulate patient phenotypes, including seizures and neuromuscular defects, giving insight into mechanisms underlying prominent clinical phenotypes. This work not only advances the understanding of *PIGA*-CDG pathophysiology but also paves the way for developing innovative treatments for patients.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2005F: A genotype-first approach in over 170,000 participants from a large healthcare population identifies treatable Mendelian conditions

Authors:

K. Murphy[1], R. Torene[2], T. Brandt[1], M. Kelly[1], C. Martin[1], K. Retterer[1]; [1]Geisinger, Danville, PA, [2]Geisinger, North Bethesda, MD

Abstract:

Historically, clinical genomics has been driven by a phenotype-first approach. However, many Mendelian conditions (MC) are under-diagnosed due to atypical presentation and challenges in accessing expert care. The Geisinger MyCode Community Health Initiative is an unselected healthcare population with exome sequencing (ES) and linked electronic health records (EHR). MyCode currently returns ACMG Secondary Findings to participants and is uniquely positioned to pursue a genotype-first approach to diagnosis. Given the potential for improved health outcomes and utilization, we analyzed treatable MCs.

We compiled 484 MCs (across 469 genes) not currently screened by MyCode or on the PA newborn screening list where there is a treatment (e.g., medication, diet, surgery) and non-molecular confirmatory test for the condition. We categorized all ClinVar variants across MCs into tiers. Tier 1 variants represent the highest confidence for pathogenicity (P/LP with multiple submitters and no conflicting evidence). Tier 2 variants also have evidence of pathogenicity (P/LP with single submitter and variants with conflicting evidence but three times more P/LP assertions than benign).

Of 170,857 MyCode participants with ES, 6,445 had at least one significant finding: 3,831 (2.2%) participants had a Tier 1 genotype while 2,677 (1.6%) participants had a Tier 2 genotype. These are referred to as *suspected* treatable Mendelian condition (STMC) participants. Of the 469 MC genes evaluated, we observed P/LP genotypes in 129 (27.5%).

We queried EHRs to see if STMC participant diagnostic codes (ICD) matched the MC (by MONDO ID and word-embedding similarity). Out of the 6,445 STMC participants, 271 participants (4.2%)

had an ICD consistent with the suspected MC. Rates of ICD to MC matches in STMC participants were significantly higher than in controls (Fisher p=5.6e-25).
Targeted EHR review uncovered cases where clinical history was consistent with the MC and the participant did not have an ICD match, indicating that ICD alone is insufficient for evaluating STMCs and a large portion remain underdiagnosed. Molecular review of variants also underscored the complexity of associating P/LP variants with the correct condition at scale. Based on expert review, we extrapolate that 1,407 (695-2,118 95% CI, 0.82%) of participants are clinically affected with a treatable MC and 1,172 (509-1,835 95% CI) are under or misdiagnosed. In sum, nearly 1% of the MyCode population has a treatable MC not currently screened. Future efforts to improve variant curation and integrate a genomics-first approach into routine healthcare will enable earlier diagnosis and better clinical management.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2006F: A multi-omics approach improves the diagnosis and understanding of inborn errors of immunity

Authors:

D. Bonner[1,2], S. Marwaha[1,2], J. Carter[1,2], C. M. Reuter[1,2], H. Mendez[1,2], T. M. Maurer[1,2], R. A. Ungar[1,2], P. Goddard[1,2], S. Emami[1,2], B. M. Tucker[1,2], N. Gomez-Ospina[1,2], J. Hom[1,2], J. Sampson[1,2], J. Shon[3], S. B. Montgomery[1,2], J. A. Bernstein[1,2], M. T. Wheeler[1,2]; [1]Stanford Univ., Stanford, CA, [2]Stanford Ctr. for Undiagnosed Diseases, Stanford, CA, [3]Serimmune, Goleta, CA

Abstract:

Inborn errors of immunity (IEI) can present as increased susceptibility to infections, autoimmunity, autoinflammation, allergy, bone marrow failure, and/or malignancy. While our understanding of the genetic basis of IEIs has rapidly advanced in recent years, the interpretation of exome or genome sequencing (ES/GS) results remains a challenge due to the prevalence of variants and/or genes of uncertain significance (VUS/GUS). Resolving this uncertainty is necessary to provide timely diagnosis, reduce unnecessary testing, and guide medical management. At the Stanford Center for Undiagnosed Diseases (CUD) we developed a multi-omic approach that integrates clinical phenotyping, genomics (ES/GS), transcriptomics (RNA-seq), and immunomics (cytokine profiling, serum epitope repertoire analysis [SERA], and mass cytometry [CyTOF]) to improve diagnosis and elucidate disease mechanisms of IEIs. We evaluated 266 probands between 2016 to 2024, of which 36 (13%) have a predominant immune presentation and 84 (32%) have one or more immunologic features. Probands with an immune phenotype (120) underwent ES and/or GS (112), RNA-seq (87), cytokine profiling (37), CyTOF

(38) and SERA (15). ES/GS and multi-omics results were integrated and manually reviewed for diagnostic significance. Overall, 34% of probands (41/120) received a highly likely and/or candidate diagnosis including 44 molecular, 3 clinical, and 1 antibody-mediated phenocopy of a Mendelian condition. Five probands received more than one molecular diagnosis. We identified several novel IEI candidate genes/loci including *EIF2AK2*, *CFLAR*, *SATB1*, and an intergenic regulatory region at Xq28. In addition, we expanded the phenotype and mode of inheritance for known IEI genes *UNC93B1* and *TRAF3*, and we identified novel variants in known IEI genes *DNASE1L3*, *IL6ST*, *STXBP2*, *RAD51*, *CBL*, and *POLR3A*. Results from multi-omic and targeted functional assays resolved a variety of VUS including missense (11), splicing (3), and noncoding (3) variants. Structural variant analysis and RNA-seq prioritized a causal noncoding deletion at Xq28 not reported by clinical GS and microarray. RNA-seq also identified a possible mechanism for immunodeficiency caused by noncoding variants in *RNU4ATAC*. Here we present a comprehensive and integrated multi-omic approach to advance IEI diagnosis and discovery, reduce the diagnostic odyssey, and improve clinical care for patients impacted by IEI.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2007F: A Novel Homozygous Nonsense Variant in the LRRC56 Gene in a Neonate with Situs Inversus, Complex Heart Malformations, and Complete AV Block

Authors:

C. Neckelmann[1], B. Stone[2], J. Yurkoski[2], M. Mori[2], Z. Chen[1]; [1]Fulgent Genetics, El Monte, CA, [2]Nationwide Children's Hosp., Columbus, OH

Abstract:

Primary ciliary dyskinesias (PCDs) are a group of rare, genetically heterogeneous disorders affecting the structure and function of cilia. Multiple genes are responsible for cilia structure and function, and among them variants in *LRRC56* have been infrequently reported in patients with PCDs. We describe the 6th known patient with autosomal recessive *LRRC56*-related ciliopathy presenting in a male neonate as heterotaxy with severe complex heart malformations, complete heart block, small bowel malrotation, and polysplenia. Notably, complete AV block is a novel phenotypic feature that is not commonly observed in PCDs and has not been previously reported in any patients with *LRRC56*-related disease. Within this patient, genome sequencing identified a novel homozygous likely pathogenic variant in *LRRC56*, NM_198075.4.159C>A (p.Tyr53*), with no other variants found to explain the patient's clinical features. This is the first case within the literature of a patient with biallelic nonsense variants

that are consistent with complete loss of *LRRC56* function, further establishing the underlying gene-disease relationship and mechanism of disease. Furthermore, this case report underscores the necessity of considering *LRRC56* in the diagnostic evaluation of patients with suspected ciliopathy and heterotaxy disorders.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2008F: A novel *SCN1A* variant in a patient with neonatal-onset developmental and epileptic encephalopathy with arthrogryposis multiplex congenita

Authors:

Y. Kawai[1], K. Nagai[1], R. Tachioka[1], T. Nishikawa[1], Y. Enomoto[2], T. Naruto[2], Y. Kuroda[1], K. Kurosawa[1]; [1]Dept. of Med. Genetics, Kanagawa Children's Med. Ctr., Yokohama, Japan, [2]Clinical Res. Inst., Kanagawa Children's Med. Ctr., Yokohama, Japan

Abstract:

SCN1A encodes the alpha subunit of voltage-gated sodium channel NaV1.1. Loss-of-function variants (LoF) of *SCN1A* cause epilepsy such as Dravet syndrome and genetic epilepsy with febrile seizures plus. Sodium channel blockers should be avoided as they worsen epilepsy control. Recently, gain-of-function (GoF) *SCN1A* variants have been reported in several cases with severe neonatal-onset epilepsy and arthrogryposis multiplex congenita. Here, we reported a patient harboring a novel missense variant of *SCN1A* with neonatal-onset developmental and epileptic encephalopathy with arthrogryposis multiplex congenita.
The patient was a 5-year-old female born at 36 weeks' gestation with a maternal history of polyhydramnios. The birth weight was 2,532 g (+0.1 SD), birth length was 37.2 cm (-3.3 SD), and occipitofrontal circumference was 32.8 cm (+0.3 SD). Ventilation support had been required soon after birth, and tracheostomy was performed at the age of 3 months. She had frontal bossing, a low nasal bridge, a small mouth, micrognathia, a short neck, and rocker bottom feet. She had poor limb movement with multiple joint contractures and a fracture of the right humerus. She developed apneic seizures at the age of 8 days. Electroencephalography showed a suppression-burst pattern and multifocal spikes. Although anti-seizure medications reduced apneic seizures, intractable myoclonic seizures and focal motor seizures were then developed. Phenytoin, a sodium channel blocker, was partially effective for her seizures. She never controlled her head and did not achieve any developmental milestones until 5 years old. Whole exome sequencing identified a novel heterozygous variant in *SCN1A*, NM_001165963:c.2956C>G (p.Leu986Val). This variant was confirmed to be de novo by trio

Sanger sequencing. The variant was not found in the population databases, gnomAD and jMorp (Japanese Multi Omics Reference Panel), and was predicted to be gain-of-function by in silico analysis tool, funNCion. The CADD score of the variant was 25.3.
Gain-of-function *SCN1A* variants can cause severe neonatal onset developmental and epileptic encephalopathy with arthrogryposis multiplex congenita. Functional effect to sodium channel will be different from Dravet syndrome due to LoF variants, since sodium channel blocker is effective in these types of epilepsy associated with GoF variants and in this patient. Further functional analysis of *SCN1A* variants may lead to better pharmacological approach.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2009F: A pilot project moving the initiation of exome sequence panel testing for patients with neurodevelopmental disorders into the developmental and behavioral pediatric clinic identifies both actionable diagnostic findings and a high rate of variants of uncertain significance.

Authors:

M. Beltran Quintero[1], C. Conrad[1], Y. Wu[1], L. White[1], R. Adams[1], T. Froehlich[1], J. Ehrdhardt[1], A. Shillington[2]; [1]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [2]Cincinnati Children s Hosp., Cincinnatti, OH

Abstract:

Introduction: Genetic testing is vital for diagnosing and managing neurodevelopmental disabilities (NDDs). The American College of Medical Genetics and Genomics (ACMG) recommends exome sequencing for children with autism spectrum disorder (ASD), intellectual disability (ID), and/or global developmental delay (GDD). Typically, clinical geneticists or genetic counselors perform this testing. However, increasing demands on clinical genetic expertise for patients with NDDs have created delays in diagnosis. To address this, we initiated a pilot project to start exome testing in the Developmental Behavioral Pediatrics (DDBP) clinic. Our study aims to compare the diagnostic yield of exome testing (ASD/ID exome sequencing panel) with chromosomal microarray (CMA) when initiated by a DDBP clinician. **Methods:** A retrospective analysis of electronic health records was conducted to examine the diagnostic yield of the ASD/ID exome sequencing panel for patients at the Cincinnati Children's Hospital Medical Center (CCHMC) Division of Developmental and Behavioral Pediatrics (DDBP). Patients were included if they had been evaluated by a developmental clinician, diagnosed with ASD, ID, and/or GDD, and had genetic testing ordered by a DDBP provider between April 2021 and July

2023. A total of 633 patients were included. CMA was ordered in 593 cases, and a CCHMC ASD/ID exome sequencing panel was ordered in 44. Genetic results were categorized as normal, variants of unknown significance (VUS), or pathogenic. **Results:** The ASD/ID exome sequencing panel yielded normal results in 11% of cases, pathogenic variants in 7% of cases, and VUS in 82% of cases. CMA yielded normal results in 76% of cases, pathogenic variants in 8% of cases, and VUS in 16% of cases. **Conclusions:** The diagnostic yield of pathogenic findings was similar between microarray and exome-based studies at 7- 8%. However, there was a significant difference in the identification of VUS: CMA identified VUS in 16% of cases, while the ASD/ID exome sequencing panel identified VUS in 82% of cases. Initiating genetic testing in the DDBP clinic benefits patients with certain pathogenic or normal findings, but the high VUS rate in exome studies poses challenges for DDBP providers who are not trained in variant interpretation or results disclosure. Future work includes developing a pathway for interpretation and counseling of VUS findings, as these may be reclassified as pathogenic or clinically diagnostic with further functional or biochemical testing, or more detailed phenotype and literature review.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2010F: A quick diagnosis of Alagille syndrome 1 of a Romanian patient with cholestatic liver disease

Authors:

I. Sabau[1,2], A. Curici[1,2], M. Mitroi[1]; [1]Synevo Romania, Bucharest, Romania, [2]Carol Davila Univ. of Med. and Pharmacy, Bucharest, Romania

Abstract:

Background: This study presents a patient referred to our clinic by the attending gastroenterologist. The patient is 17 y.o. with cholestatic liver disease and facial specific dysmorphism, raising the suspicion of Alagille syndrome (MIM# 118450). Methods: Throughout the genetic consultation, we examined the patient's phenotypical traits and symptoms, meticulously constructed the family tree, and delved into the intricacies of his personal medical history. The assessment of the patient's symptoms guided the recommendation for targetted NGS testing to uncover a potential underlying genetic cause of the suspicion of Alagille syndrome. Results: NGS testing focused on two genes (JAG1 and NOTCH2) and returned positive for one heterozygous variant: a likely pathogenic variant, JAG1: NM_000214.3:c.97G>A p.(Gly33Ser). This variant is correlated with the patient's distinct phenotype, thus making possible the diagnosis of Alagille syndrome. Conclusion: Overall, this study emphasizes the

importance of genetic consultations and advanced genomic technologies in elucidating the genetic basis of complex disorders, offering crucial information for accurate diagnosis and personalised management of the patients. This case also highlights the utility of genetic testing, small targetted panel based on NGS technology when the clinical picture is very relevant. We want to highlight the importance of having a multidisciplinary approach when treating patients with a suspected rare disease.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2013F: Arimoclomol upregulates expression of genes belonging to the coordinated lysosomal expression and regulation (CLEAR) network

Authors:

H. Shammas[1], N. Havnsøe Torp Petersen[2], P. Klein[3], A. koustrup[4], M. Terndrup Pedersen[5], A. Bie[6], T. Mickle[1], C. Kolster Fog[7], T. Kirkegaard Jensen[8], S. Guenther[1]; [1]Zevra Therapeutics, Celebration, FL, [2]BioInnovation Inst., Copenhagen, Denmark, [3]Argobio Studio, Copenhagen, Denmark, [4]Novo Nordisk A/S, Copenhagen, Denmark, [5]Muna Therapeutics, Copenhagen, Denmark, [6]Orphazyme A/S, Copenhagen, Denmark, [7]Gubra, Copenhagen, Denmark, [8]Amplify Therapeutics ApS, Copenhagen, Denmark

Abstract:

Background: Niemann-Pick disease type C (NPC) is an ultra-rare, fatal neurodegenerative disease caused by mutations in the *NPC1* or *NPC2* gene leading to lysosomal dysfunction due to impaired intracellular transport of lipids. NPC is a heterogeneous disease with progressive symptoms and no FDA-approved therapies. Arimoclomol, an orally bioavailable molecule that crosses the blood-brain barrier, is being investigated as a potential treatment for NPC.
Methods: The effect of arimoclomol on the nuclear vs cytosolic distribution of the transcription factors EB (TFEB) and E3 (TFE3) was assessed in different cell types by immunofluorescence staining. Binding of TFE3/TFEB to the promoter regions of coordinated lysosomal expression and regulation (CLEAR) genes was assessed with chromatin immunoprecipitation and qPCR in wild-type (WT) fibroblasts treated with arimoclomol. The effects of arimoclomol on the expression levels of 7 selected CLEAR genes were examined by qPCR in WT and NPC1 fibroblasts. EndoH assays and Western blotting were conducted to assess NPC1 protein level and maturation. Filipin staining was used to evaluate cholesterol clearance from lysosomes following arimoclomol treatment.
Results: Arimoclomol increased activation of TFEB and TFE3 by promoting their translocation to

the nucleus of NPC1 fibroblasts. Binding of TFE3/TFEB to CLEAR motifs was enhanced by arimoclomol resulting in transcriptional upregulation of CLEAR genes including *NPC1*. Importantly, arimoclomol also raised the level of a more mature complex form of NPC1 mutant protein and decreased unesterified cholesterol in the lysosomes of NPC patient fibroblasts. **Discussion:** Arimoclomol activation of TFEB/TFE3 resulted in the upregulation of CLEAR gene expression including *NPC1*. CLEAR genes encode various proteins involved in lysosomal and autophagosomal functions, suggesting that arimoclomol may provide treatment effects through both NPC1-dependent and NPC1-independent processes. The observed downstream effects of arimoclomol treatment were increased biosynthesis of NPC1 resulting in higher levels of mature protein and ultimately in improved clearance of cholesterol from the lysosomes. Moreover, the upregulation of other lysosomal genes suggests that arimoclomol may generally improve lysosomal function and autophagy flux to promote cell health.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2015F: Biallelic null variants in *PNPLA8* cause microcephaly by reducing the number of basal radial glia

Authors:

Y. Nakamura[1], I. S. Shimada[2], R. Maroofian[3], M. Falabella[3], M. Zaki[4], M. Fujimoto[1], E. Sato[1], A. Miyauchi[5], E. Koshimizu[6], S. Miyatake[6], Y. Arioka[7], M. Honda[8], T. Higashi[9], F. Miya[10], Y. Okubo[11], PNPLA8 study group, K. Haginoya[11], T. Matsunaga[12], H. Osaka[5], N. Matsumoto[6], N. Ozaki[7], Y. Ohkawa[13], S. Oki[8], T. Tsunoda[14], R. Pitceathly[3], Y. Taketomi[9], H. Houlden[3], M. Murakami[9], Y. Kato[2], S. Saitoh[1]; [1]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Dept. of Cell Biology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [3]Dept. of Neuromuscular Diseases, UCL Queen Square Inst. of Neurology, Univ. Coll. London, London, United Kingdom, [4]Clinical Genetics Dept., Human Genetics and Genome Res. Inst., Natl. Res. Ctr., Cairo, Egypt, [5]Dept. of Pediatrics, Jichi Med. Univ., Tochigi, Japan, [6]Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [7]Pathophysiology of Mental Disorders, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [8]Dept. of Drug Discovery Med., Kyoto Univ. Graduate Sch. of Med., Kyoto, Japan, [9]Lab. of Microenvironmental and Metabolic Hlth.Sci., Ctr. for Disease Biology and Integrative Med., Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [10]Ctr. for Med. Genetics, Keio Univ. Sch. of Med., Tokyo, Japan, [11]Dept. of Pediatric Neurology, Miyagi Children's Hosp., Sendai, Japan, [12]Dept. of Clinical Pharmacy, Graduate Sch. of Pharmaceutical Sci., Nagoya, Japan, [13]Div. of Transcriptomics, Med. Inst. of Bioregulation, Kyushu Univ., Fukuoka, Japan, [14]Lab. for Med. Sci. Mathematics, Dept. of Biological Sci., Sch. of Sci., The Univ. of Tokyo, Tokyo, Japan

Abstract:

PNPLA8, one of the calcium-independent phospholipase A_2 enzymes, is involved in various physiological processes through the maintenance of membrane phospholipids. Biallelic variants in *PNPLA8* have been associated with a range of pediatric neurodegenerative disorders. However, phenotypic spectrum, genotype-phenotype correlations, and the underlying mechanisms are poorly understood.Here, we newly identified 14 individuals from 12 unrelated families with biallelic ultra-rare variants in *PNPLA8* presenting with a wide phenotypic spectrum of clinical features. Analyzing the clinical features of current and previously reported individuals (25 affected individuals across 20 families) showed that *PNPLA8*-related neurological diseases manifested as a continuum ranging from variable developmental and/or degenerative epileptic-dyskinetic encephalopathy to childhood-onset neurodegeneration. We found that complete loss of PNPLA8 was associated with more profound end of the spectrum with congenital microcephaly.Using cerebral organoids generated from human induced pluripotent stem cells, we found that loss of PNPLA8 led to developmental defects by reducing the number of basal radial glial cells (bRGCs) and upper-layer neurons. Spatial transcriptomics revealed that loss of PNPLA8 altered the fate specification of apical radial glial cells, as reflected by the enrichment of gene sets related to cell cycle, bRGC, and neural differentiation. Neural progenitor cells lacking PNPLA8 showed a reduced amount of lysophosphatidic acid (LPA), lysophosphatidylethanolamine, and phosphatidic acid. The reduced number of bRGCs in patient-derived cerebral organoids was partially rescued by the addition of LPA.Our data suggest that PNPLA8 is critical to meet phospholipid synthetic needs and to produce abundant bRGCs in human brain development.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2017F: Biallelic variants in consanguineous families causing neurodevelopmental disorders

Authors:

T. Froukh, A. Nazer; Philadelphia Univ., Amman, Jordan, Jordan

Abstract:

Since the introduction of whole exome sequencing (WES) and new genetic variants linked to neurodevelopmental disorders (NDD) are continuously identified, published and catalogued to be used in routine genetic diagnostics. In our continuous research, WES was used for ten unrelated consanguineous families from Jordan characterized by one or two affected members with NDD. Six homozygous variants were identified by cross-link to the pathogenic variants in

ClinVar and in our in-house database. Three of them were previously reported in Jordan in unrelated families compared to the families in this study, the variants are: NM_018359.5:c.344T>A, p.Val115Glu in the gene *UFSP2*, NM_000487.6:c.256C>G, p.Arg86Gly in the gene *ARSA* and NM_017882.3:c.144G>A, p.Trp48* in the gene *CLN6*. The new bi-allelic variants are missense NM_001457.4:c.609G>T, p.Gln203His and NM_152722.5:c.503G>A, p.Cys168Tyr in the genes *FLNB* and *HEPACAM* respectively, frameshift NM_022725.4:c.109dup, p.Trp37Leufs*54 in the gene *FANCF* and one in the splice acceptor NM_000191.3:c.498-1G>A in the gene *HMGCL*. The results of this study stress the importance of continuous research using WES for NDD in Jordan to identify new variants.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2018F: Case report of PPOX mutation within intron-exon boundary junction inducing aberrant splicing in a patient with porphyria variegata

Authors:

A. Chen[1], E. C. Chen[2], D. Chen[3]; [1]Troy High Sch., Fullerton, CA, [2]Chen and Lee Med. Clinic, Montebello, CA, [3]Arbelos Genomics, Diamond bar, CA

Abstract:

Porphyria Variegata (OMIM#176200) is a rare metabolic disorder characterized by deficient function of protoporphyrinogen oxidase (PPOX gene product). The inheritance pattern is autosomal dominant with variable expressivity. Protoporphyrinogen oxidase is involved in the biosynthesis of heme by catalyzing the oxidation of protopoprphrinogen IX to form protoporphyrin IX. Deficient production of PPOX gene product results in the accumulation of protporphrinogene IX, which is responsible for the clinical manifestation. Clinical presentation includes skin photosensitivity, blistering, and hyperpigmentation. In addition, acute attack includes abdominal pain, and neurological signs are also presenting signs.
Here, we present a familial cohort of porphyria variegata. The proband is an East Asian female age 43, with 15 years of unexplained blisters and hyperpigmentation of her hands and sun-exposed facial area. Acute attacks of abdominal pain and vomiting of unknown etiology are also present in irregular frequency. Lab indicates elevated uroporphyrin I and III. Due to the suspicion of porphyria cutanea tarda (PCT), an NGS based genetic test containing 10 genes (ALAD, ALAS2, CLPX, CPOX, FECH, GATA1, HMBS, PPOX, UROD, UROS) involved in the PCT are ordered. The NGS identified a variant within PPOX gene (c.807+2T>C (splice donor)). We have assign this mutation a variant of likely pathogenicity (VLP) because of the location of the SNV

within splice junction and additional computational analysis.
Due to the paucity of additional functional evidence, we set out to conduct RNA sequencing study to ascertain the effect of the PPOX c.807+2T>C mutation causing potential splice variants. First, we conducted expanded genetic analysis of 3 related individuals (children and mother of the proband) by Sanger sequencing. The related individuals are found to carry the mutations. Second, the mRNA from buffy coats of proband and negative related indiviuals is evaluated by RNA targeted NGS sequencing using a bait capture panel focusing on the porphyria associated genes. The RNA sequencing results is analyzed using a custom designed bioinformatic pipelines. Our results indicated that proband's mRNA results show evidence of skipping of exon 8 in the mRNA of PPOX genes. Such abberant splice pattern is not identified in the negative samples. The result suggest the potential of PPOX c.807+2T>C mutation playing a role in the deficiency of PPOX gene function. Together, this bolster the possibility of the mutation's involvement in proband's clinical manifestation. Expanded study are currently underway.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2019F: Characterization of hearing loss and skin cancer in individuals with Rothmund-Thomson syndrome

Authors:

A. Hansen[1], S. Chinapuvvula[2], T-T. Rideau[3], K. Ward[1], C. Hansen[1], L. L. Wang[4]; [1]Geneial, Houston, TX, [2]The Univ. of Texas at Austin, Austin, TX, [3]Baylor Coll. of Med., Houston, TX, [4]Baylor Coll. Med., Houston, TX

Abstract:

Rothmund-Thomson syndrome (RTS) is a rare, autosomal recessive multisystem disorder characterized by poikiloderma, small stature, sparse hair, abnormal bones, cataracts, and cancer predisposition, particularly for osteosarcoma. RTS is classified into two groups: Type 2, caused by *RECQL4* pathogenic variants and Type 1, associated thus far with pathogenic variants in *ANAPC1*, *CRIPT* and *DNA2*. Genotype-phenotype correlations have been established for some findings such as osteosarcoma and juvenile cataracts. However, many other clinical features have not been fully characterized given the rarity of the disorder and challenges associated with collecting adequate clinical data worldwide.
The goal of this study is to characterize two phenotypes that have not been well studied in RTS: skin cancer and hearing loss. All individuals in the RTS Registry provided informed consent to enroll in a Baylor College of Medicine IRB approved protocol which has enrolled 155 RTS probands from 1999 to 2023. Skin cancers have been reported in RTS, but the actual prevalence

and other clinical features are unknown, and there are no published reports on hearing loss. Surveys were sent to 93 of 155 enrolled RTS probands for whom contact information was available. Survey response rate was 40%.
12 out of 47 subjects had skin cancer resulting in a prevalence of 25.5% (age range 18-77 years). Five (42%) were squamous cell carcinomas, 2 (17%) were basal cell carcinomas (16.67%), 1 (8%) was melanoma, 2 (17%) were combined basal and squamous cell carcinomas, and 2 (17%) were unknown type. Six (50%) had Type 1 RTS, 4 (33%) had Type 2 RTS, and 2 (17%) had unknown mutation status.
For hearing loss, 30 out of 56 subjects with available data had hearing loss yielding a prevalence of 53.6% (age range 1 month-59 years). Three (10%) were conductive, 5 (17%) were sensorineural, and 22 (73%) were of unknown type. Ten (33%) were bilateral, 3 (10%) were unilateral (3%) and 17 (57%) were unknown location. Four subjects (13%) had Type 1 RTS, 21 (70%) had Type 2 RTS, and 5 (17%) had unknown mutation status. Hearing loss tended to occur early; therefore, prompt diagnosis may improve childhood development and quality of life.
In order to accelerate and improve our data collection for these and other phenotypic features, we are partnering with Geneial to utilize its universal registry platform which will allow for higher quality longitudinal data collection. Ultimately, we aim to increase longitudinal research engagement while enhancing individual privacy protections and data availability for clinicians and investigators, expediting research and patient care for RTS as a model for other rare Mendelian conditions.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2020F: Characterization of pathogenic variant disease-phenotypes throughout the human genome

Authors:

D. Stein[1], M. Kars[1], B. Milisavljevic[2], M. Mort[3], P. Stenson, D[3], P. Zhang[2], B. Boisson[2], D. N. Cooper[4], A. Schlessinger[1], Y. Itan[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Rockefeller Univ., New York, NY, [3]Cardiff Univ., Cardiff, United Kingdom, [4]Cardiff Univ, Cardiff, Wales, United Kingdom

Abstract:

Despite advancements in methodology, key limitations impede computational tools for variant assessment. Firstly, many methods are designed exclusively or primarily for missense variants while some others attempt to interpret coding SNVs and/or indels in general or offer interpretations for non-coding variants. Secondly, many methods consider pathogenic variants

as a homogeneous group, and may systematically under-perform for some subpopulations of variants. Recent efforts have been made to investigate more granular characteristics of pathogenic variants, such as their mechanism of effect or mode of inheritance. Yet, many qualities of pathogenic variants, such as their phenotypic effects, remain under-explored. Currently, variant effect predictors most often provide estimations only of pathogenicity/deleteriousness generally, while most efforts to predict relationships between pathogenic genotypes and phenotypes have been carried out at the gene- or protein-level. Efforts to design pathogenicity prediction methods for specific diseases/phenotypes have reported enhanced performance compared to methods predicting pathogenicity generally. However, these methods, by definition, are limited to a single/small number of disease(s)/phenotype(s). As a result, variant-type agnostic, genome-wide pathogenicity predictors that elucidate the unexplored landscape of pathogenic variant characteristics and expand on the success of disease/phenotype specific pathogenicity prediction approaches are needed.
Here, we present Variant-to-Phenotype (V2P), a multi-task, multi-output machine learning model to jointly predict pathogenicity and phenotypic effect for SNVs and indels throughout the genome that exploits relationships between pathogenic variants and disease-phenotypes to improve the quality of phenotype and pathogenicity predictions globally. With V2P, we generate, for the first time, a complete mapping of all human SNVs and a large number of indels to disease-phenotypes. We thoroughly assess V2P's performance on several high-quality datasets. We compare V2P to variant effect prediction methods, illustrating the benefit of considering variant phenotypic outcomes during training. Additionally, we demonstrate V2P's utility for the identification of causal disease variants in real and simulated patient sequencing data. Finally, we identify the features of pathogenic variants underlying each human disease group. We precomputed V2P scores for every SNV in the human reference genome, along with indels from gnomAD to provide a freely accessible atlas of variant effect and disease-phenotypes for the community.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2021F: Clinical and Molecular Genetic Studies of a Filipino Family with a Peculiar Muscle Disease

Authors:

K. Porto[1], H. Ishiura[1], J. Mitsui[1], S. Zamora[2], M. Zamora[2], R. Mendoza[2], T. Toda[1], S. Tsuji[3]; [1]The Univ. of Tokyo, Tokyo, Japan, [2]Premiere Med. Ctr., Cabanatuan, Philippines, [3]Graduate Sch Med, Univ Tokyo, Tokyo, Japan

Abstract:

We conducted a comprehensive genetic study of a Filipino family with an unusual myopathy. The proband, a 48-year-old Filipino male from a non-consanguineous marriage, exhibited gait difficulties at 17, described as “walking like a duck.” Over the years, his ambulation worsened, also involving proximal upper extremity weakness and hoarseness by age 46. By 47, he required continuous BiPAP support, but non-compliance led to ICU admissions due to carbon dioxide retention. Cardiac abnormalities included inferolateral wall ischemia and hypokinesia of the inferior interventricular septum detected on ECG, though connection to the myopathy was unclear. Neurological examination showed proximal upper extremity weakness, scapular winging, and absent gag reflex, while sparing facial and extraocular muscles. CK levels remained normal. Further investigation revealed an unusual pattern of muscle weakness running through four generations of the family. Eleven members were identified as affected, including five alive at the time of examination. The mode of inheritance appeared autosomal dominant, though X-linked inheritance couldn’t be excluded. The weakness was proximal, with severity ranging from mild scapular winging to severe respiratory muscle involvement, affecting only the proband, who also had possible cardiac involvement. The mean age of onset was 36 years old (18-74 years old). The proband died at 49 from severe respiratory failure before a muscle biopsy could be performed. Targeted gene testing excluded LGMD 1B, Emery-Dreifuss syndrome, FHL1-related myopathy, and Pompe disease while copy number variation analysis using aCGH did not reveal copy number abnormalities. Southern blot hybridization analysis showed no pathogenic D4Z4 contracted repeats. Whole exome sequence analysis did not show pathogenic variants in genes for myopaties with autosomal dominant-inheritance or those for X-linked. Linkage analysis was conducted recruiting 5 affected and 5 unaffected family members employing Affymetrix Genome-Wide SNP Array 6.0. Parametric linkage analysis assuming an autosomal dominant inheritance revealed several genome regions with LOD scores greater than 1.0. Whole genome sequence analysis of the proband revealed candidate variants co-segregating with the disease in the genes with abundant expressions in muscle tissues, which are located in candidate regions of chr7:150,013,343-152,668,058, chr7:70,511,436-78,137,469, and chr1:206,618,625-208,657,809 (GRCh37). Future directions involve studying independent families with similar clinical presentations and collaborating with global myopathy centers to expand our investigation.

---PAGEBREAK---

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2022F: Clinical features of HNRNP-related disorders

Authors:

N. Okamoto[1], E. Nishi[1], Y. Hasegawa[1], K. Yanagi[2], T. Kaname[3]; [1]Osaka Women s and Children s Hosp., Izumi, Japan, [2]Natl Ctr Child Hlth Dev (NCCHD),, Tokyo, Japan, [3]Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan

Abstract:

TThe HNRNP gene group comprises more than 20 members of the heterogeneous nuclear ribonucleoprotein (hnRNPs) family. HNRNP polypeptides bind to heterogeneous nuclear RNAs, the transcripts produced by RNA polymerase II. Members of the HNRNP family localize to the nucleus and shuttle pre-mRNA transcripts between the nucleus and cytoplasm for processing and transport. Recent advances in genetic analyses revealed that HNRNP-related disorders are associated with various neurogenetic syndromes and congenital anomaly syndromes. Neurodevelopmental disorder with craniofacial dysmorphism and skeletal defects is caused by a heterozygous mutation in the HNRNPH1 gene. The brain type of X-linked syndromic intellectual developmental disorder is caused by pathogenic variants in the HNRNPH2 gene. Au-Kline-Okamoto syndrome (AUKS) with HNRNPK variants is characterized by hypotonia, global developmental delay, characteristic facial features, and internal organ anomalies. Okamoto syndrome has features that overlap with Au-Kline syndrome and is also caused by HNRNPK variants. Variants in HNRNPU have been reported in patients with epileptic encephalopathy. The clinical manifestation of 1q43-44 deletion syndrome is associated with haploinsufficiency for several genes, including HNRNPU. We encountered 11 patients with HNRNP-related disorders, including 1q43-44 deletion syndrome, in one institution. Many of the variants were revealed by whole exome sequencing or whole genome sequencing in the Initiative on Rare and Undiagnosed Diseases Study. We herein report the clinical features of HNRNP-related disorders. Further investigations are needed to elucidate the natural history of HNRNP-related disorders.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2023F: Clinical, genetic and bioinformatic analysis of Saudi families with Joubert Syndrome and Related Disorders

Authors:

D. Aljeaid[1,2], A. Riman[1], N. Issa[1], N. Shaik[1]; [1]King Abdulaziz Univ., Jeddah, Saudi Arabia, [2]King Abdulaziz Univ. Hosp., Jeddah, Saudi Arabia

Abstract:

Background: Joubert syndrome and related disorders (JSRD) are clinically and genetically heterogeneous ciliopathies caused by pathogenic variants in over 40 genes, mainly encoding ciliary proteins. However, data on JSRD in the Saudi population is limited.
Objective:This study aimed to identify novel and known JSRD-causing variants in Saudi families and analyze their functional impact on protein structure using advanced computational tools.
Methods:Following ethical protocols, participants with confirmed/suspected JSRD based on genetic testing or clinical features (molar tooth sign on MRI, developmental delays/intellectual disability, hypotonia) were recruited. Comprehensive clinical data were obtained. Exome sequencing were performed on DNA samples, followed by Sanger validation of candidate variants. Bioinformatics tools like VEP, HSF, and SpliceAI assessed variant pathogenicity. Conservation analysis and protein expression data were obtained from Ensembl and Human Protein Atlas databases. RNAfold and ViennaRNA web services were used to predict variant impact on RNA secondary structure and stability.
Results:Three distinct homozygous variants causing JSRD were identified in three consanguineous Saudi families. A novel stop-gain variant in *KIF7* (c.2992C>T; p.(Gln998Ter)), a novel and a rare splice site variants in *CEP104* (c.489+1G>A) and *TMEM237* (c.869+1G>A), respectively. All were predicted to be pathogenic by in silico analysis. These variants resided in highly conserved regions across primates, suggesting potential functional impact on protein structure. RNA expression analysis revealed tissue-specific patterns, with *KIF7* enriched in female reproductive organs, *CEP104* in the nervous system, and *TMEM237* in the retina and ciliated cells. Secondary structure analysis suggested potential alterations in RNA folding due to the variants.
Conclusion:This study identified two novel and one rare pathogenic variants in *KIF7*, *CEP104*, and *TMEM237* genes, respectively, in JSRD patients from Saudi families. Computational analysis suggests these variants may disrupt RNA folding and protein function. These findings expand the known variant spectrum for JSRD, particularly in the Saudi population, and warrant further functional studies to validate their pathogenicity and elucidate their role in JSRD pathogenesis.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2024F: Complete nanopore repeat sequencing of SCA27B (GAA-FGF14 ataxia) in Japanese

Authors:

N. Matsumoto[1,2,3], H. Doi[4], H. Yaguchi[5], E. Koshimizu[1], I. Yabe[5], S. Miyatake[1,2]; [1]Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [2]Dept. of Clinical Genetics, Yokohama City Univ. Hosp., Yokohama, Japan, [3]Dept. of Rare Disease Genomics,

Yokohama City Univ. Hosp., Yokohama, Japan, [4]Yokohama City Univ Grad Sch Med, Yokohama, Japan, [5]Dept. of Neurology, Faculty of Med. and Graduate Sch. of Med., Hokkaido Univ., Sapporo, Japan

Abstract:

Background Although pure GAA expansion is considered pathogenic in SCA27B, non-GAA repeat motif is mostly mixed into longer repeat sequences. This study aimed to unravel the complete sequencing of *FGF14* repeat expansion to elucidate its repeat motifs and pathogenicity. **Methods** We screened *FGF14* repeat expansion in a Japanese cohort of 460 molecularly undiagnosed adult-onset cerebellar ataxia patients and 1022 controls, together with 92 non-Japanese controls, and performed nanopore sequencing of *FGF14* repeat expansion. **Results** In the Japanese population, the GCA motif was predominantly observed as the non-GAA motif, whereas the GGA motif was frequently detected in non-Japanese controls. The 5'-common flanking variant was observed in all Japanese unexpanded GAA repeat alleles, demonstrating its meiotic stability against repeat expansion. In both patients and controls, pure GAA repeat expanded up to 400 units, whereas non-pathogenic GAA-GCA repeat expanded more, up to 900 units, but they evolved from different haplotypes, as rs534066520, located just upstream of the repeat sequence, completely discriminated them. Both $(GAA)_{\geq 250}$ and $(GAA)_{\geq 200}$ were enriched in patients, whereas $(GAA\text{-}GCA)_{\geq 200}$ was similarly observed in patients and controls, suggesting the pathogenic threshold of $(GAA)_{\geq 200}$ for cerebellar ataxia. We identified 14 patients with SCA27B (3.0%), but their SNP genotype indicated different founder alleles between Japanese and Caucasians. The low prevalence of SCA27B in Japanese may be due to the lower allele frequency of $(GAA)_{\geq 250}$ in the Japanese population than in Caucasians (0.15% vs. 0.32%-1.26%). **Conclusions** *FGF14* repeat expansion has unique features of pathogenicity and allelic origin, as revealed by a single ethnic study. **Acknowledgments** Patients and families, and Drs. Kihara N, Matsubara T, Mori Y, Kunieda K, Shimizu Y, Toyota T, Shirai S, Matsushima M, Okubo M, Wada T, Kunii M, Johkura K, Miyamoto R, Osaki Y, Miyama T, Satoh M, Morino H, Izumi Y, Shimohata T, Yoshida K, Adachi H, and Tanaka F are highly appreciated for their contribution to this study.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2026F: Comprehensive Observational Study on the Genetic Basis of CADASIL: Insights into a Complex Neurological Disorder

Authors:

A. Key[1], **N. M. Bruehl**[1], N. C. Minix[2], J. Bockholt[3], M. D. Geschwind[4], J. S. Paulsen[1]; [1]Univ. of Wisconsin, Madison, Madison, WI, [2]Univ. of Wisconsin-Madison, Madison, WI, [3]Georgia State Univ., Atlanta, GA, [4]Univ. of California, San Francisco, San Francisco, CA

Abstract:

CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy) is a genetic disorder caused by mutations in the NOTCH3 gene, leading to recurrent strokes, cognitive decline, and psychiatric disturbances. This ongoing North American natural history study aims to better characterize CADASIL by following a cohort of affected individuals and family controls at three timepoints over five years. Current enrollment exceeds 300 participants and will continue until a total of 500 is reached. The primary objective is to monitor the progression of CADASIL. Secondary objectives include lifestyle factors, neurological and clinical outcomes, neuropsychological and behavioral assessments, and identifying biomarkers for disease progression. The study involves 400 NOTCH3-positive individuals and 100 family controls. Each participant will be assessed at baseline, 12-18 months, and 24-36 months. Each timepoint includes a neurological evaluation, neuropsychological and behavioral assessments, as well as an environmental and lifestyle interview. Controls must be related to a genetically positive individual within specified generations. Whole exome sequencing will be performed at the conclusion of baseline enrollment. All participants must have a pre-existing genetic report or be tested by the study- the unknown group includes those undergoing testing. MRI scans will be conducted at each visit to monitor structural and functional brain changes. Wet biomarkers will be collected to identify indicators of disease progression. Participants were divided into four groups: positive, unknown, VUS (Variant of Uncertain Significance), and negative. The positive group, with 136 participants, had a mean age (MA) of 51.56 years, predominantly white (126), 83 females (F) and 53 males (M). Unknown included 77 participants, MA=46.87 years, mostly white (70), 47 F and 30 M, who are currently in the process of being tested. The VUS group comprised 6 participants, all white, MA=57.15 years, and mostly F (5). Lastly, the negatives had 19 participants, MA=52.08 years, predominantly white (17), with 12 F and 7 M. The majority of participants were non-Hispanic. Future work will involve comparison of mutation locations and correlation between mutation sites and their lethality. Initial observations suggest that more lethal variants occur on codons 1-6, while less lethal variants are found on codons 7-32. This aims to enhance our understanding of the genetics of CADASIL and its impact on disease progression. The integration of basic science and cognitive data will help identify key factors of CADASIL and inform future therapeutics.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2027F: De novo variants in GTF2H1 underlie variable syndromic developmental delay

Authors:

K. Patterson[1], M. Kircher[2], K. Buckingham[1], R. Wallerstein[3], E. Gabor-Fourcade[4], M. Hannibal[5], S. Casadei[1], O. Sommers[1], C. T. Marvin[6], M. M. Morrow[7], University of Washington Center for Rare Disease Research, L. Starita[6], C-L. Wei[1], J. X. Chong[1], M. J. Bamshad[1]; [1]Univ. of Washington, Seattle, WA, [2]Berlin Inst. of Hlth.@ Charité, Berlin, Germany, [3]UCSF, San Francisco, CA, [4]Univ. of California, San Francisco, San Francisco, CA, [5]Univ. of Michigan Med. Sch., Ann Arbor, MI, [6]Univ of Washington, Seattle, WA, [7]GeneDx, Gaithersburg, MD

Abstract:

In 2010, we identified heterozygous variants in KMT2D as the cause of Kabuki syndrome using exome sequencing to study a cohort of ten families. Only one of the ten remained unsolved, with no variants identified in KMT2D despite manual review and Sanger sequencing of all 54 coding exons. Two years later, reanalysis of that individual's exome identified a heterozygous de novo nonsense variant (c.1417C>T, p.Arg473*) in a little-studied gene, GTF2H1. We conducted an exhaustive search for other cases, reaching out to Kabuki syndrome support groups worldwide and to investigators studying neurodevelopmental delay but did not identify a second. We entered this candidate gene into Matchmaker Exchange, but did not match with other cases with a similar phenotype and the same candidate gene until 2023, when a second case with a de novo 2-bp deletion (c.891_892del, p.N298*) was identified. Both variants are predicted to result in loss of function, are exceedingly rare in the gnomAD v4.0 and RGC-ME population databases (i.e., among >1.6 million individuals, six heterozygotes for p.Arg473* and zero for p.N298* were present), and are predicted to be pathogenic (CADD v1.7 = 39.0 and 34.0 respectively). GTF2H1 is highly constrained (top ~12%ile) for loss of function variation, with a gnomAD v2 LOEUF (o/e) of 0.127 (pLI = 0.990), shet of 0.124, and pHaplo of 0.98, scores consistent with a gene that is haploinsufficient and intolerant of loss of function variation. Both individuals exhibit mild to moderate cognitive and motor delay and myopia. Other clinical findings present in the first individual include cardiac defects, hypotonia, long palpebral fissures, arched eyebrows, cleft palate, sacral dimple, thin nails, prominent fingertip pads, short stature/growth deficiency, left diaphragmatic hernia, and eventration of the right diaphragm. No cardiac abnormalities were identified upon physical exam of the second individual, but behavioral abnormalities and an autism diagnosis were reported. GTF2H1, also known as general transcription factor IIH subunit 1 or p62, is a subunit of the general transcription and DNA repair factor IIH (TFIIH) core complex. TFIIH plays a role in nucleotide excision repair of damaged DNA and in regulating RNA transcription by RNA polymerase II, which transcribes nearly all nuclear-encoded mRNAs. Because so few cases with GTF2H1 variants have been

identified, we are pursuing saturation genome editing experiments to assess the effects of the variants in our cases and of all possible single nucleotide changes and in-frame deletions in all 14 coding exons of the gene.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2028F: Drosophila Models Uncover Substrate Channeling Effects on Phospholipids and Sphingolipids in Peroxisomal Biogenesis Disorder

Authors:

M. Wangler[1], Y-H. Chao[1], M. Roth[2], R. Welti[3], J. McNew[4]; [1]Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Kansas State Univ., Manhattan, KS, [3]Kansas State Univ., Manhatten, KS, [4]Rice Univ., Houston, TX

Abstract:

Peroxisomal Biogenesis Disorders Zellweger Spectrum (PBD-ZSD) disorders are a group of autosomal recessive defects in peroxisome formation that produce a multi-systemic disease presenting at birth or in childhood. Well documented clinical biomarkers such as elevated very long chain fatty acids (VLCFA) are key biochemical diagnostic findings in these conditions. Additional, secondary biochemical alterations such as elevated very long chain lysophosphatidylcholines are allowing newborn screening for peroxisomal disease. In addition, a more widespread impact on metabolism and lipids is increasingly being documented by metabolomic and lipidomic studies. We have developed animal models of PBD using *Drosophila pex2* and *pex16* and we compare results from these models to human subjects with *PEX1* mutations. Here we identify phospholipid abnormalities in *Drosophila* larvae and fly brain characterized by differences in the quantities of phosphatidylcholine (PC) and phosphatidylethanolamines (PE) with long chain lengths and reduced levels of intermediate chain lengths. For diacylglycerols (DAG), the precursor of PE and PC through the Kennedy pathway, the intermediate chain lengths are paradoxically increased suggesting an imbalance between DAGs and PE and PC that suggests the two acyl chain pools are not in equilibrium. Altered acyl chain lengths are also observed in PE ceramides in the fly models. Interestingly, plasma from human subjects exhibit phospholipid alterations similar to the fly model. Moreover, human plasma shows reduced levels of sphingomyelin with 18 and 22 carbon lengths but normal levels of C24. Our results suggest that peroxisomal biogenesis defects alter shuttling of the acyl chains of multiple phospholipid and ceramide lipid classes, whereas DAG species with intermediate fatty acids are more abundant. We propose a model in which these genetic

disorders are characterized by an imbalance between de novo synthesis of PC and PE through the Kennedy pathway and remodeling of existing PC and PE through the Lands cycle. This imbalance is likely due to overabundance of very long and long acyl chains in PBD and a subsequent imbalance due to substrate channeling effects. Given the fundamental role of phospholipid and sphingolipids in nervous system functions, these observations suggest widespread cell membrane lipid abnormalities are part of the pathogenesis of PBD-ZSD. In summary, *Drosophila* models can uncover disease-relevant biochemical derangements in peroxisomal disease.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2029F: Elucidating the functional relationship between ATAXIN1 and ATAXIN1-Like in the brain

Authors:

E. Villavicencio Gonzalez[1], M. A. Durham[1], A. K. Engstrom[1], J-P. Revelli[1], H. Zoghbi[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. of Med./HHMI/NRI, Houston, TX

Abstract:

Spinocerebellar ataxia type 1 (SCA1) is a devastating neurodegenerative disorder caused by an abnormal CAG expansion that encodes the polyglutamine tract in ATXN1. It is characterized by motor coordination deficits due to loss of Purkinje cells, difficulty breathing and swallowing, and eventually, premature death due to loss of brain stem neurons. ATXN1 forms a native complex with the transcriptional repressor Capicua (CIC), and this interaction has a key role in SCA1 pathogenesis. Interestingly, increasing the levels of a paralog of ATXN1, known as ATXN1-Like (ATXN1L), suppresses the cerebellar pathology in a mouse model of SCA1. ATXN1L is an evolutionarily conserved protein that shares high sequence identity with ATXN1 in its functional domains and also interacts with CIC. However, despite the similarities between these two paralogs, their loss-of-function phenotypes are vastly different. Homozygous loss of *Atxn1L* in mice results in hydrocephalus, lung defects, and abdominal wall abnormalities, which cause early lethality. On the other hand, homozygous loss of *Atxn1* only results in a learning and memory deficit. This suggests that ATXN1L is more critical for viability than ATXN1; however, the mechanisms underlying the functional differences between both paralogs have not yet been elucidated. We hypothesize that ATXN1 and ATXN1L have close functional relationships, but they also have unique roles regulating different transcriptional targets of their respective CIC complexes. To identify the transcriptional targets of the CIC-ATXN1 and CIC-ATXN1L complexes, we performed RNA-seq in the forebrain of *Cic*$^{-/-}$, *Atxn1*$^{-/-}$, *Atxn1L*$^{-/-}$, and wildtype mice at

embryonic day 18.5 (E18.5). The differentially expressed genes (DEGs) in *Cic*$^{-/-}$ forebrain showed high directional concordance in the expression profiles of *Atxn1*$^{-/-}$ and *Atxn1L*$^{-/-}$mice, suggesting that many of the CIC targets are regulated by both paralogs. We also observed a trend in the effect sizes for most of the canonical CIC targets: largest in *Cic*$^{-/-}$, followed by *Atxn1L*$^{-/-}$, and smallest in *Atxn1*$^{-/-}$. Our results establish that, while there is a functional overlap between both complexes, the CIC-ATXN1L complex has a more dominant role in the regulation of CIC targets during early stages of brain development. Additionally, we performed Cleavage Under Targets and Release Using Nuclease (CUT&RUN) to identify the direct targets of these complexes. Preliminary data show that both ATXN1 and ATXN1L are binding most of the CIC canonical targets, which is consistent with the RNA-seq data. Future work will focus on exploring whether the functional overlap of these complexes is maintained in the adult mouse brain.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2030F: Establishing prevalence and phenotypic spectrum of purine and pyrimidine disorders using genotype-first approach to mining electronic health record biobanks

Authors:

K. Chaillet, S. Haines, J. Denny, H. Mo, O. Shchelochkov; NIH/NHGRI, Bethesda, MD

Abstract:

Introduction: Large electronic health record (EHR) biobanks present an opportunity for studying rare disorders. We propose to implement a genotype-first strategy to prospectively investigate disorders of purine and pyrimidine metabolism (DPPMs) in EHR biobanks, starting with the *All of Us* Research Program. This work aims to complement a DPPM natural history protocol (NCT06092346) by expanding our understanding of the prevalence and phenotypic spectrum of these disorders. To start, we applied a genotype-first approach to DPPM genes with three different inheritance patterns: *HPRT1* (X-linked recessive), *IMPDH1* (autosomal dominant), and *ADSL* (autosomal recessive). We hypothesize that the prevalence of DPPMs is higher and the phenotypic spectrum more variable than has previously been reported. **Methods:** For each gene, we used a variant filtering procedure that incorporated published case reports, ClinVar classifications of pathogenic/likely pathogenic, and *in silico* predictors (SpliceAI and REVEL) to assemble a pool of variants present in the *All of Us* database (n = 245,394) that are known or predicted to be pathogenic. We computed estimated carrier frequency and prevalence and reviewed the EHR data of participants. **Results:** For *HPRT1*, we identified two hemizygotes with phenotypes characteristic of Lesch-Nyhan disease (e.g., recurrent gout, psychiatric disorders,

chronic kidney disease, open bites of arms/hands) without recorded EHR diagnoses (estimated prevalence in *All of Us* = ~ 1:120,000 vs. published prevalence = 1:380,000 (PMID: 20301328)). We did not observe any individuals in *All of Us* with known pathogenic missense variants in *IMPDH1*. Finally, we found an estimated carrier frequency for *ADSL* (1:438) higher than what has been reported previously (1:557 (PMID:28127531)), with no individuals with biallelic known/likely pathogenic variants. **Discussion:** These findings suggest that several DPPMs may be more common than previously reported. This approach has the advantages of reducing ascertainment bias, linking genotypic to phenotypic EHR data, and generating hypotheses that can be further tested through deep-phenotyping in collaboration with the NHGRI Reverse Phenotyping Core. This project will complement our natural history protocol by aiding in hypothesis generation, elucidating the prevalence, and improving understanding of the clinical and genomic characteristics of these understudied disorders.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2031F: Establishment the transgenic pig model and exploration the molecular mechanisms of rare disease white sponge nevus

Authors:

W. Cai; Tongji Univ. Sch. of Med., Shanghai, China

Abstract:

White sponge nevus (WSN) is an autosomal dominant inherited rare disease and associated with the mutation of Keratin 4 (KRT4) and Keratin 13 (KRT13) gene. Till now, almost all WSN study are focus on case report or mutation report. The mechanism of WSN is still unclear. In this project, we recruited WSN Chinese family, and oral lesions biopsy with hematoxylin and eosin staining showed that patients had significantly pathological characteristics. The mutations of KRT4 and KRT13 gene were detected by PCR and direct sequencing. Sequencing analysis shows that the mutation of KRT13 was T332C, and amino acid change was Leu111Pro. Preliminary molecular research showed that the abnormal degradation of KRT13 protein of WSN probably associate with abnormal ubiquitination process. However, the molecular mechanism of this process in vivo is not comprehensive. In this project, using the CRISPR/Cas9 to build the KRT13 T332C WSN transgenic mini pig model and phenotyping, RNA-seq sequencing expression profiles provide important signaling pathways important to focus on gene transgenic mini pig cell transcriptome and genome comparative analysis to explore the pathogenesis of WSN. The platform of phenotype reversal comprehensive treatment and drug screening to interpretation and validation of molecular pathogenesis mechanism WSN in vivo. It provides an important

reference for further explore the molecular mechanism of WSN, clinical treatment and drug screening to interpretation and validation of molecular pathogenesis mechanism WSN in vivo. It provides an important reference for further explore the molecular mechanism of WSN, clinical treatment and drug screening.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2032F: Estimating the disease burden due to genetic variants associated with hypocalcemia

Authors:

J. B. Chang[1], A. M. Apostolov[2], M. M. Soliai[1], C. P. Barnhill[1], L. M. Stapleton Smith[1], A. S. Mathew[1], J. L. Nierenberg[1], J. Diao[3], C. D. Fernando[3], B. W. Dulken[1], A. Petukhov[1], A. V. Sridhar[1], C. Lange[4], R. Altman[5], T. M. Josephs[3], C. M. Gorvin[6], M. Roberts[1], S. H. Adler[1], J. C. Fox[1], S-G. Ji[7]; [1]BridgeBio Pharma, Inc., Palo Alto, CA, [2]Dept. of Bioengineering, Stanford Univ., Stanford, CA, [3]Drug Discovery Biology Theme, Monash Inst. of Pharmaceutical Sci., Monash Univ., Victoria, Australia, [4]Harvard Sch Publ. Hlth., Boston, MA, [5]Dept.s of Bioengineering, Genetics, and Med., Stanford Univ., Stanford, CA, [6]Universities of Birmingham and Nottingham, Birmingham, United Kingdom, [7]BridgeBio Pharma Inc., Palo Alto, CA

Abstract:

A challenge to the accurate diagnosis and treatment of rare diseases is that they present with common symptoms. Genetic testing can aid diagnosis but is hindered by incomplete understanding of the pathogenicity, penetrance, and expressivity of detected variants. For instance, the monogenic disease autosomal dominant hypocalcemia type 1 (ADH1) is caused by variants in the calcium sensing receptor (CaSR) thought to be extremely rare and primarily studied in small, family-based, clinically ascertained cohorts. This limits our understanding of the penetrance and expressivity of these *CASR* variants. To clarify the genotype-phenotype relationship of *CASR* variants and estimate ADH1 prevalence, we examined two large population-based biobanks, the UK Biobank (UKB) and All of Us (AOU).
We first replicated association of 8 of 11 known ADH1-associated variants found in UKB/AOU with ADH1 phenotypes. Interestingly, relatively few individuals (44% in AOU and 9% in the UKB) had a diagnosis (hypocalcemia or hypoparathyroidism) that suggested awareness of the underlying cause of disease. Consistent with previous reports, we found that the frequency of these variants was 2.8-4.6:100,000 across four biobanks (UKB, AOU, TOPMed, and gnomAD). Next, we identified novel ADH1-associated variants by developing a scoring method to quantify the association of a variant with ADH1 phenotypes. We discovered 115 variants (Group 1) that

scored above threshold across either biobank and six variants (Group 2) that scored above threshold in both biobanks. Interestingly, Group 2 variants had intermediate effects on ADH1 phenotypes. We further validated these variants by examining results from a genetic testing program for nonsurgical hypoparathyroidism and further tests performed by Invitae® as well as sensitivity to calcium *in vitro*. Notably, in the UKB, symptomatic individuals with Group 2 variants were found at approximately double the frequency of known ADH1-associated variants (7.4:100,000). 17% of these individuals had a diagnosis of hypocalcemia or hypoparathyroidism. Finally, we explored the phenotypic spectrum of Group 2 variants using a pheWAS and found associations with bone (1.7E-8), cardiac (4.4E-17), and other disorders.
Our work clarifies the genotype-phenotype relationship of CASR variants and enables more accurate diagnosis of ADH1. Furthermore, the intermediate effect variants complete a CASR allelic series for serum calcium, further supporting that CaSR is a master regulator of the serum calcium set point. Finally, our work suggests that genes causing monogenic disorders may underly a substantially greater disease burden than previously appreciated.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2033F: Exploring the genetic architecture of orofacial clefts in Filipino trios identifies GRHL2 as a novel gene associated with clefting

Authors:

S. Curtis[1], C. Yang[2], J. Murray[3], C. Padilla[4], T. Beaty[5], L. Moreno[6], E. Feingold[7], J. Shaffer[8], S. Weinberg[7], M. Marazita[9], M. Epstein[10], D. Cutler[2], H. Brand[11], E. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Emory Univ. Sch. of Med., Atlanta, GA, [3]U of Iowa, Iowa City, IA, [4]Univ. of the Philippines Manila, Manila, Philippines, [5]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [6]Univ Iowa, Iowa City, IA, [7]Univ of Pittsburgh, Pittsburgh, PA, [8]Univ. of Pittsburgh, Pittsburgh, PA, [9]Ctr. for Craniofacial and Dental Genetics, Unive, Pittsburgh, PA, [10]EMORY Univ., Atlanta, GA, [11]MGH, Boston, MA

Abstract:

Orofacial clefts (OFCs) are one of the most common structural birth defects, occurring in 1 in 700 live births. OFCs have a strong genetic basis, with genetic variants of all types and frequencies increasing risk. However, studies on the genetic basis of OFCs in cohorts with non-European ancestry are limited, and therefore, the genetic underpinnings in diverse populations are understudied. Here, we utilized data from the Gabriella Miller Kids First (GMKF) Research Program, which had whole genome sequencing data on 419 parent-child trios from the Philippines to better study OFC genetic architecture. Due to their potentially high deleterious

effect, we first focused on de novo variants (DNVs). After rigorous quality control, we identified 581 coding DNVs, with there being a significant enrichment in both LoF (N=105; p = 3.79e-17) and protein-altering DNVs (N=365; p = 1.35e-12) among OFC probands. Compared to expected mutation rates, several novel genes were enriched for LoF DNVs, including GRHL2 (p = 6.60e-6), where there were two DNVs, a stop-gain and a frameshift deletion. Combining this data with OFC trios from other populations in GMKF revealed an additional 89kb de novo deletion in GRHL2 in a proband from Colombia, giving further evidence of its role in OFC development. GRHL2 is a conserved transcription factor involved in embryonic development, with truncating mutations causing autosomal dominant progressive hearing loss and missense variants causing autosomal recessive ectodermal dysplasia. Interestingly, heterozygous variation in its homolog, GRHL3, is known to cause Van der Woude syndrome and isolated cleft palate. Additionally, mice deficient for either Grhl2 or Grhl3 have craniofacial anomalies, including facial and palatal clefts. Next, to further explore the genetic architecture of OFCs in Filipinos and the role of GRHL2, we also conducted a common variant (MAF > 5%) transmission disequilibrium test (TDT), and found suggestive association at GRHL2 (lead SNP: chr8:101536977:A:T; p = 7.8e-5) as well as genome-wide significance at a known clefting locus, NTN1 (lead SNP: chr17:9020401:A:T; p = 5.78e-9). Overall, we found evidence for a role for GRHL2 in the etiology of OFCs in Filipino populations and expanded upon our understanding of the genetic architecture of OFCs in people of non-European ancestry.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2034F: Familial TBX5 duplication associated with Currarino-Silverman syndrome

Authors:

D. Chitayat[1,2], M. Mackley[3], R. Silver[1], E. Greenfeld[4], M. Roifman[1], S. Redmond[3], K. Davidge[5]; [1]The Prenatal Diagnosis and Med. Genetics Program, Dept. of Obstetrics and Gynecology, Univ. of Toronto, Toronto, ON, Canada, [2]The Hosp. for SickKids, Div. of Clinical and Metabolic Genetics, Toronto, ON, Canada, [3]Div. of Clinical and Metabolic Genetics, Dept. of Pediatrics, The Hosp. for Sick Children, Univ. of Toronto, Toronto, ON, Canada, [4]Dept. of Pathology and Lab. Med., Mount Sinai Hosp., Toronto, ON, Canada, [5]Div. of Plastic and Reconstructive Surgery, Hosp. for SickKids; Univ. of Toronto,, Toronto, ON, Canada

Abstract:

Currarino-Silverman (CS) syndrome is a rare condition characterized by prominent sternal angle of Louis with manubrial pectus carinatum and lower sternal pectus excavatum and an increased

incidence of cardiac abnormalities. The TBX5 gene code for the T- binding box 5 protein, a transcription factor that plays an important role in the formation of tissues and organs during embryogenesis, regulating the activity of other genes by binding to specific regions of the genome, and activating genes mainly involved in the normal development of the upper limbs and heart. Pathogenic variants in TBX5 result in Holt-Oram syndrome (HOS), a rare autosomal dominant syndrome characterized by upper limb and cardiac defects. Most reported pathogenic variants are single nucleotide variants. Here we report an infant with an intragenic TBX5 duplication inherited from his father with the same but de novo TBX5 variant, associated with Currarino-Silverman syndrome. Case: The mother was of Asian descent and the father was of Spanish descent. The mother was healthy and the father had a history of ventricular septal defect (VSD) and Currarino-Silverman syndrome. During the pregnancy with their son fetal ultrasound showed upper limbs' abnormalities and CMA done on amniocentesis showed a heterozygous partial duplication the TBX5 including exons 2 through 7. The variant was paternally inherited and de novo in the father. The baby was born at term. Postnatal physical examination was notable for bilateral absent pectoralis muscles. The right thumb was digitalized, proximally inserted and with narrow first interdigital space. The left thumb was rudimentary with syndactyly with the second finger. Both shoulders were narrow and adductable and the xiphoid was prominent. The father was operated on in childhood and had modified Ravitch-type thoracoplasty. On examination he had normal hands, mild pectus carinatum (post-surgery) and absent pectoralis major muscle. Conclusions: To our best knowledge, this is the first reported case of Currarino-Silverman (CS) syndrome caused by an intragenic duplication in TBX5. Importantly, this represents only the third published case of intragenic TBX5 duplication (PMIDs: 22333898; 33930582). The baby has asymmetric upper limb abnormalities, without a congenital heart defect and notably, the father has a VSD and an unusual chest wall deformity without limb anomalies. This case shows the intrafamilial variability of the clinical manifestations associated with TBX5 duplication and raises the possibility that Currarino-Silverman (CS) syndrome is the result of TBX5 duplication.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2035F: Female carriers of X-linked myotubular myopathy (XLMTM) in Germany: Clinical characterization of a cohort extends the phenotypic spectrum and knowledge about the impact of heterozygous variants in the *MTM1* gene

Authors:

F. Braun[1,2], F. J. Kaiser[2], U. Schara-Schmidt[1]; [1]Dept. of Pediatric Neurology, Ctr. for Neuromuscular Disorders, Ctr. for Translational Neuro- and Behavioral Sci., Univ. Duisburg-Essen, Essen, Germany, [2]Inst. of Human Genetics, Univ. Hosp. Essen, Univ. Duisburg-Essen, Essen, Germany

Abstract:

The mode of inheritance of X-linked myotubular myopathy (XLMTM) is still considered recessive, although female carriers are frequently symptomatic. We aimed to characterize the phenotypic spectrum and burden of disease in a carrier cohort.
We performed a cross-sectional study among XLMTM carriers in Germany, recruiting predominantly through a patient association. We used a custom-made questionnaire on medical and family history, the Checklist Individual Strength (CIS20-R), Short Form 12 (SF-12) health survey, Pain Detect, Short-Form McGill Pain Questionnaire (SF-MPQ-D), Graded Chronic Pain Scale (GCPS), and Depression Anxiety Stress Scale (DASS21). All participants received a neurologic examination including manual muscle testing (MMT), a 6-minute walking test (6MWT), and timed up-and-go (TUG) test. Where possible, we performed ultrasound of the abdomen and vastus lateralis muscles, echocardiography, pulmonary function testing, blood count, and further laboratory testing. Carriers were classified as manifesting based on the degree of muscular weakness and ambulatory status.
In our study population (n=14), most carriers were classified as manifesting with a mild (n=8; independent ambulation, with limb or axial weakness) or minimal phenotype (n=3; only facial muscle weakness). We consider only the minority non-manifesting (n=3). Frequent features include a high-arched palate, fatigue, muscle pain and stiffness, and pulmonary problems. Many carriers are caregivers of boys with XLMTM (n=11). Families are burdened with stillbirth and death in early childhood (n=8).
The majority of female carriers of XLMTM show neuromuscular and additional symptoms. This should be considered in diagnostics, clinical care, and future clinical trials.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2036F: *FGF14*, *ZFHX3*, *NOTCH2NLC* gene repeat variant found in clinically diagnosed multiple system atrophy.

Authors:

M. Matsushima[1,2,3], H. Uwatoko[1], S. Shirai[1], I. Iwata[1], H. Yaguchi[1,2], S. Miyatake[4], N. Matsumoto[4], T. Yamada[2], I. Yabe[1,2]; [1]Dept. of Neurology, Faculty of Med. and Graduate Sch. of Med., Hokkaido Univ., Sapporo, Japan, [2]Div. of Clinical Genetics, Hokkaido Univ. Hosp., Sapporo,

Japan, [3]Hokkaido Univ. Hlth.Care Ctr., Sapporo, Japan, [4]Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan

Abstract:

[Background/Purpose] Approximately 30% of spinocerebellar degeneration (SCD) cases in Japan are hereditary, while many of the sporadic cases are referred to as multiple system atrophy (MSA). We have been continued to report on the proportion of disease type of SCD in Hokkaido, Japan. We also examined the genetic variants of newly reported conditions in recent years, such as SCA27B, SCA4, and neuronal intranuclear inclusion disease (NIID). NIID is a type of progressive dementia with leukoencephalopathy, characterized by a variety of clinical symptoms including ataxia. Since NIID, along with SCA27B and SCA4, can be part of the differential diagnosis for MSA, it is necessary to consider these conditions in genetic testing.[Subjects/Methods] We conducted genetic analysis on 398 cases clinically diagnosed with MSA at medical institutions in Hokkaido from January 1997 to December 2023. Using stored DNA from consenting patients, after excluding known variants of hereditary SCD, we examined the presence of pathological variants in SCA27B, SCA4, and NIID. Specifically, we assessed for abnormal repeat expansions of the GAA in *FGF14*, the GGC in *ZFHX3*, and the GGC in *NOTCH2NLC* using methods such as RP-PCR.[Results] Among the cases excluding known hereditary SCD, two instances showed an expansion of more than 200 GAA repeats in *FGF14*. Additionally, with respect to *NOTCH2NLC*, one case demonstrated abnormal expansion of the GGC repeat. None of the cases had a family history of similar symptoms. The clinical symptoms, progression, and imaging findings were consistent with MSA. No GGC repeat expansions were observed in *ZFHX3*.[Discussion/Conclusion] Cases clinically diagnosed with MSA were found to have GAA repeat expansions in *FGF14* and GGC repeat expansions in *NOTCH2NLC*. However, further investigation is needed to understand the significance of these findings. As in the present cases, pathogenic variants of hereditary diseases can sometimes be detected in sporadic neurological disorders. However, it is necessary to conduct a medically appropriate evaluation that includes not only the results of genetic testing but also other clinical information.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2037F: Generating pituitary cells from iPSCs for pituitary disease modeling and drug discovery

Authors:

J. Marques[1], F. Fattahi[2], L. Silveira de Carvalho[1]; [1]Hosp. das Clinicas da Faculdade de Med. da Univ.e de São Paulo, São Paulo, Brazil, [2]Univ. of California San Francisco, San Francisco, CA

Abstract:

Introduction: Reduced pituitary hormones due to genetic or organic causes is known as combined pituitary hormone deficiency (CPHD) and *PROP1 variants* are frequent causes of CPHD. Induced pluripotent stem cells (iPSC) can be used to generate patient pituitary cells for disease modeling and therapy improvement. **Aim:** To differentiate pituitary cells from healthy and CPHD individuals and use them for disease modeling and drug discovery. **Methods:** Peripheral blood was collected from patients harboring *PROP1* pathogenic variants and controls. PBMCs were isolated and expanded prior to nucleoporation and transfection with pluripotency plasmids (*OCT4* and *SOX2*; *cMYC* and *KLF*). Cells were transferred to Geltrex coated plates and fed with E8, NaB and βFGF until iPSC colonies were formed. iPSCs were differentiated into cranial placode cells using pituitary placode conditioned medium (PPCM) until day 15. For pituitary cell maturation, pituitary conditioned medium (PPCM without SB431542) was used until day 30. qPCR and flow cytometry using the Kit BD Stemflow (BD) for stem cells (SOX2, NANOG and OCT3/4) and the eBioscience™ Foxp3/ Fixation/Permeabilization (Thermo Fisher) for the placode and pituitary markers (PAX6, SIX1, LHX3, PROP1) were used for cell characterization. Pituitary hormonal levels from cell supernatant were measured. Pituitary cells were passaged into 384 well plates and treated with FDA-approved library with 3113 compounds at 1µM for 72h (#L1300, Selleckchem). Cells were stained with ACTH and imaged on ImageXpress® HCS. ACTH intensity was normalized to total cell number using MetaXpress Software and compounds z-score was calculated considering ±2.5 for analysis. Gene and pathway enrichment analysis were performed to determine target drugs based on similar target gene. **Results:** We generated iPSCs from patient and control cells with more than 90% positivity for pluripotency markers SOX2, NANOG, OCT3/4. Pituitary cells were generated with positivity for the placode marker SIX1 (58%) and pituitary markers, as LHX3 (56%) and PROP1 (59%). Control cell lineages expressed pituitary hormones at days 15 and 30 while patient cells failed to induce hormone-producing cells properly, as expected. We performed a high throughput drug screen using pituitary cells leading to the discovery of 17 compounds that were validated *in vitro* and *in vivo* and might be used pituitary disease treatment. **Conclusion:** We generated pituitary hormone producing cells. Patient cells mimic human phenotype and can be used for drug discovery. The present study elucidates CPHD basis in patients specific background, develop pituitary disease cell model as well as possible new treatments.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2038F: Genetic Analysis of Patients with Intellectual Disability and Congenital Anomalies Presenting with Hypomelanosis of Ito

Authors:

S. Mizuno[1], T. Uehara[1], S. Otsuji[1], N. Nakamura[1], H. Kurahashi[2], K. Kosaki[3], S. Hayashi[4], M. Inaba[1]; [1]Central Hosp., Aichi Dev.al Disability Ctr., Kasugai, Aichi, Japan, [2]ICMS, Fujita Hlth.Univ., Toyoake, Japan, [3]Keio Univ. Sch. of Med., Tokyo, Japan, [4]Inst. of Dev.al Res., Aichi Dev.al Disability Ctr., Kasugai, Aichi, Japan

Abstract:

Hypomelanosis of Ito (HI) is a rare congenital pigmentary disorder characterized by irregular hypopigmented patches (macules) on the trunk and limbs, often following the lines of Blaschko. It is occasionally associated with congenital malformation syndromes and developmental delay. The pathogenesis is presumed to involve mosaic mutations during the early division of the zygote, although the specific etiology often remains unclear.
In this study, we retrospectively examined 25 patients with intellectual disability and multiple congenital anomalies who presented with HI on the trunk and were seen at our facility. This study was conducted in accordance with ethical guidelines. We evaluated the genetic test results and skin findings of these patients. Chromosomal analysis was performed on all cases, revealing autosomal abnormalities in nine cases (structural abnormalities in two, aneuploid mosaicism in two, and additional structural abnormalities in five) and an X chromosome abnormality in one case. No significant findings were obtained from microarray analysis of the five cases with normal karyotypes. However, comprehensive next-generation sequencing (NGS) analysis identified X chromosome gene disorders in two out of three cases, with heterozygous missense variants in TFE3 and USP9X.
Mosaic cell formation during development and random X chromosome inactivation are suggested as potential pathogenic mechanisms. Due to the limited number of cases, the morphological differences in macules between these mechanisms could not be confirmed. The presence of HI may aid in the genetic investigation of causes and variant evaluation in unexplained intellectual disability and congenital anomaly syndromes.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2040F: Genetic etiologies and diagnostic yield of exome sequencing in pediatric motor speech disorders

Authors:

M. Mitchel[1], M. Oetjens[2], A. S. F. Berry[1], A. Johns[3], A. Moreno-De-Luca[1], R. I. Torene[4], N. T. Strande[1], M. T. DiStefano[5], L. H. Dyer[6], T. Brandt[7], B. M. Finucane[1], D. H. Ledbetter[8], K. Retterer[4], C. Martin, L.[9], S. M. Myers[1]; [1]Geisinger Autism & Dev.al Med. Inst., Lewisburg, PA, [2]Geisinger Autism & Dev.al Med. Inst., Lewisburg, PA, [3]Geisinger Dept. of Population Hlth.Sci., Danville, PA, [4]Geisinger, Danville, PA, [5]The Broad Inst., Cambridge, MA, [6]GeneDx, Gaithersburg, MD, [7]Gene Dx, Gaithersburg, MD, [8]U. Florida Coll. of Med.-Jacksonville, Jacksonville, FL, [9]Geisinger Autism & Dev.al Med. Inst., Danville, PA

Abstract:

Background: Pediatric motor speech disorders (MSDs), including apraxia and dysarthria, are caused by deficits in motor planning, programming, or execution. Compared to common childhood speech problems, MSDs tend to be severe and intractable. Few studies have investigated the genetic underpinnings of MSDs, and they have been limited by small sample sizes and an exclusive focus on apraxia. We aimed to identify pathogenic genomic variants in known and new candidate genes for MSDs and determine the diagnostic yield of exome sequencing (ES) in a large cohort of children with MSDs.**Methods:** This retrospective cross-sectional study identified patients with apraxia and/or dysarthria, ages 2- 21 years, who underwent clinical ES between 2012 and 2022. The diagnostic yield, defined as the percentage of pathogenic/likely pathogenic (P/LP) findings considered causative of the MSD phenotype, was calculated and delineated by sex and neurodevelopmental disorder (NDD) comorbidity. For each gene with a P/LP variant, the number of patients who had MSD diagnoses was compared to the number who did not. Statistical over-representation of an MSD diagnosis was calculated with Fisher's Exact test with Bonferroni correction for multiple comparisons to identify the genes with the highest likelihood of an MSD phenotype. Multivariable logistic regression adjusting for age and sex was used to estimate the associations between comorbidities and positive genetic findings. **Results:** Among 2004 patients with an MSD diagnosis (mean 9.6 years, S.D. 5.0 years; 39.3% female), a positive diagnostic result was detected in 527, resulting in a diagnostic yield of 26.3%. The yield was higher for females than for males (31.6% vs 22.8%; $p < 0.001$). Individuals with NDD comorbidities were more likely to have a P/LP finding than those without comorbidities (26.7% vs 13.3%, $p = 0.03$). Out of 527 individuals with positive genomic findings, 495 (93.9%) had sequence variants and 32 (6.1%) had copy number variants (CNVs). Twenty-nine different CNVs were identified, and P/LP sequence variants were detected in 262 genes, including 102 genes with variants in 2 or more patients. Thirty genes were nominally enriched for a MSD phenotype when the number of patients with P/LP variants with and without an MSD diagnosis were compared. Three of these genes survived study-wide correction, two of which also survived exome-wide correction. **Conclusion:** Over a quarter of patients with pediatric MSDs harbored P/LP variants, and many genes were newly implicated in MSDs. These findings

suggest that clinical ES should be considered for children with MSDs, similar to the standard of care for other NDDs.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2041F: Genetic Investigation of Consanguineous Pakistani Families Segregating Rare Neurological Disorders

Authors:

M. Ilyas[1], H. Houlden[2]; [1]Islamia Coll. Univ., Peshawar, Pakistan, [2]Inst. of Neurology UCL, London, United Kingdom

Abstract:

Neuro-developmental disorders have become a major public health problem in Pakistan in recent years. Its management and treatment strategies are a daunting problem in Pakistan due to poorly developed primary and basic health care facilities, lack of funding, and weak political processes. This increasing rate depends upon certain rare, neglected and often difficult to understand neurological conditions. Therefore, these disorders require strict attention, particularly in rural areas where these disorders remain unexplored. Identification of specific genetic markers may provide a useful explanation for molecular characterization, disease etiology and pathogenesis. We used Next generation sequencing to investigate the novel causative gene(s) in hundreds of families and understand the role in order to develop effective treatments for neurological and neuro-developmental disorders. Medical examination was performed of about 150 families of the same kinship group with affected children. Whole exome sequencing of 70 families was performed and the causative genetic variants were clarified in the families, ten convincing candidate genes were identified. We aim to provide insights into the processes that have modeled the extant genetic diversity in the Pakistani populations as well as the consequences of the diversity observed, including their relation to disease susceptibility. Our goal is to pursue this mission by emphasizing individual and collaborative faculty investigations in confluence with this genetic research cycle paradigm.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2042F: Genome Sequencing of 400 rare disease patients through the Korean Regional Rare Disease Diagnostic Support Program

Authors:

R. Khang[1], J. Kim[1], D. Moon[1], S. Jang[1], E. Lee[1], Y. Song[1], S. Ryu[1], S. Lee[1], H. Han[1], S. Kim[1], S. Jang[1], S. Oh[2], J. Lee[3], G. Seo[1], C. Cheon[4], H. Lee[1]; [1]3billion, Seoul, Korea, Republic of, [2]Dept. of Lab. Med., Pusan Natl. Univ. Yangsan Hosp., Pusan Natl. Univ. Sch. of Med., Yangsan, Korea, Republic of, [3]Div. of Rare Disease Management, Korea Disease Control and Prevention Agency, Cheongju, Korea, Republic of, [4]Rare Disease Ctr. of Gyeongnam Region, Dept. of Pediatrics, Pusan Natl. Univ. Children's Hosp., Pusan Natl. Univ. Sch. of Med., Yangsan, Korea, Republic of

Abstract:

Exome and genome sequencing (ES/GS) has proven to be a powerful diagnostic solution by enabling the simultaneous screening of thousands of rare genetic disorder genes. However, still, ES/GS are still not approved as diagnostic tests in many countries, which limits patients' access to these tests. The Republic of Korea is one such country and, in response, the Korea Disease Control and Prevention Agency (KDCA) has established a pilot study, the Korean Regional Rare Disease Diagnostics Support Program, a collaborative initiative involving 11 regional rare disease centers (RRDC) across Korea. The objective of the study was to provide patients residing outside Seoul, who have even more limited access to genetic testing because they must travel afar, with the opportunity to receive genetic testing through their RRDC. This study examines the diagnostic utility of GS within this program. A total of 400 patients with a wide range of clinical manifestations were enrolled in the study, with the neurodevelopmental delay group being the largest. The overall diagnostic yield was 36% (145/400), with 11% (15/145) of diagnosed patients harboring variants beyond the scope of detection by chromosomal microarray or ES. To achieve the greatest cost-effectiveness, only the proband was subject to GS, and therefore, subsequent Sanger sequencing (SS) was performed for segregation analysis, which played a critical role in reclassifying the pathogenicity of 32% variants of uncertain significance. SS reports were also generated for all positive cases as SS is accepted as a diagnostic test method in Korea and SS reports can be used for insurance purposes. The overall diagnostic rate across the RRDC was not significantly different ranging from 25% to 46% (p=0.08). For the 145 diagnosed patients, it took a median of 4 years from disease onset to molecular diagnosis. For 7 patients, the time to diagnosis exceeded 30 years. Finally, the implementation of a centralized system streamlined the diagnostic workflow, facilitating timely reporting within a reasonable turnaround time (TAT) of less than 35 days. This study demonstrates the successful implementation of the program for the 11 RRDCs across South Korea, with a practical workflow, comprehensive testing, comparable diagnostic yield to previous reports, and a reasonable TAT. Most importantly, the study was highly evaluated by both the treating physicians in each RRDC and the patients and their families, as it eliminates the need for travel to Seoul to receive genetic testing and still provides comprehensive testing within a reasonable TAT. In light of these encouraging outcomes, the KDCA has decided to continue the program in the upcoming year.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2043F: Genome-wide association study reveals novel loci associated with acute chest syndrome in sickle cell disease patients

Authors:

A. Kulshrestha, M. E. Garrett, M. J. Telen, A. E. Ashley-Koch; Duke Univ. Med. Ctr., Durham, NC

Abstract:

Sickle cell disease (SCD) results from a genetic mutation in the beta hemoglobin gene. Despite the same underlying causal mutation, patients exhibit highly heterogeneous clinical outcomes. Acute chest syndrome (ACS) is the leading cause of mortality among individuals with SCD, accounting for 25% of all deaths, and is the second most common reason for hospitalizations after vaso-occlusive pain. It is associated with extended hospital stays, higher risk of respiratory failure, and the potential for developing chronic lung disease after recurrent episodes. Many previous genome-wide association studies (GWAS) have failed to identify risk loci associated with ACS due to its multifactorial etiology and the absence of a well-defined, consensus phenotype.

To identify genetic loci associated with ACS, we performed a GWAS with 196 adult SCD patients from the Outcome Modifying Genes in SCD (OMG-SCD) cohort. We selected the extreme phenotype, whereby patients who had never had any episode of ACS were categorized as controls (n=114), while patients who had had more than 5 episodes of ACS were categorized as cases (n=82). All patients included in the study either had HbSS (homozygous sickle) or HbSβ^0 (sickle beta-0 thalassemia) genotype. Genotype data were generated by whole genome sequencing by Trans-Omics for Precision Medicine - TOPMed (phs001608). We performed the GWAS using the R package SAIGE, incorporating treatment (hydroxyurea) status and 2 genomic principal components as covariates. Linkage-disequilibrium clumping was performed using PLINK 1.9 to identify lead single nucleotide polymorphisms (SNPs).

Our top SNP associated with ACS was rs7319269 on chromosome 13, reaching the genome-wide threshold of $5x10^{-8}$. This intergenic SNP lies between *SNORA107* and *LINC00375*. In total, 175 SNPs with p-value less than $5x10^{-5}$ were nominally associated with ACS. After adjusting for False Discovery Rate (FDR), all these SNPs had q-values less than 0.05. Interestingly, one of our top SNPs (rs1356744 on chr 2) has been previously associated with severe acute respiratory syndrome in COVID-19, suggesting a common biological pathway in ACS and COVID-19.

Our study highlights the importance of a well-defined phenotype with distinct cases and controls for successful identification of risk loci through GWAS. To further understand the biological implications of these SNPs, we are deploying post-GWAS methods for performing

functional analyses. Our goal is to decipher the intricate biological mechanisms and pathways involved in severe SCD so that we can enhance treatment approaches and provide valuable insights into the molecular profiles specific to severe SCD.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2044F: Homozygous variants of *ELOVL1*presenting with severe liver diseases

Authors:

C-H. Tsai, s. krysinski, a. zalan, E. Falsey; Univ. of Illinois Chicago, Chicago, IL

Abstract:

ELOVL1, an enzyme crucial for the elongation of fatty acids into very-long-chain fatty acids (VLCFAs), plays a vital role in various biological processes. Defects in ELOVL1 can lead to a spectrum of related disorders. Previous cases reported resulted from de novo, autosomal dominant variants. We herein report an autosomal recessive case presenting with severe liver disease. Patient is a 6-month-old female, born full-term to a consanguineous couple, flagged by newborn screening due to undetectable C26:0. At birth, she presented with hyperbilirubinemia (Total/Direct 16/5). She was supplemented with fat soluble vitamins, DHA, and Ursodiol. Liver biopsy revealed mild cholestasis, few apoptotic hepatocytes, and mild fibrosis of portal tracts. She had marked elevation in urinary bile acid excretion consistent with severe cholestasis. There was a non-specific elevation in bile alcohol glucuronides indicating underlying hepatic dysfunction. Ursodiol was discontinued after 2 months due to worsening cholestasis and hyperbilirubinemia of 35/30. After addition of Chobham supplementation, bilirubin decreased to 25/20. Cholangiogram at 3 months of age was reassuring against biliary atresia and a second liver biopsy showed marked hepatocellular and canalicular cholestasis without fibrosis. Chromosomal microarray identified multiple regions of homozygosity. Karyotype was normal. A neonatal crisis gene panel identified a heterozygous variant of uncertain significance (VUS) in *SOS2*, associated with Noonan syndrome type 9. Whole exome sequencing identified a homozygous likely pathogenic variant in *COL9A2* associated with Stickler syndrome, compound heterozygous VUS in *SPTB* associated with hereditary spherocytosis and a homozygous VUS in *ELOVL1* **(*ELOVL1 (c. 458G>A), homozygous*.**) Given the C26:0 level of 0 and the clinical constellations, we consider the variant pathogenic. Family history is significant for an older brother who died at 2 months due to liver failure. Our patient's clinical features include mildly dilated ascending aorta, peripheral pulmonary stenosis, and patent foramen ovals, ichthyosis, severe hyperbilirubinemia and hearing deficits. She also has three episodes of salmonellosis. She is followed by ophthalmology and hematology. Her severe liver disease associated with ELOVL1 that is not previously reported, and potentially exacerbated by hereditary spherocytosis and other unidentified AR traits by WES. Our patient also demonstrated inappropriate cholic acid can worsen liver disease. The patient is currently on liver transplant wait list.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2045F: Identification of a novel splicing variant in *LEPR* associated with severe early onset obesity

Authors:

A. Butt[1], A. H. Soomro[2], S. Siddique[3,4,5]; [1]Dept. of BioSci.s, COMSATS Univ. Islamabad (CUI), Islamabad, Pakistan, [2]Dept. of Paediatrics, Maryam Mem. Hosp., Rawalpindi, Pakistan, [3]Sch. of Med. and Dentistry, Univ. of Central Lancashire, Preston, United Kingdom, [4]American Univ. of the Caribbean (AUC), Sch. of Med. - UK Track, Univ. of Central Lancashire, Preston, United Kingdom, [5]Molecular Pathology Div., MEDFER Diagnostics, Islamabad, Pakistan

Abstract:

Background: The leptin receptor (*LEPR)* plays a crucial role in regulating body weight and energy balance. This study aims to genetically characterize a pediatric case of severe early-onset obesity with complaints of persistent hyperphagia and mild speech delay. **Case Presentation:** A full-term male child, born via spontaneous vaginal delivery to consanguineous parents, with a birth weight and height of 3.5 kg and 46 cm respectively, presented with history of hyperphagia and rapid weight gain since birth. At 19 months, he weighed 35 kg with elevated serum triglyceride levels (2.62 mmol/L). Sub-clinical hypothyroidism was diagnosed based on thyroid profile and treated with levothyroxine (50 mcg). Initially suspecting congenital leptin deficiency, the case was referred for genetic counseling. A monogenic obesity NGS panel covering 68 genes identified two variants: a known pathogenic c.67_69del (p.Phe23del) in exon 1 of *SLC22A5* in a heterozygous state, and a novel c.995-2A>G (splice acceptor) variant in intron 8 of *LEPR* in a homozygous state. While the *SLC22A5* variant is pathogenic for primary carnitine deficiency, it is insufficient in a heterozygous state, implicating the *LEPR* variant in this case of monogenic obesity. SpliceAI and Human Splice Finder also predicted the deleterious nature of the intronic variant. Familial testing showed both parents were heterozygous carriers of the above noted mutations in *LEPR* and *SLC22A5*. Over a 31-month follow-up, the index continued to gain weight despite dietary restrictions and had linear growth without dysmorphic features or cognitive impairment. At present, his weight and height are 72 kg and 124 cm respectively. The use of setmelanotide was considered but is currently unfeasible due to age criteria and unavailability in Pakistan. **Conclusion:** This is the first report of pathogenic intronic variant and the third splicing variant in *LEPR* in the ClinVar database. Overall, clinical features, segregation, and computational analysis support its pathogenicity. While many previously reported mutations in *LEPR* correlate with severe early-onset obesity, their frequency is lower compared to mutations reported for *LEP* and *MC4R*. A recent report by Saeed et al. (2023) in a Pakistani cohort also highlighted high morbidity and mortality risks in children with *LEP* and *LEPR* mutations. As obesity therapies advance, identifying causative variants is essential for personalized clinical management, particularly in high-risk populations.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2046F: Identification of de Novo and novel mutations in LTBP2

Authors:

S. Irshad, r. S. saleem; Sch. of Biochemistry and Biotechnology, PU, LHR, Lahore, Pakistan

Abstract:

Background: Primary congenital glaucoma (PCG) is a leading cause of childhood blindness dueto optic nerve damage, enlarged globe and visual field defects. PCG is usually diagnosed withinthe first year of life.Objective:The aim of current study was to assess the involvement of Latent transforming growth factor 2(LTBP2) gene in PCG families from Pakistan and to find novel mutations.Methods: Genomic DNA was extracted from the whole blood of n=20 families. Affected andunaffected individuals of the families were genotyped with a whole genome single nucleotidepolymorphism microarray (SNP). Homozygosity mapping analysis was performed for theselected members of these families. LTBP2 gene was screened using Sanger Sequencing in n=20consanguineous Pakistani families diagnosed with PCG by standard ophthalmologicalexamination.Results: Novel homozygous mutations were identified in three families with PCG in LTBP2gene. We found a new de novo frameshift mutation c.1762_1763del; p. (Leu588Valfs*14). In asecond family, we identified a splice site mutation c.2531-2A>C, and in third family, a splicedonor site c.1686G>A; p. (Gln562Gln) mutation.Conclusions: In this study, we report the involvement of novel de novo frameshift mutation andtwo genetic variants that affect splicing in PCG families from Pakistan. The current study willhelp us to extend our understanding towards the role of LTBP2 in PCG.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2047F: Identification of novel pathogenic variants in the neurodevelopmental gene *LNPK*

Authors:

R. Doss[1], S. Wirth[1], J. Pitsch[2], A. Gropman[3], M. Breuss[4]; [1]Univ. of Colorado Anschutz, Aurora, CO, [2]Univ. of Colorado Anschutz Med. Campus, AURORA, CO, [3]Children s Natl. Hosp., Washington, DC, [4]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

LNPK encodes for lunapark, a conserved membrane-associated protein suggested to play several important biological roles as the curvature-stabilizing protein within tubular three-way junctions of the endoplasmic reticulum. Recently, several individuals with homozygous variants in *LNPK* who share a range of atypical neurodevelopmental phenotypes have been reported. Major phenotypes include moderate to severe intellectual disability, epilepsy, movement disorders, speech delays, hypotonia, corpus callosum hypoplasia, and ear-of-the-lynx sign on MRI. Here we present a child with novel compound heterozygous variants of unknown significance (VUS) within intronic splice site regions of *LNPK*, who displays related but milder phenotypes. We aimed to investigate the functional significance of these VUS on gene expression and possible defective splicing patterns. To this end, we obtained skin biopsies from the mother and the affected child and cultured fibroblast lines. Following the confirmation of genotypes by Sanger sequencing, we performed RNA extractions for RT-qPCR analysis and protein extractions for Western Blot assessment. This gene expression analysis revealed that, indeed, the mRNA and protein levels were significantly reduced in the child, whereas the mother's cells exhibited a genetic compensation allowing for typical lunapark expression. Additionally, we performed targeted PCR on cDNA samples followed by TOPO cloning and subsequent Sanger sequencing to identify possible splicing variations. This analysis revealed several unexpected transcript products, whereas the initially predicted intron retention was undetectable; this finding could impact potential therapeutic approaches in this or related cases, suggesting that molecular assessment of predicted splicing defects remains an important addition to validation workflows. The observed splice aberrations triggered nonsense-mediated mRNA decay (NMD) which was directly demonstrated through pharmacological inhibition of this pathway. Overall, our study supports that these novel *LNPK* variants are pathogenic. The milder phenotype presentation compared to the previously identified cases is possibly due to an incomplete loss of protein function.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2048F: Interpretation of neurological disorders cohort through in-house developed tool, FestiVAR, showed an improved diagnostic yield: data from over 1000 Indian patients

Authors:

A. Mannan[1], B. Forona[1], S. Sur[1], K. Karmalkar[1], N. K[1], J. S. Vasudevan[1], R. Dave[1], S. Nayanala[1], R. H. J.[1], S. Sridharan[1], S. Phalke[1], C. Judith[1], S. Katragadda[2], A. Janakiraman[3], V. Veeramachaneni[4], **J. Singh**[5], R. Hariharan[6]; [1]Strand Life Sci., Bangalore, India, [2]Strand Life Sci.,

Bangalore, India, [3]Strand Life Sci. Pvt Ltd, Bangalore, India, [4]Strand Life Sci. (P) Ltd., Bangalore, India, [5]Strand Life Sci., Sacramento, CA, [6]Strand Life Sci. Private Limited, San Francisco, CA

Abstract:

Background: Neurological disorders are one of the major causes of disability and death worldwide. These disorders are clinically heterogeneous group and a significant proportion of these are 'neurogenetic disorders', which include disorders, such as epilepsy, ataxia, leukodystrophy, neuropathy, dystonia, myopathy, autism, intellectual disability and others, that are mostly caused due to defects in one or more genes. A precise and confirmatory diagnosis in the patients in a timely manner is essential for appropriate therapeutic and management strategies. Due to the complexity of the clinical presentations across various neurological disorders, arriving at an accurate diagnosis remains a challenge. Methods: Whole exome sequencing (WES) at >100x coverage was used to sequence >1100 unrelated patients from India with suspected neurological disorders. Small nucleotide variants (SNVs) were identified using the Strand NGS software. Copy number variants (CNVs) were determined by taking a consensus between the Strand NGS and Illumina Dragen enrichment CNV results. Interpretation was done by using in-house tools: FestiVAR and StrandOmics platform.Results: The diagnostic yield in our cohort was 40% (pathogenic and likely pathogenic variants) and in 26% of cases, we detected VUS (variant of uncertain significance). The highest diagnostic rate was observed among patients with muscular dystrophy (65%) followed by leukodystrophy (51%). Interestingly, the detection rate in autism cases was lowest (22%). In terms of mutation types, we detected all types of variants with 50.4% of pathogenic/likely pathogenic variants being missense, 15.3% nonsense, 12.8% frameshift, 8.6% splice-site, 0.7% in-frame and 12.1% were structural variants (CNV). Conclusion: In our study, we observed an improved performance of whole exome testing, with an overall diagnostic yield of 40%. Furthermore, we show that our recently developed in-house interpretation tool, FestiVAR, is fast and efficient, which can minimize the time required to perform interpretation and it can be very effective in identifying causative genes/variants in complex neurological cases.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2049F: Investigating Missing Heritability in Inherited Retinal Dystrophy

Authors:

L. Ahmed[1,2], R. Osman[1,3], C. Eileen[1,4], K. Green-Sanderson[1,2], E. Tavares[1], K. Ahmed[5], E. Heon[1,2], A. Vincent[1,2]; [1]SickKids, Toronto, ON, Canada, [2]Inst. of Med. Sci., Univ. of Toronto, Toronto, ON,

Canada, [3]Queen's Univ., Kingston, ON, Canada, [4]Faculty of Arts & Sci., Univ. of Toronto, Toronto, ON, Canada, [5]Sickkids, Toronto, ON, Canada

Abstract:

Background Inherited Retinal Dystrophy (IRD) is a group of degenerative disorders causing vision loss. Around 300 genes are associated with IRDs, but 15-50% of cases remain "gene-elusive." This study aims to investigate missing heritability in "gene-elusive" cases. **Methods** Six cases with presumed autosomal/X-linked recessive IRD & inconclusive/negative clinical panel tests underwent ocular exams and genome sequencing (GS). Following annotation & validation, variants were filtered using phenotype-driven panels, followed by whole genome filtering. Structural variants (SVs) & single-nucleotide variants (SNVs) defined as rare based on population frequencies (<0.5%) were kept. SNVs in coding regions were screened for pathogenicity using an in-lab pathogenicity prediction pipeline, and homozygous variants in regulatory regions and untranslated regions (UTRs) were kept. **Results** Two children & four adults with IRDs were included. Diagnoses included Leber's congenital amaurosis (LCA), early onset retinal dystrophy (EORD), retinitis pigmentosa (RP), fundus albipunctatus & multifocal vitelliform dystrophy (MVD). After GS analysis, candidate variants were identified in 5/6 cases, which included deep intronic & 3' UTR variants. Case 2, a 14-month-old male with LCA, had 4 RPGRIP1 variants. Maternal variants included 1 pathogenic variant (c.1468-2A>G) & 2 variants of uncertain significance (VUS) (c.3100-229_3238+234dup, c.491-112G>A). One paternal VUS (c.906+26T>C) was identified. Functional validation of the paternal variant is being conducted. Case 3, an 11-year-old male with EORD had compound heterozygous variants in CRB1, 1 pathogenic (c.3415C>T/p.Q1139X) & 1 VUS (c.849-26A>G). Minigene confirmed the VUS to cause exon 4 skipping & a premature stop codon (p.[Arg284fs6*]). Case 4, a 33-year-old male with X-linked RP, had a VUS in RPGR (c.*78_*3782delins13). Functional validation is being conducted. Case 5, a 22-year-old female with MVD, had 3 VUSs in IMPG1; 1 maternal (c.563-174T>G) & 2 paternal (c.2045-7421C>T, c.1291+13335G>A). Functional validation is being conducted. Case 6, a 22-year-old female with RP, had compound heterozygous VUSs in HGSNAT (c.1634C>A/p.T545K, c.235-4486T>C). RT-PCR of patient fibroblast RNA confirmed the exonic variant to cause exon 18 skipping & premature stop codon (p.[Ser550Asnfs12*]). Intronic variant causes intronic retention & premature stop codon (p.[Cys78Aspfs16*]). **Conclusions** This study strongly supports the notion that gene-elusive cases may be partially explained by non-coding variants, highlighting the importance of leveraging comprehensive genome analysis for accurate diagnosis.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2050F: Investigating the genetic and phenotypic landscape of ectodermal dysplasia

Authors:

C. Munderloh[1], R. Belanger Deloge[1], H. Du[1], J. Rosenfeld[1,2], R. Lewis[1], J. Posey[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics Lab., Houston, TX

Abstract:

Ectodermal dysplasias (EDs) encompass a diverse group of inherited disorders that demonstrate substantial genetic and phenotypic heterogeneity. Currently, 49 EDs have a known molecular etiology involving genes encoding critical components of the *TP63*, *WNT*, and *EDA* developmental pathways (~65%), genes in other varied pathways (~21%), and genes encoding structural framework proteins (~14%). The molecular etiologies of the remaining 40 clinically defined EDs are unknown. We recognize that many other ED subtypes and syndromes meet the criteria for an ED clinical diagnosis through phenotypic presentations of two or more affected ectodermal derivatives. Thus, a full characterization of ED's genetic and phenotypic spectrum is needed urgently to better guide diagnostic testing. In this study to date, we have recruited a cohort of 196 individuals in 48 ED families with observed and suspected modes of inheritance consistent with previously described autosomal dominant, autosomal recessive, and X-linked forms of ED. We collected blood samples from 196 individuals, of which 68 are affected and 128 are unaffected. Exome and genome sequencing data are being generated, and family-based rare variant analyses performed to identify potentially deleterious variants in both established and/or candidate novel ED genes. Analyses of the first eight genomes have revealed two notable findings in known ED disease genes. These include the identification of a pathogenic variant in the gene *TP63* (NM_003722.5:c.953G>A (p.R318H)) and a likely pathogenic variant in the gene *KDF1*, which has a role in epidermis formation (NM_152365.3:c.795C>A (p.D265E)). *KFD1* variants are associated with phenotypes ranging from non-syndromic tooth agenesis to ectodermal dysplasia. We are building an additional cohort to investigate genotype-phenotype correlations in *KDF1* disease biology. Future directions include rare variant and further discovery analyses in the remaining 40 ED families, candidate gene pathogenicity testing, and phenotypic clustering analyses to reveal the phenotypes that drive ED subtypes, identify potential phenotypic expansions, and inform genotype-phenotype correlations.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2051F: Investigating the genetic etiologies underlying septo-optic dysplasia (SOD)

Authors:

S. Beheshti[1], A. Jolly[1], H. Du[1], J. A. Rosenfeld[1], J. R. Lupski[1,2], R. A. Lewis[1,2], J. E. Posey[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX

Abstract:

Septo-optic dysplasia (SOD) is a rare congenital disorder with an estimated incidence of 1 in 10,000 newborns. SOD is characterized by cerebral midline anomalies, and its classic triad of features includes optic nerve hypoplasia (ONH), agenesis of the septum pellucidum or corpus callosum, and hypoplasia of the hypothalamo-pituitary axis (HPA). The presence of two or more components of the classic triad is considered "diagnostic" for SOD. While ONH is the most common feature in SOD, the condition shows substantial phenotypic heterogeneity with additional systemic abnormalities like neurodevelopmental disorders and structural brain abnormalities. Several transcription factor genes, including *HESX1*, *SOX2/SOX3*, *OTX2*, and *PAX6*, have been identified to play crucial roles in the embryonic development of the eyes, optic nerves, and pituitary gland, thereby influencing the expression of SOD. These genes explain only 10% of SOD cases, leaving many families without a molecular diagnosis. The genetic heterogeneity and varying severity of phenotypes pose significant diagnostic challenges in SOD and emphasize the need to establish SOD phenotypes as a spectrum and to study gene-disease correlations. Through the Baylor Hopkins Center for Mendelian Genomics (BHCMG) and Baylor College of Medicine Genomic Research to Elucidate the Genetics of Rare (BCM-GREGoR) databases, we have identified 35 unrelated SOD probands and their families with extant exome sequencing (ES) data. Initial analysis of phenotypes demonstrated that 77% of cases have midline brain abnormalities and ONH without HPA. Using a rare variant, family-based analysis approach leveraging existing genomic knowledge sources, we have identified putatively damaging, rare variants (MAF ≤ 0.1%) predicted to impact protein function in genes that demonstrate biological roles relevant to the SOD phenotypes (*KIF26A, NFIA, TUBB2B*, *SLCO1A2,* and *DOCK4*). These early molecular findings will encourage efforts to refine SOD clinical diagnoses and foment understanding of the biological underpinnings of this developmental disorder.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2052F: Leveraging large scale biobank data to enrich phenotypic understanding of Mendelian disorders

Authors:

A. Scalici[1,2], T. Miller-Fleming[1,2], E. Brokamp[1,2], M. Shuey[1,2], E. Knapik[1,2], N. Cox[1,2]; [1]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Monogenic disorders are phenotypically heterogeneous and clinically present with phenotypes that span multiple organ systems. Deficiency in extracellular matrix (ECM) proteins can often result in Mendelian disorders and be caused by variation in genes encoding extracellular matrix (ECM) proteins (COL2A1, COL1A1) and proteins in the secretory pathway that traffic collagen to the ECM (RIC1, SEC23A, SEC24D, SEC13, SEC31 SAR1B). While databases such as OMIM catalogue these gene-phenotype associations, they can often be incomplete due to differential diagnoses or unknown gene function. Because Mendelian disorders are associated with rare variation in these genes, we are often have insufficient statistical power and small sample sizes in population-based studies such as genome-wide association studies (GWAS). We implemented an approach that leverages electronic health records (EHR), genetically predicted gene expression (GPGE) data from an EHR-linked biobank, BioVU, and known Mendelian disorders cataloged in OMIM to gain greater phenotypic understanding of Mendelian disorders associated with ECM deficiency and the protein trafficking, by aggregating GPGE across biologically related genes into a multi-gene score (MGS), representative of a weighted sum of GPGE for an individual. We constructed MGS for trafficking genes, ECM genes, and ECM and trafficking genes combined. Using this MGS in a PheWAS will increase statistical power to detect phenome and identify distinct phenotypic patterns between ECM and trafficking genes. We demonstrated that our MGS increases power especially for neurological phenotypes compared to single gene analyses. Additionally, we found distinct phenome associated with our ECM score such as dementia compared to the trafficking score such as schizophrenia, and addiction related phenotypes. This novel application of a transcriptomic based score to a group of related genes has the potential to expand phenotypic associations and target phenotypes to functionally validate in animal models.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2053F: Low concordance of variant effect predictors to determine the pathogenicity of genetic variants in core immunity genes

Authors:

E. Carroll[1], T. Bergquist[2], V. Pejaver[1], **S. Asgari**[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Inst. for Genomic Hlth., Icahn Sch. of Med., Kent, WA

Abstract:

Background: Primary immunodeficiencies (PIDs, aka Inborn Errors of Immunity) are a group of >450 rare, chronic conditions that affect the core functions of the immune system and have high mortality and morbidity. Diagnosis of PIDs using routine lab tests at birth can be challenging due to the dynamic and evolving immune system of newborns. genetic testing can help with early diagnosis, however for this approach to be effective we need to be able to predict the clinical consequence of identified variants. To this end, we aim to assess the utility of existing tools to assess the clinical impact of genetic variants in PID genes. **Methods:** We used WGS data from 4,150 healthy adults from the 1000 Genomes and the Human Genome Diversity Projects. We used GATK and VEP to select and annotate genetic variants overlapping PID 450 genes. Next, we investigated gene-level characteristics (e.g. resistance to disrupting mutations) and variant-level metrics (e.g. predicted consequence and allele frequency) followed by a thorough comparison of predicted variant effects, clinical consequences, and severity using 8 variant annotation tools plus ClinVar. For variants with multiple consequences, we limited our analyses to the MANE Select followed by the Canonical transcript, or the most severe outcome if the MANE transcript was unavailable. **Results:** In total, 1,688,213 variants overlapped PID genes. We observed a high correlation between allele frequencies in our cohort and gnomAD (r=0.99). Only 2.4% were common in all populations and 94.8% of the variants were rare (minor allele frequency < 5%) and specific to 1 or 2 populations. African, East Asian, and South Asian populations had the largest proportion of population-specific variants (22%, 14%, and 13% respectively). 1.35% of the variants were missense; however, their predicted effects varied widely (mean AlphaMissense score=0.237±0.246). The 8 pathogenicity predictors we used showed low concordance in determining the most deleterious variants in our dataset with none of the top 2% of highest-ranked variants in each method overlapping with the top 2% most pathogenic of other methods. This lack of concordance was even more pronounced when we looked ClinVar "variants of uncertain significance" (VUS, 25% of all variants) with no overlap among the top 5% of each effect predictor. As expected, PID genes showed a low tolerance to putative loss-of-function (LoF) variants on average but wide variations were seen among them (mean LoF tolerance Z-score=2.93 ±2.13). **Conclusion:** Variant effect prediction models for PID genes need to be improved and better calibrated in order to include these genes in newborn screening programs.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2054F: Management of inborn errors of metabolism in Ecuador.

Authors:

V. Romero[1], J. Pozo[2]; [1]Univ. San Francisco de Quito, Quito, Ecuador, [2]Univ. de Cuenca, Cuenca, Ecuador

Abstract:

Inborn errors of metabolism (IEM) are genetic disorders that can cause the accumulation of toxic substances and result in harmful health consequences if left untreated. In developed countries, genetic testing and counseling services for EIMs exist, but developing countries often lack resources and professionals to provide these services.In Ecuador, only 4 conditions are in neonatal metabolic screening and the rest of the individuals are not studied. We've looked at phenylketonuria, which can cause severe mental disability and life-threatening seizures if left untreated. While PKU is detected in Ecuador through national neonatal screening, genetic testing is not available for patients and their families. We investigated this condition and found that in Ecuador they found unique allele frequencies due to the country's mestizo ancestry. The study will be the first to correlate the PKU variant with the ancestry of affected individuals in the region, being highly relevant to other Latin American populations.Another disease we study is Wilson's disease. The clinical characteristics of Wilson's disease vary from asymptomatic patients to those with symptoms related to hepatic, neurological and/or psychiatric involvement. At the time of diagnosis, most patients present with some degree of liver disease, ranging from asymptomatic abnormalities in liver biochemistry to cirrhosis or acute liver injury. In addition, neurological symptoms, such as dysarthria and movement disorders, as well as nonspecific psychiatric symptoms, may be observed.
Wilson's disease commonly occurs between the ages of 3 and 55, although it has also been diagnosed in younger and older patients. The clinical presentation ranges from asymptomatic abnormalities to severe manifestations. We found patients with possible Wilson's disease in the southern region of Ecuador, we ruled out some patients with similar conditions and found new variants.
Although both are EIMs, they are not managed effectively, which can lead to complications of and from the treatment that should be considered as a public health policy.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2055F: Mental health conditions, physical functioning, and health-related quality of life in adults with OI: a cross-sectional study

Authors:

E. Carter[1], C. Derocher[1], J. Mulcrone[1], C. L. Raggio[2]; [1]Hosp. for Special Surgery, New York, NY, [2]Hosp. for Special Surgery, New York City, NY

Abstract:

Osteogenesis imperfecta (OI) is a group of heritable connective tissue disorders characterized by brittle bones, short stature, blue sclera, cardiopulmonary problems, dental problems, and hearing loss. Clinical care focuses on managing physical aspects of the disorder, and existing literature presents mixed conclusions regarding the degree of psychological stress and pain in adults. This study attempts to conduct a thorough analysis of the health-related quality of life for adults with OI by studying responses to 7 questionnaires covering pain, grit, depression, anxiety, physical function and disability, and overall physical and mental wellbeing. This IRB-approved, prospective, cross-sectional study included 26 adults (23 F, 3 M, 18+ years) with an OI diagnosis seen between May 2021-October 2023. Demographics included education level, employment status, surgical history, psychiatric history, and assistive device use. Questionnaires included the RAND 36-Item Short Form Health Survey (SF-36), Patient-Reported Outcomes Measurement Information System 29+2 Profile v2.1 (PROMIS-29), Pain Catastrophizing Scale (PCS), 7-item Generalized Anxiety Disorder questionnaire (GAD-7), Patient Health Questionnaire Depression Scale (PHQ-8), Stanford Health Assessment Questionnaire Disability Index (HAQ-DI), and the 12-item Grit Scale. Statistical analysis was conducted using SPSS 29 for Windows (IBM Corp., Armonk, NY, USA). Mean OI questionnaire scores were compared to the general population and between OI types. Significance was set at $p<0.05$. Participants with OI reported lower scores in physical functioning and health domains than the general population ($p<0.001$ on SF-36, PROMIS, and HAQ-DI). Those with the most severe form (type III OI) had a higher mean HAQ-DI score than types I or IV OI ($p<0.001$). Mental health-related domains were not lower (SF-36 and PROMIS-29), and the percentage of OI participants with clinically-significant anxiety or depression was not different than in the general population. The mean pain intensity (scale of 1-10) experienced over the last four weeks was 4.0 (SD=2.6), with the lower extremities (46.2%), lower back (34.6%), and neck (30.8%) as most common sites. Pain catastrophizing was not different between the OI cohort and the general population, and individuals with OI had higher Grit scores ($p<0.001$). These results give us a broader understanding of health-related quality of life in adults with OI, and should inform future studies aimed at improving overall care. It is important for health teams to use a comprehensive approach to care that recognizes both physical and mental aspects of living with OI.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2057F: Novel compound heterozygote variants in *AARS1* cause developmental and epileptic encephalopathy in a Malian family

Authors:

A. Camara[1], A. Maiga[2], S. Bamba[3,4], O. Traore[3], M. Goita[5], S. Diarra[6,4], A. Yalcouyé[7,1], C. Guinto[1], W. Ji[8], M. Khokha[8], S. Lakhani[8], G. Landoure[9]; [1]Faculty of Medecine and Dendistry, Univ. of Sci., Techniques and Technologies of Bamako (USTTB), Bamako, Mali, [2]USTT-B, Bamako, Mali, [3]USTTB, Bamako, Mali, [4]Pediatric Genomics Discovery Program, Dept. of Pediatrics, Yale Univ., New Haven, Connecticut, USA, New haven, CT, [5]Univ. of Sci. Techniques and Technologies of Bamako (USTTB), Bamako, Mali, [6]USTTB and Yale Univ., Bamako, Mali, [7]JHU, USA, Baltimore, MD, [8]Yale Univ., New Haven, CT, [9]Université des Sci., des Techniques et des Technologuies de Bamako, Bamako, Mali

Abstract:

Introduction: Developmental and epileptic encephalopathy (DEE) is a severe neurological disorder characterised by seizures, developmental delay, and progressive cognitive impairment. Its genetic basis remains largely unelucidated, especially in Africa. However, recent advances in genetic technologies have provided new opportunities for gene identification in Africa. We report novel compound heterozygous variants in the *AARS1* gene associated with DEE in a Malian family. **Method:** The study was conducted with ethical approval and consent from participants. Comprehensive evaluations including clinical examination, blood chemistries, EEG, and brain imaging were performed. DNA was extracted for genetic analysis including Exome sequencing (ES) and Sanger sequencing. *In slico* analyses were performed to confirm deleteriousness of putative variants. **Results:** Two male siblings from a non-consanguineous family of 9 and 22 years old were enrolled for early-onset epileptic seizures and delayed motor development. The ages of onset were 7 and 3 years, respectively. Neurological examination for the proband showed symptoms consistent with attention deficit, mood disorder and aggressiveness, while his brother presented with comprehension and speech disorder, spasticity and hypertonia in all four limbs, spastic gait, and the Babinski sign. EEG showed patterns of cerebral palsy in the proband. MRI shows pronounced cerebellar atrophy. ES revealed a novel compound heterozygous variant in the AARS1 gene (c.2470_2472del; Lys824del and c.2080G>A; Val694Met). This variant segregates with the disease status in the family and was showed to be deleterious by several in slico prediction tools.**Conclusion:** We report a Malian family presenting with DEE caused by a novel variant in AARS1, expanding the genetic epidemiology of this

condition in Africa. Our study further highlights the importance of including populations with diverse genetic background in the global variant databases.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2058F: Novel genetic findings in Malian families with tuberous sclerosis complex.

Authors:

R. Kane[1], A. Toure[1,2], S. Bamba[3,4], K. Dembele[5], W. Ji[6], A. Maiga[7], A. Kone[1,2], S. Diarra[8], A. Yalcouyé[9], A. Fomba[2], O. Traore[3], C. Cisse[10], M. Goita[11], C. Guinto[1,12], S. Lakhani[6], M. Khokha[6], G. Landoure[13,12]; [1]Faculté de Médecine et d'Odontostomatologie, Univ. of Sci., Techniques and Technologies of Bamako (USTTB), Bamako, Mali, [2]Département de Pédiatrie, Ctr. Hosp.ier Univ.ire « Gabriel Touré », Bamako, Mali, Bamako, Mali, [3]USTTB, Bamako, Mali, [4]Pediatric Genomics Discovery Program, Dept. of Pediatrics, Yale Univ., New Haven, Connecticut, USA, New Haven, CT, [5]Hôpital du District de la Commune IV, Bamako, Mali, Bamako, Mali, [6]Yale Univ., New Haven, CT, [7]USTT-B, Bamako, Mali, [8]USTTB and Yale Univ., Bamako, Mali, [9]JHU, USA, Bamako, Mali, [10]Univ. of Sci. Techniques and Technology, Bamako, Mali, [11]Univ. of Sci. Techniques and Technologies of Bamako (USTTB), Bamako, Mali, [12]Service de Neurologie, Ctr. Hosp.ier Univ.ire Point G, Bamako, Mali, Bamako, Mali, [13]Université des Sci., des Techniques et des Technologuies de Bamako, Bamako, Mali

Abstract:

Introduction: Tuberous Sclerosis Complex (TSC) is a mTORopathy, with autosomic dominant transmission, caused by variants in TSC1 and TSC2 genes. It affects multiple organs, primarily characterized by the development of hamartomas, preferentially in the brain. Despite its well-established clinical profile, prevalence and genetic characteristics of TSC remain underreported in sub-Saharan Africa. This study presents the first molecularly diagnosed cases of TSC in Malian families.**Methodology**: Clinical assessment was performed by neurologists and pediatric neurologists after informed consent. Then, patients underwent imaging including brain CT-scan and MRI, EEG and kidney ultrasounds. DNA was extracted for genetic analysis including exome sequencing (ES) and Sanger sequencing (SS). Prediction tools were used to assess deleteriousness of putative variants.**Results**: Six patients, from four non-consanguineous families, were enrolled in the study. Pedigree analysis suggested an autosomal recessive inheritance and sporadic cases with onset ages ranging from 1 to 12 months. Clinical examination revealed attention deficit and hyperactivity disorder, impulsivity, facial angiofibroma, hypochromic, and fibrous cephalic tags in all patients. Brain imaging showed

subependymal nodules (SEN) and encephalic dysplasia in one patient, and cortical tubers and SEN in four patients. EEG found hypsarrhythmia in three patients, and diffuse right spike waves with slow delta activities in two. Kidney ultrasound showed angiomyolipoma with normal function in three patients. Exome sequencing identified various variants in three families in the TSC2 gene. Notably, a novel heterozygous frameshift variant (c.1095_1096ins CTGATGTTCATCATCATGCT; p. E366Lfs*30) was found in all three patients from family 1. Neither parent carried the variant, suggesting germline mosaicism. Additionally, a known de novo heterozygous pathogenic splicing variant (c.2245-1G>A) was detected, along with two nonsense variants (c.4375C>T; p. Arg1459Ter and c.598C>T; p. Gln200Ter), each present in one patient from different families. These variants co-segregate with disease status in the respective families and were shown to be deleterious by several in silico prediction tools.**Conclusion**: We report novel and known variants in the TSC2 gene causing TSC in the Malian population, expanding the global epidemiology of the disease. Larger cohort studies may uncover new TSC related genes and/or variants in understudied populations such as Mali. Cell and/or animal-based studies may provide more insights in the pathogenicity and therapeutics.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2059F: Novel germline mutation in *BCORL1* gene identified in a family with three Shukla-Vernon syndrome patients

Authors:

A. Alkhateeb; Jordan Univ. of Sci. and Technology, Irbid, Jordan

Abstract:

Germline variants in *BCORL1* have been proposed to underlie an X-linked disorder of intellectual disability, dysmorphic features, and behavioral abnormalities, annotated in OMIM as Shukla-Vernon syndrome (OMIM #301029). All reported variants are inherited from mothers who were either unaffected or had mild signs, such as learning disability. Here, we report on a family with three sons having seizures, intellectual disability, learning difficulties, ataxia, abnormal vision, and abnormal gait. We carried out whole exome sequence analysis of all protein coding genes in the genome for the proband and affected/unaffected family members, coupled with whole exome deletion/duplication (CNV) Analysis. All patients were found to be hemizygous for BCORL1 c.3366C>T (NM_021946.4), p.(Gly1122=), which is a variant of uncertain significance (VUS). The variant is inherited from the patient's mother, who is heterozygous for the variant. Sanger sequencing confirmed the variant. No other candidate variants were found. This variant is absent in gnomAD, a large reference population database (n>120,000 exomes and >15,000

genomes). *BCORL1* c.3366C>T, p.(Gly1122=) is a synonymous variant that does not alter the protein sequence, however, in silico splice prediction tools (SSF, MaxEntScan, NNSPLICE) predict this substitution will create a new cryptic donor splice site. The variant may therefore lead to abnormal splicing. Transcriptional analysis is underway to determine this variant's effect on splicing. To the best of our knowledge, this variant has not been described in the medical literature or reported in disease-related variation databases such as ClinVar or HGMD.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2060F: Novel pathogenic variant of *TBCK* reveals broadened phenotypic spectrum of rare TBCK syndrome

Authors:

S. Jenkins[1], N. Arvin[1], A. L. Shillington[2], E. D. Sperry[2], G. Hizaji[2], R. J. Hopkin[2], Y. Wu[2], C. Nagaraj Bhimarao[2]; [1]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [2]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Abstract:

TBCK syndrome is characterized by infantile hypotonia with psychomotor retardation and characteristic facies 3, and is an autosomal recessive condition characterized by mutation(s) in the *TBCK* gene. The condition presents with severe neurological deficits, is caused by biallelic pathogenic variants in the *TBCK* gene, and presents with pathogenic variants implicated in the mTOR pathway. The condition presents with developmental delay, impaired motor functions., speech, seizures, and decreased reflexes. Due to a low number of reported cases, the anticipated phenotypic spectrum and diagnostic criteria is not well-defined, nor consistent across all patient cases. Furthermore, novel variants identified in this gene are challenging to classify due to limited data on the pathogenic variants of *TBCK*. Here, we are reporting a novel variant in *TBCK* and will assist in broadening the understanding of the TBCK syndrome spectrum. We will also discuss variants of uncertain significance interpretation and investigate genetic test methodologies implicated during the diagnostic process.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2061F: Novel Strategies and Analytical Methods in Tackling Unresolved Rare Diseases: The Approach of the BCM-GREGoR Program

Authors:

S. Jhangiani[1], F. Sedlazeck[1,2], H. Du[1], R. Duan[1], D. Calame[1], R. Sutton[1,3], M. Dawood[1], M. Mahmoud[1], S. Behra[1], M. Lun[4], J. Bengtsson[4], A. Saad[1], A. Rivera[1], A. Rai[5], C. Munderloh[1], C. Grochowski[1], D. Marafi[1], D. Pehlivan[1], I. Chinn[1,3], N. Gogate[1], P. Kaur[1], M. Jamsandekar[1], M. Gandhi[1], P. Liu[1], R. Lewis[1,3], S. Beheshti[1], B. Garcia[1], Y. Li[1], S. Morris[1,3], T. Bozkurt[5], Z. Dardas[1], Z. Akdemir[5], Z. Khan[1], K. Walker[1], H. Doddapaneni[1], C. Carvalho[4], J. Lupski[1,3], D. Muzny[1], R. Gibbs[1], J. Posey[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Rice Univ., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Pacific Northwest Res. Inst., Seattle, WA, [5]Univ. of Texas Hlth.Sci. Ctr., Houston, TX

Abstract:

Rare diseases impact more than 300 million people globally, with most lacking a molecular diagnosis or effective treatments. The Genomic Research to Elucidate the Genetics of Rare disease program at Baylor (BCM-GREGoR) aims to address these unresolved challenges and has identified critical gaps in knowledge: (1) rare diseases without an established molecular etiology, (2) variants of unknown functional impact, (3) diagnostic utility of newer genomic technologies, (4) effectiveness of newer genome builds in disease diagnostics and (5) complete understanding of molecular mechanisms underlying disease. To resolve this, BCM-GREGoR has enrolled more than 1,700 families and generated sequencing data on 2K rare disease samples, harmonizing these with data from prior rare disease research programs, i.e. BHCMG. We continue to study allelic series in genes such as *FLVCR1* and *ASTN1*, the former for which we observed a phenotypic spectrum ranging from adult-onset retinitis pigmentosa to severe neurodevelopmental disorder. Deep phenotyping coupled with HPO analysis informed investigation of clinical subtypes of disease. Focused application of functional studies further uncovered molecular mechanisms leading to distinct clinical disease subtypes, providing a foundation for therapeutic strategies. We have further studied techniques such as MAVE, which resolve variant functional impacts, to reduce VUS inequities impacting under-represented populations. We employ targeted and whole genome sequencing using both short and long reads, as well as RNA sequencing and have developed cost efficient computational pipelines. In-house developed tools such as Sniffles2, HMZDupFinder and VizCNV aid in comprehensive data analysis and have led to candidate novel gene discoveries, such as a homozygous duplication found in *PAAF1*. We are also testing newer reference builds, i.e. CHM13, and performing a liftover to GRCh38 to maximize annotation resources. As we build cohorts from unique populations such as Turkey and Middle East/Northern Africa we gain a deeper understanding of

multilocus pathogenic variation, and we can investigate population substructure and contribution of identity-by-descent to rare disease. Data sharing is paramount and quarterly submissions to AnVIL encompass robust sample metadata and sequencing data, and the consortium maintains an open-access variant browser at https://variants.gregorconsortium.org/. Each discovery has showcased the diagnostic potential of our combined approaches and provided valuable insights into the genomics of disease, which continue to fuel investigations into undiagnosed cases and the pathobiology of disease.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2062F: Rare coding variants in epilepsy: A study of 9,205 individuals using NGS dataset. ⭐

Authors:

S. Ganesan[1], M. Torres Diz[2], A. Malakar[3], J. Miller[4], S. Haag[5], R. Rajagopalan[6], L. Conlin[6], I. Helbig[6]; [1]Children s Hosp. of Philadelphia, Chalfont, PA, [2]The Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Childrens Hosp. of Philadelphia, Philadelphia, NJ, [4]The Children S Hosp. of Philadelphia, Philadelphia, PA, [5]CHOP, Chalfont, PA, [6]Children s Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Next-generation sequencing (NGS) has been pivotal in identification of novel genes associated with epilepsies and neurodevelopmental disorders. Here, we leveraged NGS data collected and harmonized by the Arcus program for all individuals across the Children's Hospital of Philadelphia, alongside clinical information available extracted from electronic medical records, which also consists of diagnostic data repurposed for secondary research use. We identified 1,243 individuals (cases) diagnosed with an epilepsy-related ICD10 code (G40, G41, R56, or P90) from a cohort of 9,205 individuals (controls). We aimed to identify if rare variants in established epilepsy genes have a significant effect on clinical features, potentially highlighting novel genotype-phenotype associations. We focused on variants in genes previously associated with epilepsy. Variants with an MAF less than 0.01% in the gnomAD database were selected. Applying these filters, we discovered protein-truncating variants (PTVs) in 87 genes, with variants in *DYNC1H1 (n=6), and SYNGAP1 (n=5)*, and missense variants in 292 genes, including *RELN (n=74)*, *CACNA1A* (n=42) and *DEPDC5* (n=35). Although our initial analysis with variant classes did not reveal any significant associations, it may suggest that the relationship between rare variants and epilepsy phenotypes is complex and may require more detailed investigation. Future analyses will focus on evaluating whether individuals with rare variants in

the same epilepsy genes exhibit similar clinical presentations. This could improve the clinical interpretation of NGS data and aid in natural history studies and precision medicine trials. Overall, our analyses indicate importance of continuous exploration and validation of genotype-phenotype associations to enhance clinical relevance of genomic data in epilepsy. These results have the potential for improving the clinical interpretation of variants from NGS data (exome and genome sequencing) and help with natural history studies and precision medicine trials.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2063F: SATB2 Variants in Glass Syndrome: Discussing the Clinical Impact of Novel Variants in the context of Misleading Diagnoses.

Authors:

M. Thompson[1], E. Ehmke[2], D. Horn[2], P. Krawitz[3], K. Alexej[3]; [1]Univ. of Toronto, Toronto, OH, Canada, [2]Charité - Univ.smedizin, Berlin, Germany, [3]Inst. for Genomic Statistics and Bioinformatics, Univ. of Bonn, Bonn, Germany

Abstract:

Background/Objectives: We present a discussion of two patients with *SATB2* variants that may be associated with SATB2 associated syndrome (SAS), or Glass syndrome, for whom Mabry syndrome was initially considered in the differential diagnosis. SAS (MIM 612313), which entails multisystemic developmental delay with dysmorphic features, can be distinguished from Mabry syndrome (MIM 239300) because SAS lacks hyperphosphatasia necessary to diagnose Mabry syndrome. While one patient with an *SATB2* variant displaying Glass syndrome features was initially considered for Mabry syndrome due to concurrent hyperphosphatasia, a second patient, with a potential Glass syndrome, revealed a *de novo* SATB2(NM_001172517.1):c.1733C>T (p.Pro578Leu) mutation of unknown significance (VUS). We ruled out the differential diagnosis of Mabry syndrome because hyperphosphatasia was absent and there were no likely biallelic pathogenic variants in the GPI biosynthesis or remodeling genes.
Methods: In the first patient, we identified the SATB2(NM_001172517.1):c.400del; (p.Ala134HisfsTer17) variant, suggesting autosomal dominant Glass syndrome. No likely biallelic pathogenic variants were identified by exon sequencing of GPI biosynthesis and remodeling pathway genes. In the second patient we identified a likely *de novo* SATB2(NM_001172517.1):c.1733C>T (p.Pro578Leu) variant.
Results: While the first patient exhibited features of Glass syndrome in the absence of identified biallelic pathogenic variants, the second patient, with additional features including seizures,

arachnodactyly, and a happy demeanor, presented with a phenotype that was distinct from his brother with Mabry syndrome.
Conclusion: Our preliminary analysis, suggests that the first patient's Glass syndrome is linked to SATB2 variant. The second *de novo* mutation identified in the second mutation requires further investigation. We discuss phenotypic features, including alkaline phosphatase levels, facial dysmorphology, bone density, and metabolic profile, that can be used to distinguish Glass syndrome and Mabry syndrome and facilitate diagnosis.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2065F: Severe hemolytic anemia, seizures, and respiratory failure in a patient with triosephosphate isomerase deficiency: a case report

Authors:

M. Weisz-Hubshman[1,2], A. J. Williams[1,2], V. C. Rossi[1,2], E. Bland[1,2], E. Mizerik[1,2], X. Luo[3], P. R. Hillman[2], K. R. Shields[4], F. Scaglia[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Baylor Genetics, HOUSTON, TX, [4]Univ. of Texas Hlth.Sci., Houston, TX

Abstract:

Triosephosphate isomerase (TPI) is a ubiquitously expressed enzyme encoded by the *TPI1* gene. It catalyzes the interconversion of the triose phosphate isomers dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate in the fifth step of glycolysis. Given its importance, this enzyme is found in almost every organism, from bacteria to humans. Global knock out of *Tpi1* is embryonic lethal in mice. Partial TPI deficiency in humans with residual enzymatic activity (less than 50% but more than 0%) is compatible with life but causes autosomal recessive hemolytic anemia due to triphosphate isomerase deficiency (MIM#615512). As an essential glycolytic enzyme, severe hemolytic anemia is a universal finding. Additionally, these individuals develop a life-limiting neuromuscular phenotype as well as increased susceptibility to infections. This often results in respiratory failure that leads to ventilator dependency. This neuromuscular phenotype was recently recapitulated in *Tpi1*-deficient mice, but the etiology is unclear. Here we present a 2-month-old proband with failure to thrive, respiratory failure, seizures, and severe hemolytic anemia, who passed away at 3 months of age. Trio whole genome sequencing uncovered compound heterozygous variants in *TPI1* consistent with a TPI deficiency. Our proband was compound heterozygous for the previously described pathogenic c.315G>C (p.Glu150Asp) variant and a likely pathogenic c.324+1G>C splice variant. The p.Glu150Asp variant accounts for 80% of patients with TPI deficiency. The second variant was previously unreported; c.324+1G>C results in a splice donor loss, possibly leading to nonsense mediated

decay. A recent study showed a similar phenotype including hemolytic anemia, neuromuscular dysfunction, possible seizures, and markedly reduced lifespan in a mouse model with a *Tpi1*$^{E105D/null}$ genotype, which is associated with partial residual enzymatic function. Given the generally fatal nature of this condition, treatment options are needed. Currently, treatments are experimental. Ketogenic diet was effective in treating seizures in a TPI mutant *Drosophila,* known as *TPI*^{sugarkill}*,* although clinical data is lacking. Additionally, bone marrow transplant has been shown to improve the hematologic phenotype in mice and has been done in an isolated number of patients, although no data indicates this could improve or reverse the neuromuscular phenotype. While there are no proven therapies available at this time, we hope this case and others will lead into future therapeutic development like gene therapy or enzyme replacement.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2066F: Siblings with Okur-Chung syndrome (OCS) due to a pathogenic mutation of *CSNK2A1* gene on chromosome 20p13: p.(R21*), and a 51 Kb duplication at chromosome Xp11.4 with variable expression.

Authors:

M. Hajianpour[1], J. Lundberg[1], C. DeJesus[1], C. Feuerstein-Simon[2], D. Sima[3]; [1]Albany Med. Coll., Albany Med Hlth.System, Div. of Med. Genetics and Genomics, Albany, NY, [2]Albany Med. Coll., Albany Med Hlth.System, Dept. of Pediatrics, Albany, NY, [3]Albany Med. Coll., Albany Med Hlth.System, Div. of Pediatric Endocrinology, Albany, NY

Abstract:

We present two brothers (17 & 12 yo) with **Okur-Chung syndrome (OCS**) due to a pathogenic mutation of ***CSNK2A1*** gene on chromosome 20p13: **c.61 C>T [p.(R21*)]** by **WES**. Both have a **51 kb duplication at Xp11.4** => arr [GRCH37] Xp11.4(40440222_40491329)x2 and maybe alcohol-related neurodevelopmental disorder **(ARNDD)**. Their 11 yo brother does not have the CSNK2A1 variant or Xp11.4 dup, but may have ARNDD. All 3 lived in orphanages (home country of **Kyrgyzstan**), witnessing and being victims of violence. They were recently adopted by foster parents in USA.
Sib 1: 17 yo, Ht & Wt<3rd%ile, HC: 40%ile, DD/LD, FSIQ 80, not communicative (quiet). Dysmorphic features: Plagiocephaly (mild-Left), widow's peak, narrow palpebral fissures, hypoplastic helices (mild), high nasal bridge, high arched palate (mild), facial asymmetry (L<R),

deviation of the mouth to the right, poor dentition, retrognathia (mild), and pes planus. He is at 10th grade, inability to finish schoolwork (partly related to the language barrier).
Sib 2: 12 yo (more severely affected), Ht & Wt<3rd%ile, HC: 45th%ile, DD/LD, FSIQ 76, behavioral problems with aggression, food seeking behavior ("eats from garbage can"), GH deficiency (Short, delayed bone age: on GH), strabismus, hypertension, "renal disease, mitral valve prolapse," and constipation.
Dysmorphic features: Dolichocephaly, epicanthic folds, short/ narrow palpebral fissures, floppy ears, hypoplastic helices, broad nasal bridge & tip, full lips, dental crowding, retrognathia, facial asymmetry (L<R), extra nipples (x2), hypoplastic nipples, pectus carinatum, shawl scrotum, hypermobile fingers, and Pes planus. He is in 6th grade, special program with accommodations and modifications.
Sib 3: 11yo, **no** CSNK2A1 mutation, **no** Xp11.4 dup, Ht: 17th%ile, Wt: 20th%ile, HC: 50th%ile, ADD/LD, IQ: WNL, anxiety, frequent bowl movement, emotional disorder (ARNDD?).
Dysmorphic Features: Dolichocephaly, epicanthic folds, wide nasal bridge & tip, anteverted nares, irregular teeth, retrognathia, hypoplastic nipples, hypermobile fingers and left knee.
Clinical features in first two siblings could be mostly related to OCS, and partly to Xp11.4 dup, and ARNDD, and in unaffected sib to ARNDD. Some facial features are ethnicity-related.
The **OCS** is characterized by delayed psychomotor development, intellectual disability with poor speech, behavioral abnormalities, and variable dysmorphic facial features (microcephaly, cortical abn., GI issues, low Ig in some).
The **Xp11.4 dup** was interpreted by the Lab as a VUS, but two brothers with Xp11.4 dup with developmental delay/ID, autism, and behavioral problems are published (So, it may contribute to the phenotype).

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2067F: Tandem repeats in genes associated with synaptic functions are frequently expanded in individuals with suspected genetic epilepsies ⭐

Authors:

J. Nguyen[1,2], A. Mitina[1], Y. Yue[1], W. Engchuan[1,3], B. Haque[1], G. Costain[1,2,4], R. Yuen[1,2]; [1]Genetics and Genome Biology, The Hosp. for Sick Children, Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]The Ctr. of Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [4]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, Toronto, ON, Canada

Abstract:

Introduction: Epilepsy is a neurological disorder characterized by recurrent unprovoked seizures, with genetic factors playing a significant role in its development. Genome sequencing (GS) captures most genetic variations, but >50% of individuals with suspected genetic epilepsies remain undiagnosed after clinical GS. GS allows the detection of tandem repeat expansions (TREs), a type of genetic variation known to be associated with over 60 human monogenic disorders, including eight types of epilepsy. TREs are also involved in other complex brain disorders such as autism spectrum disorder and schizophrenia. The extent to which genome-wide TREs contribute to epilepsy is unclear. Methods: We analyzed GS data to identify rare (<0.1% population frequency) TREs in 1,465 individuals with epilepsy from Genomics England. Subsequently, we performed a logistic regression analysis to compare the burden of rare TREs between the epilepsy cohort and a cohort of 386 unaffected individuals. To identify cis-regulatory elements and functional pathways impacted by TRE-associated genes, we performed a burden analysis on ChIP-seq data from human postmortem brains and a gene set enrichment analysis using Gene Ontology terms. Results: We identified 490 rare TREs in 182 distinct genic regions, including 10 known TRE disease-associated loci such as *STARD7* and *MARCHF6*. We found that rare TREs are enriched in the genic regions in individuals with epilepsy compared to unaffected individuals (odds ratio (OR) = 1.88, $p = 1.6 \times 10^{-7}$). Specifically, rare TREs are more likely to be found in the 5' untranslated region (OR = 3.1, $p = 2.9 \times 10^{-2}$) and in proximity to the transcription start site (OR = 3.8, $p = 4.0 \times 10^{-2}$). We demonstrated that rare TREs have high GC content ($p < 2.2 \times 10^{-16}$), often overlap with active enhancers in the developing brain (OR = 1.5, $p = 6.4 \times 10^{-3}$), and impact genes involved in synaptic processes (OR = 2.7, false discovery rate = 0.04). Individuals with an earlier age of onset (<2 years) and those without a molecular diagnosis tend to have a higher burden of rare genic TREs. We estimate that enhancer-associated TREs contribute to 3.6% of the risk in epilepsy. Conclusions: Our findings suggest that TREs contribute to the pathogenesis of epilepsy by disrupting active enhancers involved in regulating gene expression during brain development. Furthermore, our results indicate that the contribution of TREs to epilepsy is greater than previously understood. This research was made possible through access to data and findings in the National Genomic Research Library via the Genomics England Research Environment.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2068F: Targeted panel sequencing: Advancing molecular diagnosis for patients with vascular anomalies

Authors:

E. Tan, K. Rafiee, L. Wee, M. Koh; KK Women's & Children's Hosp., Singapore, Singapore

Abstract:

Vascular anomalies are a heterogeneous group of disorders that can be classified as vascular tumors and vascular malformations with overlapping phenotypic features. They are characterized by morphological, structural, and functional alterations of blood and/or lymphatic vessels. The vascular lesions can range from small transient lesions to large overgrowths of blood vessels of varying sizes associated with increased mortality or morbidity. Although most vascular anomalies can be diagnosed based on clinical features, radiological, and histopathological investigations, genetic analysis can provide an accurate diagnosis of more complex conditions for better prognosis and management, especially with the advent of targeted therapies, such as mTOR, PIK3CA and MEK inhibitors.
Approval to conduct the study was obtained from the Institutional Review Board. Venous blood or affected tissue samples were collected from patients with written informed consent. Genomic DNA was extracted with the Qiagen PureGene Kit, QIAmp DNA Mini Kit, or Cell-free Circulating DNA Purification Midi Kit. Sequencing was performed with a SureSelect custom panel and run on a MiSeq Sequencer using the Micro flow cell (v2, 300 cycles).
From the testing of nine cases, we have identified pathogenic/likely pathogenic variants for six patients. There were three variants in *EPHB4*, one in *RASA1*, one in *GNA14*, and one in *PTEN*. Besides one novel *EPHB4* (single-nucleotide duplication) variant, the *GNA14* (missense) and the *RASA1* (single-nucleotide insertion) variants are also novel. Of the three samples that did not have any variants confirmed, one was a cell-free DNA lymphatic fluid sample with no DNA leftover to validate potential variants, while no potential variants were identified for an FFPE sample of poor quality. The third sample was from the affected tissue of a patient with a vascular or lymphatic tumor.
Our preliminary results show good molecular diagnostic yield (66.7%) using our custom NGS panel for this cohort of patients with disorders of vascular anomalies which display a wide spectrum of clinical presentations. Clinical diagnosis can be challenging due to the heterogeneity of presentations and overlapping features with other pediatric conditions, a molecular finding can offer a precise diagnosis of the underlying disorder and ascertain the need for monitoring the development of malignancies and other complications.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2070F: The role of multidisciplinary care for a family without a clear cardiac phenotype identified to have an *SCN5A* variant on exome sequencing

Authors:

A. Pagani[1,2], B. N. Hughes[3], J. Schacht[4], L. Hellwig[3,1,2], M. Haigney[5]; [1]HJF, Rockville, MD, [2]Ctr. for Military Precision Hlth.(CMPH), Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [3]Dept. of Pediatrics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [4]Dept. of Pediatrics, Walter Reed Natl. Military Med. Ctr., Bethesda, MD, [5]Dept. of Med., Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Abstract:

The *SCN5A* gene is associated with a spectrum of autosomal dominant cardiac conditions including, but not limited to, Brugada syndrome (BrS), long QT syndrome (LQTS), dilated cardiomyopathy (DCM), and other arrhythmias or conduction defects. *SCN5A*-related conditions demonstrate reduced penetrance and variable expressivity, and some individuals may even present with more than one *SCN5A*-related cardiac phenotype (*SCN5A* overlap syndromes). Many of the *SCN5A*-related conditions have an increased risk for sudden cardiac death, so appropriate identification and management of individuals at risk for *SCN5A*-related conditions is vital. However, given the variable expressivity of *SCN5A*-related conditions, identification and interpretation of *SCN5A* variants can be challenging, especially in families without a clear cardiac phenotype. We report the case of a family identified to have a rare *SCN5A* variant (c.250G>A, p.Asp84Asn) which has conflicting interpretations of pathogenicity (reported as a variant of uncertain significance and as likely pathogenic). The *SCN5A* variant was first identified on exome sequencing for a 15-year-old male with multiple neurodevelopmental concerns and a history of arrhythmia with first degree atrioventricular (AV) block. A 2020 paper using patch clamp data found that this variant had partial loss of sodium channel function with 32.8% of the peak current density compared to wild type. Previously, the proband had a procainamide challenge that was nondiagnostic, but toward the end of the challenge he had changes on his ECG which may fit with the partial reduction in sodium current suggested by the functional data. There were several other family members with this variant who are asymptomatic or have had non-specific cardiac symptoms, like first degree AV block or a reduced ejection fraction that resolved spontaneously. The proband's paternal grandmother died at age 48 in her sleep and was reportedly identified to have four-chamber dilatation on autopsy, however, this individual did not have genetic testing. Given the variability in variant classification and spectrum of *SCN5A*-related conditions, multidisciplinary cardiogenetics care was critical for this family in determining appropriate recommendations for follow-up and management. For first-degree

family members, we facilitated genetic counseling, targeted variant testing, and a clinical cardiac evaluation, including ECG. For the proband's siblings that tested positive, we recommended at least annual heart rhythm monitoring and ECGs, along with intermittent echocardiograms and adherence to preventive BrS management guidelines, such as prompt treatment of fevers.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2071F: The unusual pattern of hypomethylation at IC1 and IC2 in Beckwith-Wiedemann syndrome and Russell-Silver syndrome

Authors:

S. Cyrus, L. Hasadsri, W. Shen, K. Rumilla; Mayo Clinic, Rochester, MN

Abstract:

Beckwith-Wiedemann syndrome (BWS) and Russell-Silver syndrome (RSS) are imprinting disorders caused by epigenetic or genetic alterations at the 11p15 region. The characteristic methylation pattern of hypomethylation at IC2 and/or hypermethylation at IC1 is indicative of BWS, while hypomethylation at IC1 is indicative of RSS. Rarely, loss of methylation is seen at both IC1 and IC2 in both disorders with few cases described in literature. Herein, we report 8 additional cases referred for testing at the Mayo Clinic.

We identified 7 individuals and 1 prenatal sample with hypomethylation at both IC1 and IC2. When clinical information was provided, clinical features were consistent or suggestive of BWS or RSS. Gene dosage and methylation analysis by methylation specific- multiplex ligation dependent probe amplification (MS-MLPA) was performed on all samples. No deletions or duplications were identified. The levels of hypomethylation at IC1 and IC2 were variable across samples, with most samples exhibiting near complete loss of methylation at one imprinting center, while the other imprinting center had borderline or incomplete hypomethylation. Reported phenotype of samples correlated with the region that had the greater level of hypomethylation.

Additionally recent literature has suggested that hypomethylation of both IC1 and IC2 is associated with a multi locus imprinting defect or disorder (MLID). To determine if other imprinting disturbances were present, we performed our Angelman syndrome (AS)/Prader Willi syndrome (PWS) MS-MLPA assay on two patients with hypomethylation of both IC1 and IC2. An abnormal methylation pattern at the AS/PWS region was identified in one patient who presented with a BWS phenotype, whereas the other patient with an RSS phenotype had normal methylation.

Generally, a genetic test result of hypomethylation of IC1 and IC2 is highly dependent on clinical

correlation for establishing/confirming a BWS/RSS diagnosis. However, our results suggest that the imprinting region exhibiting the higher level of disturbance could be predictive of the main clinical phenotype. Although testing of additional loci may show other imprinted disturbances, the clinical utility and clinical implications of these additional methylation disturbances are yet to be established. Therefore, further studies are required to completely characterize whole genome methylation patterns and the associations with clinical phenotype.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2072F: *UBE2I* mutations unveil the critical role of SUMOylation pathway in syndromic autism

Authors:

V. Pullano[1], S. Trajkova[1], A. Aulino[1], M. Rossi Sebastiano[2], E. Sukarova-Angelovska[3], E. Ortolan[4], H. Sczakiel[5,6], S. Hiatt[7], A. Hurst[8], B. Cogne[9,10], I. Bertrand[9,10], B. Karishma[11], A. Pichler[11,12], T. Pippucci[13], P. Dimartino[14], S. Cardaropoli[15], R. Piva[16], E. Ferrero[1], A. Mussa[15], S. De Rubeis[17], J. Buxbaum[17], G. Ferrero[18], A. Brusco[1]; [1]Univ. of Turin, Dept. of NeuroSci.s Rita Levi-Montalcini, Turin, Italy, [2]Univ. of Turin, Dept. of Molecular Biotechnology and Hlth.Sci., Torino, Italy, [3]Univ. Clinic for Children's Diseases, Faculty of Med., Univ. "Sv. Kiril I Metodij", Dept. of Endocrinology and Genetics, Skopje, North Macedonia, The Republic of, [4]Univ. of Turin, Dept. of Med. Sci., Turin, Italy, [5]Inst. of Med. Genetics and Human Genetics, Charité - Univ.smedizin Berlin, Corporate Member of Freie Univ. Berlin and Humboldt-Univ. zu Berlin, Berlin, Germany, [6]BIH BioMed. Innovation Academy, Clinician Scientist Program, Berlin Inst. of Hlth.at Charité - Univ.smedizin Berlin, Berlin, Germany, [7]HudsonAlpha Inst. for Biotechnology, Huntsville, AL, [8]Univ. of Alabama, Dept. of Genetics, Birmingham, AL, [9]Service de Génétique Médicale, Ctr. Hosp.ier Univ.ire de Nantes, Nantes, France, [10]Université de Nantes, Inserm UMR 1087 / CNRS UMR 6291, Inst. du thorax,, Nantes, France, [11]Inst. of Biochemistry, ETH Zürich, Zurich, Switzerland, [12]Max Planck Inst. of Immunobiology and Epigenetics, Dept. of Epigenetics,, Freiburg, Germany, [13]IRCCS Azienda Ospedaliero-Univ.ria di Bologna, Bologna, Italy, [14]Univ. of Pavia, Dept. of Molecular Med., Pavia, Italy, [15]Univ. of Turin, Dept. of Publ. Hlth.and Pediatrics, Turin, Italy, [16]Univ. of Turin, Dept. of Molecular Biotechnology and Hlth.Sci.,Turin, Turin, Italy, [17]Seaver Autism Ctr. for Res. and Treatment, Icahn Sch. of Med. at Mount Sinai, New York, NY, [18]Univ. of Turin, Dept. of Clinical and biological Sci.s, Turin, Italy

Abstract:

The SUMOylation pathway is a crucial post-translational modification process that regulates the function of hundreds of proteins by SUMO protein attachment. This three-step pathway

requires activation of SUMO proteins by the E1 enzyme, transfer to the E2 conjugating enzyme and SUMO-protein ligation directly or through E3 ligase. It is now emerging that defects in SUMOylation are being increasingly linked to abnormal neuronal development and synaptic function. The focus of this study is the *UBE2I* gene, which encodes the sole human E2 enzyme, also called UBC9; the goal is to evaluate the role of *UBE2I* in neurodevelopment disorders (NDDs). Trio-exome sequencing of a cohort of >600 NDD families led to the identification of a *de novo (dn)* missense variant in UBE2I [NM_194260.3:p.(Asn124Ile)] in a pediatric patient with a syndromic phenotype. In GeneMatcher, we found another three NDD patients harboring *UBE2I* variants (p.Val25Met *dn*; p.Trp53Leu *dn*; p.Glu143Gly). Comparative analysis of the patients' clinical records revealed common features: neurodevelopmental delay, hypertelorism/telecanthus, almond-shaped eyes, thick eyebrows, and hand and nasal anomalies. All variants are absent in GnomAD. The substituted amino acids are evolutionarily conserved from *C.elegans* to human and the changes are predicted to be highly damaging (CADD ≥26). All variants highly destabilize protein structure according to 3D modelling predictions (DDG< –0.5 kcal/mol); in addition, Val25 is known to be part of the RanBP2 E3 ligase binding site. When stable cell lines expressing either WT or variant UBC9-mCherry fusion proteins were analysed by western blot and flow cytometry, expression of the UBC9 variants was found to be reduced by ~50% with respect to WT. Proteasome inhibition significantly restored expression of the UBC9 variant proteins, indicating they likely destabilize protein structure, leading to proteasomal degradation. We are currently assessing the impact of UBE2I downregulation using antisense oligonucleotides in human neuroprogenitor cell lines. In conclusion, we propose that *UBE2I* be considered as a candidate gene for a novel NDD characterized by syndromic autism, a finding that will provide insights into the role of SUMOylation in brain function and development.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2073F: Unlocking the genetic code of cerebral palsy: Exploring clinical and genetic insights in modern era

Authors:

T. Lau[1], Genomics England Research Consortium, SYNaPS Study Group, Queen Square Genomics, C. Steward[2], J. Gecz[3], D. Rowitch[4], D. Murphy[1], R. Kaiyrzhanov[1], S. Efthymiou[1], J. Vandrovcova[1], R. Maroofian[1], H. Houlden[1]; [1]Univ. Coll. London, London, United Kingdom, [2]Genomics England, London, United Kingdom, [3]The Univ. of Adelaide, Adelaide, Australia, [4]Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

Background/Objectives: Cerebral palsy (CP), characterised by non-progressive movement impairments and co-morbidities, was traditionally linked to perinatal environmental risk factors. Recent evidence suggests a genetic component could play a part in the aetiology of this group of conditions, prompting an investigation into the clinical and genetic aspect of individuals with CP. **Methods:** Screened the Genomics England database (GEL, n=90,178) and the Neurogenetics Lab database at UCL Queen Square Institute of Neurology (UCL-ION, n=~43,000) to identify individuals with clinically diagnosed or suspected CP. Clinical and genetic data were collected and analysed. Selected cases with available exome/genome sequencing data underwent analysis using bioinformatic filtering pipelines. **Results:** A total of 2121 individuals were identified from both databases (GEL-CP, n=1464; UCL-ION-CP, n=657). Approximately 40% displayed at least one recorded motor subtype phenotype with spastic/spastic-related phenotype being the most prevalent (~21.3%) followed by dyskinetic/dyskinetic-related phenotype (~15.8%), hypotonic/hypotonic-related phenotype (~13.5%) and ataxic/ataxic-related phenotype (~6.9%). Neurodevelopmental features like intellectual disability (~41.3%), developmental delay (~39.9%) and seizures/epilepsy (~36.7%) were commonly observed across both cohorts. The combined diagnostic yield was ~25.6% with around 400 unique genes involved. Most reported in the GEL-CP cohort were genes associated with autosomal dominant disorders whereas in the UCL-ION CP cohort, it was genes associated with autosomal recessive disorders. Reanalysis of cases uncovered both known and potential novel genes implicated in CP and increased the diagnostic rate up by ~10%. **Conclusion:** This study highlights the diverse clinical and genetic aspects of CP, revealing its overlap with rare neurodevelopmental and movement disorders which is in line with current CP genomic studies. In addition, it underlines the importance of reanalysing unsolved cases, given the ongoing gene discovery and updated clinical description in the context of CP. **Grant References:** Research Foundation, Cerebral Palsy Alliance PhD Research Grant.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2074F: Unraveling the genetic mimics of cerebral palsy: Treatable disorders should not be missed

Authors:

C. Lourenco, A. Queiroz, R. Albuquerque; Faculdade de Med. de São José do Rio Preto (FAMERP), São José do Rio Preto, Brazil

Abstract:

Background: Cerebral palsy (CP) describes a heterogeneous group of neurodevelopmental disorders characterized by onset in infancy or early childhood of motor symptoms (including hypotonia, spasticity, dystonia, and chorea), often accompanied by developmental delay. A growing group of metabolic and genetic disorders is being recognized, mainly due to the advancement in the area of metabolic and genetic investigations, to present with exactly the same symptoms and signs as CP.
Methods: Retrospective cohort study of 254 patients referred to a National Center for Rare Diseases in Brazil. All patients underwent genetic tests with two customized Next Generation Sequencing (NGS) panels in combination with biochemical testing (organic acids and amino acids chromatography, acylcarnitine profile, 3OMD dosage, prolactin, blood lactate, CK, homocysteine, among others) in a 2 year period. Results: Overall, pathogenic variants were detected in 145/254 patients (57%). Among them, we found mutations in specific dystonia genes (18%); ATP1A3, SLC2A1, GNAO1, NKX2-1, ADCY5, ATP1A2, MCT8 were found in 35 patients with previous clinical diagnosis of neurotransmitters defect. Pathogenic mutations in PANK2, PLA2G6 and WDR45 were found in 12 patients showing features of dystonia, hypotonia and progressive neurological disorder with no specific features in their brain MRI pointing out to a NBIA (neurodegeneration with brain iron accumulation) diagnosis. Biallelic mutations in neurotransmitters genes were found in 20% of the patients , being AADC deficiency , GTPCH1 and TH deficiency the main disorders identified in this group. The oldest patient diagnosed was an adult male , 65 years, diagnosed with Sepiapterin Reductase Deficiency (as far we know the oldest patient to be diagnosed in such age). Among the patients with no genetic diagnosis , 5 patients harboring one heterozygous VOUS in DDC gene were submitted to further biochemical analysis to rule out the diagnosis of AADC deficiency. None of them showed abnormal levels of 3OMD neither decrease in DDC enzyme.Conclusion: The term “cerebral palsy mimic” is used to describe a number of neurogenetic disorders that may present with motor symptoms in early childhood, resulting in a misdiagnosis of cerebral palsy. Ts. Our results suggest that gene panel analysis proved to be a valid method to detect a growing number of metabolic and genetic diseases presenting with clinical features that resemble a CP phenotype. Identifying the etiology of CP has practical consequences, such as proper family counseling about the cause of the disease, risk of recurrence, prognosis, and in some diseases provide disease-modifying therapy

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2075F: Unraveling the Genomic Landscape of Anti-Epileptic Drug Resistance: A Bioinformatics approach ⭐

Authors:

F. Fofana[1], M. K. Goita[2,1], A. Kiprop[3], L. Bernard[4], O. Soro[5], O. I. Awe[6]; [1]Univ. of Sci., Techniques and Technologies of Bamako (USTTB)/ African Ctr. of Excellence, Bamako, Mali, [2]Univ. of Sci. Techniques and Technologies of Bamako (USTTB), Bamako, Mali, [3]Fndn. of Knowledge, Moi Univ., Kenya, Eldoret, Kenya, [4]Noguchi Mem. Inst. for Med. Res., Univ. of Ghana, Accra, Ghana, [5]Molecular Biology and Biotechnology, Pan African Univ. for Basic Sci., Technology, and Innovation, Juja, Kenya, Juja, Kenya, [6]African Society for Bioinformatics and Computational Biology, Cape Town, South Africa, Cape Town, South Africa

Abstract:

Epilepsy is a chronic brain disorder affecting people of all ages, with a subset of patients experiencing drug-resistant epilepsy (DRE) despite treatment. Genetic factors, particularly variations in genes related to drug metabolism and transport like *CYP2C9*, *CYP2C19*, and *ABCB1*, are thought to contribute to this resistance. The mechanisms behind drug resistance in epilepsy are complex and not fully understood. Key hypotheses include drug metabolism, transport to the epileptic focus, and target sensitivity. This study aimed to explore the genomic landscape of anti-epilepsy drugs resistance (AEDs) to better understand these mechanisms and identify new genetic markers. We analyzed epilepsy pharmacoresistance datasets from the NCBI GEO database (*GSE252323*, *GSE213488*, *GSE210043*, and *GSE143272*). The study included 194 individuals: 143 epilepsy patients with varying responses to carbamazepine (CBZ), phenytoin (PHT), and valproate (VPA), and 51 healthy controls. Sequencing reads from SRA GEO repositories were quality-checked, trimmed, and aligned to the human genome GRCh38. Variant calling was performed using GATK, and variants meeting specific criteria were selected. Differential gene expression (DEG) analysis using limma in R compared responses to the three AEDs. Significantly differentially expressed genes (DEGs) were identified, focusing on those related to epilepsy pathways, neurodevelopment, or neuronal signaling. We identified 95 DEGs, with 52 upregulated and 43 downregulated when comparing phenytoin and valproate non-responders to responders. Valproate non-responders showed the most DEGs, with 45 upregulated and 40 downregulated. No DEGs were found when comparing responses to carbamazepine. None of the identified DEGs were linked to valproate or phenytoin response in PharmGKB. However, 3.84% (2/52) shared the same variant, *KCNQ2*: *c.G1461C*: *p.E487D* associated with potassium conductance. This study can contribute to the development of personalized epilepsy treatments, potentially improving patient outcomes. Further research is needed to explore the relationship between these genes and AED response in epilepsy.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2076F: Using Whole Genome Sequencing to improve diagnostic and care in Critically Ill pediatric patients in Africa: A Preliminary Study from Kinshasa.

Authors:

G. Mubungu[1], C. Bifu[1], P. Makay[2], P. Fuanani[1], C. Songo[1], C. Mbongomanu[1], D. Perry[3], A. Kesari[3], M. Ngole[1], C. Nsibu[1], P. Lukusa[1,4], E. Thorpe[3,5], R. Taft[3,5], A. Lumaka[1]; [1]Univ. of Kinshasa, Kinshasa, Congo, Democratic Republic of the, [2]Université de Kinshasa, Kinshasa, Congo, Democratic Republic of the, [3]Illumina, San Diego, CA, [4]Ctr. for Human Genetics, Katholeik Univ.it Leuven, Leuven, Belgium, [5]Genetic Alliances, Damascus, MD

Abstract:

INTRODUCTIONGenetic diseases have a big share among the causes of morbidity and mortality in newborns and infants. Recent studies have indicated that the proportion of genetic diagnosis is higher among severely affected infants. An early diagnosis can guide care and prevent deaths or irreversible complications. Comprehensive analysis of NGS data can shorten the diagnostic odyssey. The present study aimed to evaluate the contribution of the NGS to the diagnosis and care management of severely ill patients in Africa.METHODSThe study was conducted at the University Hospitals of Kinshasa, in the DR Congo. Critically ill patients without clinically recognizable syndromes and without suspicion of acquired cause were recruited from the NICU and PICU. DNA was extracted in the Center for Human Genetics in Kinshasa. cWGS was performed in the Illumina Laboratory Services as part of the Illumina iHope program.RESULTSForty patients, with 22 trios and 18 duos, were recruited. Twenty-three were from PICU and seventeen from NICU. Sex ratio (M/F) was 1.8 and 1.5 in PICU and NICU, respectively. Severe hypotonia, seizures, hypoglycaemia, hepatomegaly were the frequent clinical features. A definite molecular diagnosis was reached in eight patients (yield: 20%), namely complex 1q21.1-q21.2 rearrangement, Intellectual developmental disorder with severe speech and ambulation defects (ACTL6B, de novo), Alagille Syndrome (JAG1, absent in the mother, father unavailable), EEF2-related neurodevelopmental disorder (de novo), Hyperornithinemia-hyperammonemia-homocitrullinuria syndrome (SLC25A15, compound heterozygous), SCN2A-related disorders (de novo), and Methylmalonic aciduria and homocysteinemia cblX type (HCFC1, hemizygous, de novo).CONCLUSIONThese data demonstrate the diagnostic advantage of cWGS as a first-tier test in critically ill pediatric patients in Africa. The comprehensive analysis of WGS data offers a unique opportunity to reach additional diagnostic yield in this category of patients. WGS also allows novel variants to be identified from under-represented population and to be shared with the community.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2077F: Variable Expressivity of Rare Familial 14q13.3 Deletions That Include NKX2-1, PAX9, and SLC25A21

Authors:

B. Fischer[1], K. M. Antshel[2], N. Brescia[3], S. C. Smith[4], J. Black[5], R. R. Lebel[5]; [1]Norton Coll. of Med., SUNY Upstate Med. Univ., Syracuse, NY, [2]Dept. of Psychology, Syracuse Univ., Syracuse, NY, [3]Dept. of Neurology, SUNY Upstate Med. Univ., Syracuse, NY, [4]Dept. of Pathology, SUNY Upstate Med. Univ., Syracuse, NY, [5]Ctr. for Dev., Behavior and Genetics, SUNY Upstate Med. Univ., Syracuse, NY

Abstract:

A microdeletion isolated to 14q13.3 is exceedingly rare, such that its prevalence is unknown. It may present with the clinical phenotypes of hypodontia from haploinsufficiency of the *PAX9* gene (OMIM #167416), or choreoathetosis and congenital hypothyroidism with pulmonary dysfunction; known as brain-thyroid-lung syndrome (OMIM #610978), associated with haploinsufficiency of the *NKX2-1* gene (OMIM #600635). The *SLC25A21* gene (OMIM #607571) which is less than 1000 bases from *PAX9* is often deleted along with these genes. *SLC25A21* deletions may cause synpolydactyly. Here, we describe the clinical features of a 14q13.3 deletion segregating within a family and compare their neurodevelopmental trajectories. A maternal heterozygous 627 kbp deletion of 14q13.3 involving the *NKX2-1*, *PAX9,* and *SLC25A21* genes was found to segregate to her daughter. The phenotype of the mother includes oligodontia, hypothyroidism, behavior problems, and global developmental delays identified in early childhood. The cognitive delays were moderate to severe; in junior high school her full scale IQ score was in the low 70s and clinical tests found deficits in visual spatial comprehension. Her motor development delays were mild. She was diagnosed with choreoathetosis at 15 years old, and at 24 years she appears to be evolving dementia and cognitive decline. Her adult height is low at 5th percentile, and her weight is high at the 99th percentile. Her daughter's phenotype has presented with milder clinical features, including hypothyroidism, and mild motor delay. She has not shown signs of choreoathetosis or oligodontia, and has typical cognitive milestones at 14 months. Her longitudinal and ponderal growth are above average with height at the 71st percentile, and weight at the 99th percentile. Several other mother/daughter discrepancies have been observed. Both of these individuals have had seizures, and the daughter has recurrent congenital pilonidal cysts; these features have not been reported in other case studies involving similar deletions. Many of the expected gene-deletion associated phenotype features of *NKX2-1* that were observed in the mother during early childhood are diminished or have not been observed in the daughter. Some

features may still develop in the future; *PAX9* associated hypodontia can't be accurately diagnosed until 8 years old, and choreoathetosis does not begin until the teenage years. Clinical signs of the familial *SLC25A21* gene deletion were absent in both mother and daughter.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2078F: Whole exome sequencing and metabolomics screening of a Kobberling Familial Partial Lipodystrophy cohort identifies new candidate genes and dysregulated pathways ⭐

Authors:

C. T. Marvin[1], E. E. Blue[2,3], S. Subramanian[2], K. J. Buckingham[1], J. X. Chong[1,3], F. Wong[4], D. Bradley[4], S. Lancaster[4], M. Snyder[4], University of Washington Center for Rare Disease Research, C-L. Wei[5], M. J. Bamshad[1,3,5,6]; [1]Dept. of Pediatrics, Univ. of Washington, Seattle, WA, [2]Dept. of Med., Univ. of Washington, Seattle, WA, [3]Brotman Baty Inst., Seattle, WA, [4]Stanford Univ. Sch. of Med., Dept. of Genetics, Stanford, CA, [5]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [6]Seattle Children's Hosp., Seattle, WA

Abstract:

Kobberling Familial Partial Lipodystrophy (KFPLD1), also known as Familial Partial Lipodystrophy Type 1, is one of six rare familial partial lipodystrophy (FPLD) syndromes. KFPLD1 is characterized by loss of adipose tissue from the limbs paired with central adiposity, which includes a ledge at the boundaries of the lipodystrophic regions. It is often associated with severe insulin resistant diabetes, hypertriglyceridemia which can trigger acute pancreatitis, steatotic liver disease, early onset microalbuminuria, and premature cardiovascular disease. While FPLD is thought to be exceedingly rare, with a prevalence of one in one million individuals, genetic studies of large clinical care cohorts have estimated a prevalence nearly 100x greater suggesting systemic underdiagnosis of FPLD. While the genetic basis for five of the FPLDs is known, the etiology of KFPLD1 remains unknown.
We assembled a cohort of 53 participants ascertained for clinical features of KFPLD1. Exome sequencing of 22 probands underwent rare variant analysis to identify coding and canonical splice site changes within shared genes. Potentially explanatory variants were observed for three probands in genes underlying dominant lipodystrophies (*LMNA*, *PIK3R1*, *PPARG*) and heterozygous variants in genes underlying recessive FPLDs were observed for another five probands (*AGPAT2*, *CIDEC*, *LIPE*) for which the second variant, if any, remains unknown. Candidate variants were observed in multiple probands implicating related Mendelian disorders for which our probands may represent phenotypic expansions (*ABCG8*, *LPL*, *PRDM16*) or novel

gene-disease relationships (*ABCA7*, *ABCG2*); prior GWAS and mouse model studies have linked each of these genes to phenotypes involving adipose tissue, glucose level, or lipid homeostasis. These results suggest KFPLD1 is not a single gene disorder like other FPLDs and instead exhibits genetic heterogeneity.
To find pathways that may be dysregulated in KFPLD1, 31 probands and 78 control samples underwent both metabolomics (704 metabolites) and lipidomics (849 lipid traits) screening. Differentially abundant metabolites were enriched in several pathways including methionine and alanine metabolism. The unbiased lipidomics analyses revealed the expected elevation in triglycerides in KFPLD1, but also significant differences in several subtypes of diradylglycerolipids along with two glycerophospholipids. Together, this dataset provides new candidate genes and pathways for further study to investigate how these may interact with the lipodystrophy and metabolic phenotype features of KFPLD1. Additional KFPLD1 cases are being enrolled for further studies.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2079F: Whole genome analysis in consanguineous families reveals new loci for speech sound disorder (SSD)

Authors:

M. Basra[1], T. Yasmin[1], A. Saadia[1], L. Nadeem[1], M. Rice[2], M. Raza[2]; [1]Sch. of Chemistry, Univ. of the Punjab, Lahore, Pakistan, [2]Speech-Language-Hearing: Sci. & Disorders, Univ. of Kansan, Lawrence, KS

Abstract:

Speech is the most common means of communication in humans. Any deficiency in accurate speech production ability results in the development of speech sound disorder (SSD), a most common communication condition in developing children that can significantly impair an individual's academic performance, social interactions, and relationships with peers and adults. We performed family-based genome-wide parametric linkage analysis and homozygosity mapping in 3 consanguineous families with SSD from the Punjab province of Pakistan. The Test for Assessment of Articulation and Phonology in Urdu (TAAPU) was used to analyze the speech articulation data and determine the Percentage Correct Consonants (PCC) score. The TAAPU was locally designed for the Urdu-speaking population of Pakistan. The PCC score was used to assess speech articulation in the available individuals of each family. The speech articulation data showed that substitution errors are dominant among other articulation errors in families with SSD. Parametric linkage analysis revealed a linkage to chromosome 5 (5q21.3-5q23.1) with a

significant logarithm of the odds (LOD) score of 3.13 in a family PKSLI-97 under an autosomal recessive mode of inheritance. The other two families showed a suggestive linkage at 6p22.1, 14q12, and 16q12.1 under the recessive mode of inheritance. Interestingly, homozygosity mapping showed a loss of heterozygosity in the linkage region at 5q15-5q23.1, shared among seven affected (mainly in the younger generation) and one unaffected individual of PKSLI-97. Our analysis identified the 6p22 locus previously implicated in dyslexia, childhood apraxia of speech (CAS), and language impairment, confirming the role of KIAA0319 and DCDC2 in this locus. Our study confirmed previously implicated genomic regions in SSD. This study also suggested new genomic loci associated with articulation disorder and offers future opportunities to further the role of genes in speech production.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2080F: Xanthine Stones and Molybdenum Cofactor Deficiency: A Genetic Exploration of MOCS1, MOCS2, and GPHN

Authors:

I. Patel[1,2], K. Lovrenert[1], D. Bodner[3,4], F. Schumacher[1], C-H. Wu[1,2]; [1]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Univ. Hosp. of Cleveland Med. Ctr., Cleveland, OH, [3]Case Western Reserve Univ., Cleveland, OH, [4]Louis Stokes Cleveland Dept. of VA Med. Ctr., Cleveland, OH

Abstract:

Introduction: Molybdenum Cofactor Deficiency (MoCD) is a rare autosomal recessive metabolic disorder characterized by onset in infancy of poor feeding, and severe neurological deterioration due to deficiencies of enzymes that use molybdenum as a cofactor. Patients may present with xanthinuria and xanthine stones. Current treatment options include Fosdenopterin to treat MoCD Type A, and supportive care. Most affected individuals experience mortality in early childhood.
Efforts to uncover the genetic etiologies of MoCD have resulted in the identification of 3 causal genes associated with xanthine stones: *MOCS1* which causes MoCD Type A; *MOCS2* which causes MoCD Type B; and *GPHN* which causes MoCD Type C. This investigation aims to estimate the genetic prevalence of MoCD associated with xanthine stones through the *MOCS1*, *MOCS2*, and *GPHN* genes and compare it to the estimated global incidence.
Methods: Variants were identified through (1) the 1000 Genomes Project Phase 3 (1KG), which identifies variants in the general healthy population, and (2) the Human Gene Mutation Database (HGMD), which identifies pathogenic variants. Variants from both databases were intersected to identify pathogenic variants in the general population.

Results: In 1KG, non-related healthy individuals (n=2504) carried 882, 321, and 15903 unique variants for *MOCS1*, *MOCS2*, and *GPHN* respectively. HGMD contained 41, 17, and 20 unique pathogenic variants for *MOCS1*, *MOCS2*, and *GPHN* respectively. These variants include missense, nonsense, synonymous, stop-loss, insertions, deletions, and complex substitutions. The intersection resulted in 1 pathogenic variant of *MOCS1* across 8 heterozygote carriers and 0 pathogenic variants of *MOCS2* or *GPHN*.
Calculations based on Hardy-Weinberg equilibrium show disease-causing alleles to have an allele frequency of 0.16%, carrier rate of 1 in 314, and an affected rate of 1 in 391,876 for *MOCS1*, and allele frequency of 0%, carrier rate of 0 in 0, and an affected rate of 0 in 0 for *MOCS2* and *GPHN*. The combined carrier rate of the three genes is 1 in 314 and the combined affected rate is 1 in 391,876.
Conclusion: This investigation reveals 8 *MOCS1* carriers and 0 *MOCS2* and *GPHN* carriers in the healthy, unrelated 1KG population. The literature reports an estimated global incidence of MoCD is 1 in 100,000 to 1 in 200,000. Of this, 49% of patients are affected by *MOCS1* pathogenic variants, 45% by *MOCS2* variants, and 4% by *GPHN* variants. We report an affected rate of 1 in 391,876 for MOCS1 and 0 in 0 for *MOCS2* and *GPHN*. Thus, our study reveals that the prevalence of *MOCS1* estimated by genetics is comparable with clinical estimates, while *MOCS2* and *GPHN* are less common by genetics estimate.

Session Title: Mendelian Phenotypes Poster Friday Session

Board 2081F: Zhu-Tokita-Takenouchi-Kim Syndrome (ZTTK): a Novel Neurodevelopmental Disorder caused by pathogenic variants in *SON*

Authors:

G. Raines, B. Park, E. Ahn; Univ. of Alabama at Birmingham, Birmingham, AL

Abstract:

Zhu-Tokita-Takenouchi-Kim (ZTTK) Syndrome is caused by heterozygous loss-of-function mutations in SON, which encodes for a protein involved in pre-mRNA splicing of many genes critical for development. This leads to neurodevelopmental manifestations with multiorgan involvement. ZTTK syndrome is associated with a broad phenotypic spectrum characterized by developmental delay (DD), intellectual disability (ID), musculoskeletal abnormalities, dysmorphic facial features, vision abnormalities, and cerebral cortex malformations. We aim to present a summary of the collaborative efforts further characterizing the phenotypic spectrum of this novel syndrome and developing management guidelines for clinicians. Since the description of ZTTK Syndrome, a parent organized advocacy group called the SON-Shine

Foundation was formed to provide support and resources for those patients affected. The group seeks to promote awareness surrounding ZTTK Syndrome and promote research efforts. For example, patients with ZTTK Syndrome have displayed recurrent infections and low immunoglobulins; research in this domain includes further exploration of the mechanisms surrounding the hematologic and immunological phenotypes. We will also present an overview of the ongoing research efforts of the foundation and the creation of a patient registry to amplify opportunities for interaction between patients and researchers.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3041F: A 15-Year Review of Cytogenetic Studies and Analyses Done at the National Institutes of Health, University of the Philippines Manila from 2008-2022 ⭐

Authors:

M. Borcelis, M. B. Alcausin, E. G. Maceda, J. Z. Oblefias, W. R. Bagay, E. G. Salonga; Univ. of the Philippines - Manila, Manila, Philippines

Abstract:

The Cytogenetics Laboratory, established in 1991 under the Institute of Human Genetics, National Institutes of Health-University of the Philippines Manila, is the country's leading provider of cytogenetics tests and analyses through processing of peripheral blood, cord blood, bone marrow, and skin/tissue samples. This paper aims to present the results of the retrospective review of the cytogenetic studies and analyses conducted by the laboratory from 2008-2022.

The database of the Cytogenetics Laboratory was reviewed to retrieve information of samples analyzed limited only to the following variables: results of the cytogenetics tests, indications for referral, specialty of the referring physician, and region where the sample came from. Peripheral blood samples account for the greater majority of the samples processed with approximately 13,385 blood samples submitted for analyses from 2008-2022. Out of the all the patient samples processed and analyzed, 83.6% were from Luzon, 12% from Visayas, and 4.4% from Mindanao with regions NCR (National Regional Capital), Region IV-A (CALABARZON), and Region III (Central Luzon) having the highest utilization of cytogenetic services. The top indications of referral for chromosomal analysis were confirmation of chromosomal diagnosis, multiple congenital anomalies, and reproductive health disorders by physicians specializing in Pediatrics,

Obstetrics and Gynecology, and Internal Medicine. Results have shown that numerical chromosome abnormalities were much more frequently seen than structural chromosome abnormalities with 10:1 ratio. Among the numerical chromosome abnormalities, 97.4% were autosomal abnormalities and 2.6% were sex chromosome abnormalities. The autosomal numerical abnormalities consisted of 89.5% trisomy 21 while the predominant sex chromosome abnormality is made up of 64.1% turner syndrome. Among the autosomal abnormalities structural chromosome abnormalities, 16.9% were translocations and 36.2% were deletions. This 15-year review emphasizes the relevant role of cytogenetic analysis in diagnosing and managing patients with suspected chromosomal abnormalities in a limited resource setting like the Philippines. The high prevalence of numerical abnormalities highlights the need for early detection and intervention to improve patient outcomes. The regional disparities in the utilization of cytogenetic services also suggest a need to increase awareness and improve access to these tests across the country which might help in the integration of cytogenetic testing into clinical practice nationwide.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3042F: Advancing mitochondrial DNA analysis: Utility of PacBio long-read sequencing technology in accurate and precise variant detection ⭐

Authors:

T. Jadhav[1], M. Aruta[1], L. Conlin[1,2], R. Rajagopalan[1,2], J. Wang[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Purpose: Primary mitochondrial diseases (PMDs) can be caused by pathogenic variants in mitochondrial DNA (mtDNA). A wide variety of variants including single nucleotide variants (SNVs), single large-scale mtDNA deletions (SLSMD), and multiple mtDNA deletions (MMDs) contribute to the complexity of mtDNA testing. Accurate detection and differentiation of MMDs and SLSMDs, along with precise heteroplasmy quantification are imperative for clinical correlation. Clinical mtDNA tests utilize targeted Next Generation Sequencing (NGS) for variant detection and ddPCR for heteroplasmy quantification. This multi-tiered approach is necessary to address the limitations of NGS in detecting deletion heteroplasmy. However, this approach fails to identify precise deletion breakpoints. Long-read sequencing (LRS) has emerged as a powerful tool for detecting genetic variants otherwise not detectable by the current NGS technologies, but its utility for mtDNA analysis remains underexplored. This study evaluates PacBio LRS for

mtDNA SNV and large deletion detection and quantification.**Methods:** Twenty-two samples including 6 SLSMD-positive cases (4 blood, 2 muscle), 11 MMDs-positive muscle samples, and 5 deletion-negative controls (2 blood, 3 muscle) were sequenced using the PacBio Sequel II. Two samples were sequenced twice to assess reproducibility. A bioinformatics pipeline was developed for mtDNA large deletion detection and breakpoint clustering. The SNVs, structural variants, and heteroplasmy levels from LRS data were compared against the standard of care, clinical mtDNA testing assay at our hospital.**Results:** Our findings indicate LRS's efficacy in accurately detecting and distinguishing SLSMDs and MMDs in samples with deletion heteroplasmy >10%. LRS achieved 100% concordance for both MMDs and SNVs. Deletion heteroplasmy computed from LRS was highly correlated with the ddPCR estimates (r^2 = 0.87). SNV detection from LRS demonstrated high sensitivity (99% overall; 100% for SNV heteroplasmy >10%) and 100% specificity compared to SR-NGS in the three deletion-negative samples. However, LRS's sensitivity to detect SNV heteroplasmy <10% was limited. Breakpoint clustering revealed the recurrent 4977bp deletion's prevalence in MMDs samples.**Conclusion:** LRS's high accuracy in detecting and differentiating large-scale single and multiple mtDNA deletions and SNVs along with heteroplasmy levels demonstrates its potential as a comprehensive clinical test covering the broad mtDNA variation spectrum. Identifying precise deletion breakpoints in samples with multiple deletions provides the opportunity to study mtDNA replication mechanisms.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3043F: An exploratory prospective clinical trial of 120 neonates in BeginNGS, an artificial intelligence-enabled newborn genome sequencing system of 412 childhood genetic disorders. ⭐

Authors:

J. Schleit[1,2], L. Protopsaltis[3,2], L. Olsen[1,2], B. Schultz[1,2], M. Wright[4], M. Baker[5], S. Caylor[6,2], T. Defay[7], K. Ellsworth[3,2], A. Feigenbaum[8,6,2], L. Guidugli[6,2], Y. H. Kwon[1], J. Le[6,2], J. Lenberg[6,2], L. Madhavrao[6,2], R. Mardach[9,2,10], D. Oh[1,2], M. Saad[1], E. Sanford[1], G. Scharer[6], L. D. Smith[1], L. Van Der Kraan[11,2], K. Watkins[1,2], K. Wigby[8,6,2], M. Willis[12], R. Reimers[13,6], S. Kingsmore[14,2]; [1]Rady Children s Inst. for Genomic Med., Seattle, WA, [2]Rady Children’s Hosp., San Diego, CA, [3]Rady Children s Inst. for Genomic Med., San Diego, CA, [4]Rady Children's Inst. for Genomics Med., San Diego, CA, [5]5Wisconsin State Lab. of Hygiene, Ctr. for Human Genomics and Precision Med., Madison, WI, [6]Rady Children’s Inst. for Genomic Med., San Diego, CA, [7]Alexion, Astra Zeneca Rare Disease, Boston, MA, [8]Univ. of California San Diego, San Diego, CA, [9]Rady Children s Hosp. of San Diego, San Diego, CA, [10]4Dept. of Pediatrics, Univ. of California San Diego, San Diego,

CA, [11]Rady Children's Inst. for Genomic Med., San Diego, CA, [12]Rady Children's Inst. for Genomic Med., La Jolla, CA, [13]Rady Children's Hosp., San Diego, CA, [14]Rady Children s Inst. for Genomic Med., Coronado, CA

Abstract:

Newborn screening (NBS) improves outcomes in some severe childhood genetic disorders (SCGD) by early identification and therapy. Genome sequencing (GS) could extend NBS to all SCGD. We are developing a platform to accomplish this based on NBS with a derivative of rapid diagnostic genome sequencing (RDGS), artificial intelligence, and a management guidance system (BeginNGS). We report results of a pilot prospective clinical trial of BeginNGS in 120 newborns aged 1-10 in a level IV NICU who were not suspected of having a genetic disease (Clinicaltrials.gov NCT06276348). Enrollees received standard NBS, RDGS, and BeginNGS from dried blood spots, with return of results to neonatologists or primary care pediatricians. RDGS was performed in a standard manner with 14-day turnaround. BeginNGS was performed by automated analysis of GS data with a set of ~54,000 pathogenic (P) and likely pathogenic (LP) variants associated with 342 genes, 412 disorders, and 1,603 effective therapeutic interventions. Post-enrollment and post-result disclosure surveys were administered to parents. 45 RDGS findings were reported in 34 (28%) enrollees, of which 13 were expected to change management. 4 (3%) newborns had confirmed positive state NBS results. BeginNGS identified findings in 6 (5%) newborns, of which 5 were true positives and one is indeterminate. 118 and 51 parents completed the post-enrollment survey and post-result disclosure surveys, respectively. Most respondents had completed post-secondary education (57%). The most common primary languages were English (59%) and Spanish (4%). In the pre-enrollment survey, most parents (88%) felt adequately informed about possible issues related to their child undergoing GS. Similar levels of satisfaction (90%) were reported following receipt of results. 88% of parents felt their child's GS results were useful. However, 64% of respondents reported understanding their child's test results, 32% reported somewhat understanding, and 4% of respondents did not understand the results. Most parents felt more knowledgeable about their child's future health (86%), better able to manage potential symptoms from their child's condition (55%), and able to make informed decisions on reproductive planning or caring for other family members (51%). In summary, parents of NICU newborns who were not suspected of having a genetic disease were satisfied with RDGS and BeginNGS and found them to be useful. However, over one third of parents did not fully understand RDGS and BeginNGS results. This pilot study informed the design of a large adaptive clinical trial of BeginNGS that is underway (Clinicaltrials.gov NCT06306521).

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3044F: An interstitial duplication of 2q11.2q13 in a neonate with multiple congenital anomalies: A case report

Authors:

M. Regmi, F. Shakil, C. Kumar; Dept. of Pathology, Westchester Med. Ctr./New York Med. Coll., Valhalla, NY

Abstract:

Partial trisomy of the distal long arm segments of chromosome 2 are known from a number of reports. Proximal duplications were rarely reported because their determination requires successive fluorescence in situ hybridization (FISH) and molecular investigations. Large duplications of the long arm of chromosome 2 have been reported in children with developmental delay (DD), intellectual disability (ID), autism spectrum disorder (ASD), microcephaly, short stature, cleft lip and palate. In general, pure proximal duplication or triplication of 2q is rare. Due to the variable clinical features, no common pattern of a proximal duplication 2q syndrome could be defined from these cases, even with similar chromosomal breakpoints. Although diverse phenotypes are associated with these abnormalities, some common features have emerged. It has been reported that individuals with duplication 2q11.2 had developmental delay, gastroesophageal reflux and short stature, whereas dysmorphic features, DD and abnormal head size are common with duplication 2q13 carriers. Here, we present a 1-month-old female born via vaginal delivery to a 29-year-old mother with Type 1 Diabetes and polyhydramnios. Baby was diagnosed with multiple congenital anomalies like trachea-esophageal (TE) fistula, ventricular septal defect, 11 ribs, flattened broad nasal bridge and bilateral foot 2nd toe overlapping 1st toe. Chromosomal microarray and cytogenetic investigations were indicated. Chromosome analysis was performed on GTG banded metaphases prepared from cultured lymphocytes that revealed an interstitial duplication of chromosome 2q and the infant's karyotype as: 46,XX,dup(2)(q11.2q13). The microarray analysis performed on whole blood further confirmed a 10.7 megabase of interstitial duplication of 2q from 2q11.2 to 2q13 region. The duplicated interval involved 100 known genes (55 protein-coding), none of which are individually associated with abnormal phenotypes due to copy number gain. A similar case of de novo proximal 2q tandem duplication (2q11.1-q13.2) were identified by FISH studies in a 4-year-old boy with intellectual disabilities, unilateral cleft lip and palate, but otherwise with discrete minor anomalies associated with developmental delay. In conclusion, this case demonstrates the effective use of cytogenetics and microarray studies to confirm the duplicated region: dup(2)(q11.2q13) for patient's diagnosis and clinical management. Parental chromosome analyses were recommended to determine whether this abnormality is inherited or de novo in origin. Clinical correlation and genetic counseling were recommended.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3046F: Clinical relevance of mosaic variants detected from exome sequencing data in a cohort enriched in primary immune disorders.

Authors:

R. Ghosh[1], Z. Fazal[2], A. Oler[2], J. Yan[1], M. Kamen[1], M. Similuk[1], A. Paul[3], M. Cooper[3], S. Holland[1], M. Walkiewicz[1]; [1]NIAID Centralized Sequencing Program, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]Bioinformatics and Computational BioSci.s Branch, Office of Cyber Infrastructure and Computational Biology, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [3]Washington Univ., St. Louis, MO

Abstract:

Background: Mosaicism, the presence of two or more genetically distinct cell populations within an individual, can significantly impact the diagnosis and management of genetic disorders. This study investigates the prevalence and clinical significance of mosaic variants in primary immune disorders from exome sequencing data.
Methods: Exome sequencing was performed on blood-derived DNA from a cohort of 2064 patients with various primary immune disorders. Lofreq2 and Mutect2, algorithms tailored for mosaic variant detection, were employed to call mosaic variants from the exome sequencing data. Using validation by deep sequencing and/or droplet digital PCR, we established a lower limit of detection at 0.03 variant allele fraction (VAF). Clinical data were retrospectively reviewed to evaluate the phenotypic impact of identified variants.
Results: Mosaic variants resulting in a molecular diagnosis were detected in 28 out of 2064 patients, reflecting a prevalence of 1.4%. These variants were associated with a range of disorders including mastocytosis, autoinflammatory and myeloproliferative diseases. We found that 36% of these variants were missed by a germline variant calling workflow. We also detected instances of apparent parental mosaicism in 53 out of 222 (2.4%) families, defined by a parental VAF < 0.2 and proband VAF >0.4 for a given variant. We identified two families with pathogenic variants in *STAT1* and *NOD2* that could potentially allow for more accurate recurrence risk assessment. Notably, a mosaic variant in *NRAS*, implicated in chemotherapy resistance, was identified in a patient at a VAF of 0.05. We also identified new mosaic research candidates, including 58 novel predicted loss of function variants in genes constrained for such variants, 125 variants deemed pathogenic/likely pathogenic by at least two submitters in ClinVar and 86 variants implicated in clonal hematopoiesis.
Conclusion: The use of dedicated algorithms for detecting mosaic variants from exome sequencing data enhances the diagnostic yield in patients with primary immune disorders. This

method discovered new genetic candidates for prognosis and research and offered better estimates of recurrence risk, highlighting the need for characterizing mosaic variants in the accurate diagnosis and management of immune disorders.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3047F: Clinical significance of imprinting center 1 (IC1) microdeletions in Beckwith-Wiedemann syndrome

Authors:

Q. Ding, Z. Stander, E. Stelmach, L. Hasadsri, K. Rumilla, W. Shen; Div. of Lab. Genetics and Genomics, Dept. of Lab. Med. and Pathology, Mayo Clinic, Rochester, MN

Abstract:

Background: Beckwith-Wiedemann syndrome (BWS) is an overgrowth imprinting disorder, commonly associated with a constellation of clinical manifestations including macroglossia, macrosomia, hemihyperplasia, abdominal wall defects, and increased risk for certain cancers. While the vast majority of BWS patients are sporadic in nature, familial BWS has also been reported. Maternally inherited microdeletions in imprinting center 1 (IC1, *H19* DMR) at 11p15.5 have been suggested to cause aberrant IC1 hypermethylation, thereby leading to familial BWS. Nonetheless, their pathogenicity remains uncertain. This study aims to further clarify the clinical significance of these deletions. **Methods**: Retrospective review of copy number and methylation-sensitive probe results of patients who underwent multiplex ligation-dependent probe amplification (MLPA) testing for BWS or Russell-Silver syndrome (RSS) at our CLIA-certified, CAP-accredited clinical laboratory between June 2015 and April 2024 was performed. Several cases with IC1 microdeletions upstream of the *H19* gene were identified, and further investigated in light of clinical and familial history. **Results**: Among the 8,607 patient results, 12 patients (from 8 families) carrying microdeletions upstream of *H19*: NC_000011.9:g.(2020007_2022376)_(2022833_2025882)del (GRCh37/hg19) were identified. Similar deletions have not been reported in large control populations. In our cohort, 10 of the 12 patients with the microdeletion also had IC1 hypermethylation (significantly enriched, p = 4.64e-13, odds ratio: 54.53 [95% CI: 20.73-140.29], Fisher's exact test), including all five patients with known maternal inheritance of the microdeletion. Notably, we identified a three-generation family, in which all three maternal transmissions of the microdeletion co-segregated with IC1 hypermethylation and clinical features of BWS. Using Sanger sequencing, we mapped the deletion breakpoints in this family as NC_000011.9:g.2022169_2024001del (GRCh37/hg19). **Conclusions**: These findings provide additional support for pathogenicity of IC1

microdeletions in causing BWS. Similar deletions are now clinically classified as Likely Pathogenic by our clinical laboratory.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3048F: Clinical validation of RNA sequencing for Mendelian disorder diagnostics ⭐

Authors:

P. Liu[1,2,3], S. Zhao[1,2], K. Macakova[1,2,4], J. Sinson[1,2], H. Dai[1,3], J. Rosenfeld[1], G. Zapata[1,2], S. Li[1,2], P. Ward[1,2,3], C. Wang[1,2], C. Qu[3], B. Maywald[1,2], Undiagnosed Diseases Network, B. Lee[1,2], C. Eng[1,2,3]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Med. Genetics and Multiomics Lab., Baylor Coll. of Med., Houston, TX, [3]Baylor Genetics, Houston, TX, [4]Graduate Program in Diagnostic Genetics and Genomics, The Univ. of Texas MD Anderson Cancer Ctr. Sch. of Hlth.Professions, Houston, TX

Abstract:

Despite rapid advancements in clinical sequencing, over half of diagnostic evaluations still lack definitive results. RNA-seq has shown promise in research settings for bridging this gap by providing essential functional data for accurate interpretation of diagnostic sequencing results. However, despite advanced research pipelines, clinical translation of diagnostic RNA-seq has not yet been realized. The field lacks the comprehensive implementation knowledge required for such a transition.

We have validated clinical diagnostic RNA-seq in a CLIA laboratory for individuals with suspected genetic disorders who have existing or concurrent comprehensive DNA diagnostic testing. Accepted sample types include RNA from fibroblast or blood.

The clinical interpretation of RNA-seq results revolves around the detecting outliers in gene expressions and splicing patterns. We designed the validation in three phases: (1) evaluating single sample gene expression and splicing calling performance, (2) establishing reference ranges for gene expressions and splicing events across sample pools, and (3) validating diagnostic RNA abnormalities, including expression and splicing outliers. We sequenced samples as benchmarks, negative controls, and those with positive clinical findings, totaling 150 runs.

To evaluate expression and splicing calling performance, we established a provisional RNA-seq benchmark using short- and long-read RNA-seq data for HG002 (Genome in a Bottle). Validation runs using HG002 achieved high sensitivity and specificity (>99%).

We next utilized all samples from the validation runs to establish reference ranges for expression levels and splicing patterns. These quality measurements offer higher resolution for

assessing test performance at individual genes or splice sites, which traditional metrics such as TPM cannot provide.
Finally, we evaluated clinical performance using 20 blood and 20 fibroblast samples from individuals with diagnostic RNA abnormalities. In a permissive filtration mode, our pipeline demonstrated high sensitivity at 95% for both fibroblast and blood samples, but it was accompanied by a high number of non-diagnostic outliers, resulting in low positive predicted values (PPV). We tuned parameters to balance clinical sensitivity and PPV, concluding that RNA-seq clinical review may benefit from a combined strategy, which includes an agnostic analysis (balanced sensitivity and PPV) and a targeted analysis (optimized sensitivity but requiring prior DNA variant knowledge).
In conclusion, our study provides a paradigm and necessary resources for independent laboratories to validate a clinical RNA-seq test.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3049F: Comparison of prenatal cell-free DNA screen positive for monosomy X with diagnostic maternal and fetal cytogenetic studies: a single-center retrospective study ⭐

Authors:

C. Myers, S. Coe, J. Buchan, B. Shirts, W. Neufeld-Kaiser, M. Das, E. Cheng, E. Oehler, V. Sybert, T. Kolarova, C. Disteche, Y. Liu, C. Lockwood; Univ. of Washington, Seattle, WA

Abstract:

Prenatal cell-free DNA screening, also referred to as non-invasive prenatal testing (NIPT), has been clinically available since 2011 to screen for the common fetal aneuploidies, including trisomy 13, 18, 21, and sex chromosome abnormalities. In 2020, ACOG recommended that all pregnant individuals be offered prenatal genetic screening and diagnostic testing options. False positive and false negative results from prenatal cfDNA screening are caused by a variety of biologic mechanisms, including confined placental mosaicism, a vanishing twin, and maternal factors. The positive predictive values (PPV) of prenatal cfDNA screening are dependent upon the chromosome in question, laboratory sensitivity and specificity, and *a priori* risk for aneuploidy. The PPV is highest for trisomy 21 at advanced maternal age (79-99%). The PPV of monosomy X (45,X) is lower at 25-40% irrespective of age. **We sought to assess the performance of prenatal cfDNA screening to detect monosomy X in an unselected obstetrical population using a low-pass whole-genome sequencing platform and review of diagnostic cytogenetic testing. Of 15,674 individuals screened, 54 had a cfDNA screen positive result for**

monosomy X. Eight were lost to follow-up and 3 have diagnostic studies pending. Of the remaining 43 individuals, **17 (40%) had conceptions with true fetal monosomy X**. In one of these 17, a woman was diagnosed with mosaic monosomy X (mos 45,X[9]/46,XX[11] in peripheral blood) after her third pregnancy resulted in a fetus with monosomy X. **Five (12%) individuals had maternal loss of X confirmed by peripheral blood karyotype or IFISH studies.** Two of these five most likely have age-related loss of the X chromosome, but **three likely have constitutional mosaic monosomy X (45,X in 20-30% of blood cells).** Two of the three with mosaic monosomy X were carefully examined and none had obvious features of mosaic Turner syndrome. **Twenty individuals (46%) had false positive cfDNA screen results with normal fetal diagnostic studies (amniocentesis or postnatal cord blood testing).** Nine of these 20 also had normal maternal peripheral blood karyotype results. **One individual (2%) is unresolved; two independent pregnancies (2 years apart) have screened positive for monosomy X despite having normal fetal and maternal cytogenetic analyses.** Our understanding of mosaic monosomy X and its implications for fertility and medical management is evolving. This study contributes to the **increasing number of reports about individuals who are unexpectedly found to have mosaic monosomy X through prenatal cfDNA screening** of a spontaneous pregnancy and highlights some of the challenges in reporting and interpretation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3050F: Copy number Variants in Clinical Whole Exome Sequencing of First 1000 cases

Authors:

C. Kaiwar[1], T. E. Firman[2], R. Barnard[1], N. J. Burns[1], H. C. Cox[1], D. DeMille[1], N. B. Newman Edelman[1], S. Parets[1], J. Taggart[1], P. Vrana[1], J. Yuan[1], N. Meeks[1], A. Gaskell[1]; [1]Children's Hosp. of Colorado, Aurora, CO, [2]Fred Hutch Cancer Ctr., Seattle, WA

Abstract:

BACKGROUND AND AIM The diagnostic yield of chromosomal microarray (CMA) for detection of germline copy number changes in patients with developmental delay, intellectual disability, autism, and/or multiple congenital anomalies is well documented and it has long been established as the first-tier test for these constitutional disorders by the American College of Medical Genetics as well as by the Society for Maternal Fetal Medicine. Meanwhile there has been increasing use of whole exome sequencing (WES) as a parallel test for the same cohort and improvement in technologies that allow for copy number calling using WES. This retrospective analysis of the first 1000 whole exome sequencing tests at the Precision

Diagnostics Laboratory (PDL) at the Children's Hospital of Colorado (CHCO) aims to review the impact of copy number variants called from WES data, on diagnostic yield, rate of multiple diagnoses, and ordering trends in a pediatric practice.**RESULTS**A total of 1007 WES tests were reported by PDL between 1st January 2023 to 31st May 2024. Of these 91 had copy number variants on the report, a yield of 9% percent. The size of copy number variants ranged from 12kB to whole chromosome copy number gains or losses. Copy number losses accounted for 55% of detected events while the rest were copy number gains. About 14% of detected events were below 200kB in size, making them less likely to be detected by CMA. More than half of the detected events (56%) were classified as Pathogenic or Likely pathogenic based on the ACMG/AMP criteria. In 5% cases a copy number variant along with a sequence variant provided a comprehensive molecular diagnosis that explains the total phenotype of the proband. Turner and Klinefelter syndromes were detected in 10 cases. The detected events were confirmed de novo in 20% (19/91) cases. In 46% (42/91) cases, no chromosomal microarray testing had been performed, WES was used as the first-tier test. More insights into the overall impact of this data set are forthcoming, along with a cost-effectiveness analysis of WES over a combination of WES and CMA. **CONCLUSIONS**The diagnostic yield of copy number variants in the first 1000 PDL WES cases falls within the expected range of detection of these variants from WES data and is dependent on the primary indication for testing. Utilizing familial samples to perform trio or duo exome testing disambiguates inheritance patterns for these events making for more comprehensive counselling and management. Moreover, WES facilitates finding multiple diagnosis that explain the complete phenotype of individuals with complex presentations. Further analysis will provide additional insights.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3051F: Cytogenetics in the era of genomic medicine: a retrospective study of 750 patients

Authors:

J. Levy[1], X. Latypova[1], S. Drunat[1], N. Couque[1], Y. Vial[1], C. Collet[2], A. Bonnard[1], N. Derive[3], P. Blanc[3], L. Goujon[4], L. Ruaud[1], L. Perrin[1], Y. Capri[1], A. Delahaye-Duriez[5], M. Heulin[5], M. Houlier[5], C. Mignot[4], B. Chaumette[6], A. Maruani[1], D. Germanaud[1], P. Gaignard[7], C. Dupont[1], A. Verloes[1], A-C. Tabet[1]; [1]AP-HP, CHU Robert-Debré, Paris, France, [2]AP-HP, CHU Necker, Paris, France, [3]AP-HP, Laboratoire SeqOIA, Paris, France, [4]AP-HP, CHU Pitié-Salpétrière, Paris, France, [5]AP-HP, CHU Jean Verdier, Bondy, France, [6]GHU Psychiatrie & NeuroSci.s, Paris, France, [7]AP-HP, Hôpital de Bicêtre, Kremlin-Bicêtre, France

Abstract:

Constitutional cytogenetics focuses on identifying numerical and structural chromosome anomalies involved in genetic pathologies. Historically dependent on manual, time-consuming techniques such as karyotyping and FISH, the field has advanced with the advent of chromosomal microarray analysis (CMA), enhancing diagnostic yield by 5 to 15%. As part of the *Plan France Médecine Génomique 2025* (PFMG 2025), genome sequencing is emerging as the preferred method for identifying and characterizing genomic alterations at the nucleotide level. In this study, we analyzed cytogenetic anomalies in the first 750 patient cases (about 2250 trio genome) using short-read genome sequencing in SeqOIA laboratory, addressing three clinical indications: 1) intellectual disability (ID), 2) malformation syndromes, and 3) autism spectrum disorder/neurodevelopmental disorders without ID. Genome sequencing revealed a broader spectrum of genomic alterations than traditional methods, uncovering rearrangements below CMA detection thresholds and gene-level structural anomalies such as translocations, inversions and ring chromosomes. It enabled a more detailed characterization of the genomic architecture of complex chromosomal rearrangements (CGRs), particularly through the detection of newly reported recurrent chromosomal patterns (dup-trip-inv dup, dup-normal-dup, etc.) Our study also highlights the limitations of short-read sequencing to determine the topological organization of structural anomalies at chromosomal level. Therefore, we complemented the analysis with optical genome mapping (Bionano), highlighting the complementary nature of these technologies in characterizing structural anomalies.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3052F: De novo variant calling from duo long-read sequencing: improving equitable variant interpretation for diverse family structures

Authors:

L. Boukas[1], E. Delot[1], G. Pitsava[2], UCI-GREGoR Consortium, E. Vilain[2], S. Berger[1]; [1]Children's Natl. Hosp., Washington, DC, [2]Univ. of California, Irvine, Irvine, CA

Abstract:

The widespread implementation of sequencing technologies in the clinical diagnostic setting has led to the realization that a substantial fraction of Mendelian disease phenotypes are caused by de novo variants. Thus, the ability to detect de novo variants via trio sequencing is associated with higher diagnostic yield compared to proband only or duo (proband and single parent) sequencing. However, there are scenarios where sequencing both parents is not possible. There are approximately 10 million single-parent families in the United States alone; another example are families with children born from egg or sperm donations. The inability to detect de novo

variants in these cases makes these families less likely to receive a diagnosis from genetic testing and benefit from precision treatments, thereby creating an equity issue.
To address this unmet challenge, we developed a method that enables identification of de novo variants from duo sequencing. Our method, called duoNovo, leverages recent improvements in the accuracy of long-read sequencing which permit the reliable detection together with phasing of single nucleotide variants. duoNovo uses blocks of phased single nucleotide variants spanning tens of kilobases in order to determine whether a candidate variant arose on the haplotype inherited from the sequenced parent - and is thus a de novo variant - or not.
We performed HiFi long-read sequencing on 16 trios, and applied duoNovo to each of the 32 duos. From the 16 father-proband duos, we tested an average of 518,313 candidate variants (reliable heterozygous phased variant call in the proband; reliable homozygous reference call in the father). We first identified de novo variants with duoNovo, and then examined the maternal sequence to evaluate duoNovo's performance. duoNovo identified an average of 30 (range 18-47) de novo variants and achieved an average precision (positive predictive value) of 92.2% (range 86 - 100%). As expected given the elevated contribution of the paternal germline to the pool of de novo variants, duoNovo detected on average 3.5 times as many de novo variants in father-proband duos compared to mother-proband duos. The precision in mother-proband duos remained satisfactory, albeit lower than in father-proband duos (average 73.4%; range 50-100%). Finally, we performed short-read sequencing on each trio to obtain a ground truth set of de novo variants, and found that *duoNovo* has a false negative rate of 0% in all of the 32 duos.
In summary, duoNovo is the first systematic method that can identify de novo variants from duos with high accuracy, and has the potential to increase the diagnostic yield of genetic testing for millions of families.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3053F: Downgrading Partial Duplications of Haploinsufficient Genes from VUS to Likely Benign

Authors:

K. Rudd[1], J. Zhao[1], Z. K. Lewis[1], B. Hong[1,2], E. F. Andersen[1,2]; [1]ARUP Lab., Salt Lake City, UT, [2]Univ. of Utah, Salt Lake City, UT

Abstract:

Copy number variation (CNV) is a major contributor to genetic disorders. Chromosomal microarray analysis (CMA) reliably detects intergenic and intragenic CNV but does not sequence breakpoints to determine the genomic structure of a variant. While deletion of a

haploinsufficient (HI) gene is relatively simple to interpret, duplications are more complicated. Breakpoint sequencing studies have shown that most interstitial duplications are in direct tandem orientation relative to the original locus and are predicted to leave the gene intact. However, duplications with one breakpoint in a HI gene raise the possibility for gene disruption and loss-of-function. Thus, without additional evidence to support or refute pathogenicity, many partial duplications of HI genes are classified as variants of uncertain significance (VUS). To better inform classification and reporting practices of these CNVs, we analyzed and catalogued duplications intersecting HI genes frequently observed by genomic microarray testing. To capture a large collection of partial gene duplications, we analyzed >50,000 germline duplications less than 5 Mb in our CMA database at ARUP. Data were collected from prenatal, postnatal, and oncology (germline only CNV) specimens analyzed on the CytoScan HD microarray. We searched for partial duplications involving genes classified by the ClinGen Dosage Sensitivity working group as HI=3, indicating sufficient evidence for haploinsufficiency. We find >600 partial gene duplications that have one breakpoint in a HI=3 gene. The most common partial duplications involve the genes ABCD1, TRIO, KMT2C, TRIP12, POGZ, MAGT1, SHOX, STS, ZMYND11, GRIN2A, MYT1L, CHEK2, NFIA, CHD2, and AUTS2. Partial duplications of the 5' or 3' ends of genes were equally frequent. We compared CNVs from our clinical dataset to DGV, ClinVar, gnomAD, and DECIPHER databases. Many partial gene duplications of HI=3 genes are similar across the population-based and clinical databases, suggesting they are likely benign (LB) variants that do not disrupt gene function. Though many duplications from our clinical population overlap clinically relevant HI=3 genes, they are not predicted to lead to loss-of-function. Partial duplications with common breakpoints, affecting the same exons, have discordant classifications of LB and VUS in data submitted to ClinGen. Recognizing these CNVs as common variants will pave the way to more consistent LB classification and reduce patient anxiety and follow up testing for VUS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3054F: Elucidating the diagnostic yield and allelic characteristics of FGF14 repeat expansion for adult ataxia through whole genome sequencing

Authors:

H. Dai[1,2], H. Du[2], C. Wu[1,2], W. Jin[1], K. Schoch[3], V. Shashi[3], N. Abreu[4], C. Steigerwald[4], J. Rucker[4], J. Cedeno[5], E. Moukheiber[6], L. Rosenthal[6], A. Butala[6], W. Mu[5,6], C. Chambers[7], K. Schulze[1,2], G. Sirugo[8], D. Rader[9], M. Risman[8], Y-A. Ko[8], A. Raper[8], E. Macnamara[10], C. Toro[11], L. Vossaert[1,2], L. Meng[1,2], P. Ward[1,2], C. Eng[1,2], Undiagnosed Diseases Network, F. Xia[1,2], B. Yuan[1,2]; [1]Baylor

Genetics, Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Duke Univ. Sch. of Med., Durham, NC, [4]NYU Grossman Sch. of Med., New York City, NY, [5]Johns Hopkins Univ. Genetic Med., Baltimore, MD, [6]Johns Hopkins Univ. Neurology, Baltimore, MD, [7]Univ. of Virginia Neurology, Charlottesville, VA, [8]Univ. of Pennsylvania, Philadelphia, PA, [9]Univ of Pennsylvania, Philadelphia, PA, [10]NIH, Bethesda, MD, [11]NIH, Bethesda, MD

Abstract:

Background: Adult ataxia, affecting 2.2 to 12.4 per 100,000 people, is a progressive neurodegenerative disorder with diagnostic yields below 35% [1]. Recent long-read sequencing has identified a GAA expansion in FGF14, causing late-onset spinocerebellar ataxia-27B (SCA27B), with occurrences ranging from 15 to 60% in different patient groups [1,2,3]. This large FGF14 repeat expansion poses detection challenges for short-read sequencing. We present molecular findings from a clinical whole genome sequencing (WGS) cohort and a general population cohort to further characterize the allelic architecture and phenotypic correlation of this genetic cause of adult ataxia. Method: We combined short tandem repeats (STR) bioinformatic calling and clinical ascertainment to prioritize potential positive cases. Using ExpansionHunter (EH), we performed STR calling on paired-end 150bp whole genome sequencing (WGS) at 40X coverage. STR results were correlated with phenotypes for index cases and confirmed using repeat-primed (RP) PCR and gel sizing. We also examined unsolved adult ataxia cases suspicious for SCA27B using HPO terms. Population characterization of FGF14 repeats was conducted on 265 unrelated individuals from our WGS database, and 202 individuals from a subgroup of the All of Us (AoU) v7 cohort using TRGT caller on PacBio HiFi long-read data with 5x coverage. Results: We identified 17 new adult ataxia cases for SCA27B, with GAA repeats ranging from 252 to 550 in ataxia patients aged 53 to 90 years. Other observed neurological phenotypes included unsteady gait (10), nystagmus (8), dysarthria (6), diplopia (6), sensory dysfunction (5), vestibular hypofunction (5), cerebellar atrophy (5), and dysphagia (5). For a single center sending for adult ataxia, FGF14 GAA expansion contributed to 29.6% (8 out of 27) of total cases, and 80.0% of diagnosed cases (8 out of 10). With underestimation of EH calling, it was estimated about 18.4% (49 out of 265) could have GAA expansion >100. Comparatively, a 320 GAA pathogenic repeat (0.5%) and a 160 repeat were identified in the AoU group. Conclusion: We used computational and clinical evidence to detect FGF14 expansion related to adult ataxia through WGS, RP PCR, and gel sizing. We also estimated its occurrence in the general population. The results showed a promising diagnostic yield for SCA27B due to the GAA expansion and suggested a high prevalence of FGF14 expansions. This underscores the importance of incorporating innovative methods into routine diagnostic workflows for adult ataxia, potentially leading to improved clinical management and targeted therapies.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3055F: Evaluating Dosage Sensitivity Predictions for Multigenic Copy Number Variants to Facilitate Clinical Interpretation

Authors:

E. Riggs[1], E. Spector[1], E. F. Andersen[2,3], J. Foreman[4], J. Herriges[5], A. W. Higgins[6,7], S-H. L. Kang[8], B. Levy[9], C. Lowther[10], C. L. Martin[1], L. Matyakhina[11], D. McMullan[12], D. Pineda-Alvarez[13], L. Shao[14], S. Shetty[15], A. Vaags[16,17], E. Thorland[18]; [1]Geisinger, Danville, PA, [2]ARUP Lab., Salt Lake City, UT, [3]Univ. of Utah, Salt Lake City, UT, [4]European Molecular Biology Lab., European Bioinformatics Inst., Hinxton, United Kingdom, [5]Children's Mercy Hosp., Kansas City, MO, [6]UMass Chan Med. Sch., Worcester, MA, [7]Quest Diagnostics, Malborough, MA, [8]Univ. of California, Los Angeles, Los Angeles, CA, [9]Columbia Univ. Irving Med. Ctr., New York, NY, [10]Massachusetts Gen. Hosp., Boston, MA, [11]GeneDx, Gaithersburg, MD, [12]Birmingham Women's and Children's Hosp., Birmingham, United Kingdom, [13]Invitae Corp., San Francisco, CA, [14]Univ. of Michigan, Ann Arbor, MI, [15]Case Western Reserve Univ. and Univ. Hosp., Cleveland, OH, [16]Trillium Hlth.Partners, Mississauga, ON, Canada, [17]Univ. of Toronto, Toronto, ON, Canada, [18]Mayo Clinic, Rochester, MN

Abstract:

In the 2020 ACMG/ClinGen Technical Standards for Evaluation of Constitutional Copy Number Variants (CNVs), a small amount of supporting evidence may be applied during a CNV evaluation if multiple predictors agree that a single gene within that CNV is dosage sensitive (DS). It is currently unclear if additional weight should be applied if multiple genes within the CNV meet this criteria. The newly reconvened ACMG/AMP/CAP/ClinGen CNV technical standards working group sought to evaluate the performance of a previously published method (Huang et al. 2010) to generate LOD scores based on the predicted DS of multiple genes within a CNV. This method has been built into the DECIPHER platform (www.deciphergenomics.org), and is available to registered users. DECIPHER produces an overall DS score for the CNV and compares it to DS scores calculated for CNVs observed in a control population. Utilizing this information, the CNV under evaluation can be categorized by its sampling probability (SP), an estimate of the proportion of the general population that carries a rare CNV with a score as severe or more severe than its own. Those CNVs with the most severe scores have SPs at <1%, and those with the least severe scores fall within the >5% category.

To assess the performance of this method, we generated DS scores for a total of 338 CNVs: 144 genomic regions previously curated by ClinGen, and 194 CNVs classified by clinical laboratories. We hypothesized that CNVs with either ClinGen curation results of “sufficient evidence of

dosage sensitivity" or clinical classifications of pathogenic (P) ("known P," n= 97) should have SPs of <1%, while those with ClinGen curation results of "dosage sensitivity unlikely" or clinical classifications of benign (B) should have SPs of >5% ("known B," n= 23). Classifying true positives as known P with SP <1% (n=74), false positives as known B with SP <1% (n= 0), false negatives as known P with SP >5% (n=7), and true negatives as known B with SP >5% (n=21), we determined that the sensitivity of this metric is 91%, specificity is >99.9%, positive predictive value is >99.9%, and the negative predictive value is 75%.
We conclude that this method for accounting for the DS of multiple genes within a CNV shows promise. The group is currently independently evaluating all laboratory-reported classifications to ensure concordance for known P and B CNVs. We will also evaluate all uncertain laboratory CNVs (n= 89) to determine if the predictor can distinguish between those trending toward P or B. Future directions also include performing these analyses on large sets of publicly-available CNV data from the general population to better inform future SP estimates.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3056F: Genetic characterization and diagnostic yield of next-generation sequencing in Colombian patients with suspected skeletal dysplasias

Authors:

S. Bello Uyaban[1], Y. Suarez Useche[1], C. Baquero Zapata[1], E. Chacón Gómez[1], G. Gordillo Gonzalez[2], J. Rojas[3], M. Galvez[4]; [1]Genuino Res. Group- Gencell, Bogotá D.C., Colombia, [2]Servicio De Genética Sociedad de Cirujanos y pediatras, Barranquilla, Colombia, [3]Hosp. UNIVERSITARIO SAN IGNACIO, Bogota, Colombia, [4]Genuino Res. Group- Gencell, Bogota, Colombia

Abstract:

Introduction: Skeletal dysplasias constitute a heterogeneous group of genetic disorders with a prevalence of 2.4 cases per 10,000 births. In Colombia, a LatinAmerican country, genotypic characterization studies are limited, which is essential for identifying the diagnostic and healthcare needs of the affected population.**Objectives:** To describe the prevalence of pathogenic (P) and likely pathogenic (Lp) variants causing skeletal dysplasias and to evaluate the diagnostic yield of next-generation sequencing (NGS) panels in a group of patients with clinical suspicion of these conditions. **Methodology:** Cross-sectional observational study including patients with NGS panel targeting causal genes of skeletal dysplasias. Frequencies of clinical features as well as of affected genes were calculated. Models of the mutant protein for new variants, were generated using the RoseTTAFold method integrated into the Robetta server, and

visualization was performed in ChimeraX software to assess the structural effects of the new variant, analyzing changes in the secondary and tertiary structure. **Results:** Sixty subjects were included, with a median age of 7 years (IQR: 2-19.5), 56.67% of whom were male. Postnatal short stature was the most common clinical feature (88.33%), followed by facial dimorphism (40%) and other skeletal abnormalities of the spine and/or limbs. Additionally, 53.33% of the subjects presented positive radiological findings and 28.33% reported previous non-diagnostic molecular or cytogenetic studies. The NGS panel revealed a diagnostic yield of 30%. *SLC26A2* gene was the most frequently affected (33.33%), followed by *FGFR3, MMP13, TRPV4* (11.1% each), and *ANO5, COL2A1, COMP, CUL7, FGFR2, PEX7* (5.55% each). Variants were 77.27% missense, 13.63% nonsense, and 4.54% splicing site and in-frame variants. Since the c.1798G>A variant in the *TRPV4* gene has not been previously reported in the literature, protein modeling was performed to understand the structural effects showing that the protein is affected. **Conclusions:** Pathogenic or likely pathogenic variants associated with the diagnosis of skeletal dysplasias were detected in 30% of the cases, with the c.835C>T; p.Arg279Trp variant in the *SLC26A2* gene identified in 33.33% of the diagnosed patients. This is the first study to conduct a genetic characterization of skeletal dysplasias in the Colombian population.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3057F: Genetic insights into spina bifida: a study of 205 families

Authors:

A. Alkelai[1], J. Backman[1], J. Gleeson[2], A. Shuldiner[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Univ California, La Jolla, CA

Abstract:

The most frequent central nervous system congenital defect is spina bifida (SB). In the United States, the estimated incidence rate of SB is 3.63 per 10,000 live births. SB is considered to be a complex trait that involves a number of genes, environmental factors, and combinations of variants at different loci, as evidenced by multiple sporadic cases and the lack of clear Mendelian inheritance in families. Although prenatal folic acid supplementation has resulted in a significant reduction in the incidence of SB, the clear genetic and environmental drivers of SB remain unknown. Exome sequencing (ES) is a well-established and important clinical tool for pediatric illness diagnosis and management. Recent ES studies and candidate gene studies have discovered several SB candidate genes. Although the clinical utility of ES in Mendelian diseases has been well established, its value in the context of SB diagnosis and management is unclear. The goal of this study is to find clinically significant single nucleotide variants (SNV) and

insertions or deletions (indels) that cause SB in probands. This study presents ES in 205 families with a proband diagnosed with a severe form of SB. By utilizing a trio/duo/singleton ES approach and internal rare-disease analysis pipeline we discovered 50 pathogenic/likely pathogenic candidate variants in 42 cases out of 205 (20.5%). *TBXT, BMP4, FOLR2, VRTN, AGO2* and *SHROOM3* were among the implicated genes. These SNVs and indels were inherited (20) or *de novo* (30). Five of the probands were found to carry high confidence ClinVar pathogenic or likely pathogenic variants in genes causing known monogenic syndromes that have not been previously associated with SB: *SIN3A, IFIH1, GNAS, COL7A1, TPM2,* potentially broadening the phenotypic spectrum of these conditions. Most of the candidate variants were not found in any external or internal databases. Taken together, ES could be considered as a valuable tool in the genetic workup for SB patients. To better unravel the broad genetic architecture of SB and to validate ES findings, large-scale sequencing of cohorts from various ethnic groups is required.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3058F: Hidden yet dangerous: searching for occult Y chromosome in patients with Turner syndrome.

Authors:

M. Ponikowska[1,2], M. Wójcik[3], A. Maciejewska[4], A. Ulenberg[5], M. Jankowski[1], N. Filipowicz[3], K. Drężek-Chyła[3], E. Rychlicka-Buniowska[3], D. Birkholz-Walerzak[5], R. Pawłowski[4], M. Koczkowska[3], M. Mysliwiec[5,6], B. S. Lipska-Ziętkiewicz[7,2]; [1]Dept. of Biology and Med. Genetics, Med. Univ. of Gdansk, Gdansk, Poland, [2]Rare Diseases Ctr., Med. Univ. of Gdansk, Gdansk, Poland, [3]3P-Med. Lab., Med. Univ. of Gdansk, Gdansk, Poland, [4]Dept. of Forensic Med., Med. Univ. of Gdansk, Gdansk, Poland, [5]Dept. of Pediatrics, Diabetology and Endocrinology, Univ. Clinical Ctr. in Gdansk, Gdansk, Poland, [6]Dept. of Pediatrics, Diabetology and Endocrinology, Med. Univ. of Gdansk, Gdansk, Poland, [7]Clinical Genetics Unit, Dept. of Biology and Med. Genetics, Med. Univ. of Gdansk, Gdansk, Poland

Abstract:

Background: Turner syndrome is an aneuploidy resulting from the complete or partial loss of one of the sex chromosomes, X or Y, in phenotypic females. The loss might occur in a proportion of cells, and half of the patients exhibit mosaicism. The presence of Y chromosome in females increases the risk of gonadoblastoma, prevalence of which is comparable in individuals with the overt and cryptic Y chromosome. Standard karyotyping is often insufficient, necessitating novel approaches. **Objective:** This study examines the potential of ddPCR and STR analysis to detect

the cryptic Y chromosome. ddPCR quantifies selected X and Y chromosomal loci enabling the assessment of mosaicism level. STR analysis offers a comprehensive view of the entire Y chromosome at picogram levels. **Method:** The study involved 25 Turner syndrome patients: 14 with a 45,X karyotype, 11 mosaic for a cell line with a second X chromosome or its derivative, and a positive control patient mosaic for Y chromosome. DNA was isolated from peripheral blood, buccal swabs and urine. ddPCR was designed to target *AMELX* and *AMELY* (blood, buccal swabs, urine) and *SRY* in a separate assay (blood, buccal swabs) with the QX200 Droplet Digital PCR System (Bio-Rad). Data analysis used QuantaSoft. STR analysis (blood, buccal swabs, urine) employed the PowerPlex® Y23 System (Promega, USA) for 23 Y chromosome loci with the detection of products by ABI 3130 Genetic Analyzer (Applied Biosystems, USA) and analysis in GeneMapper ID. **Results:** In 1 of 25 patients ddPCR revealed presence of *SRY* gene in the blood and the buccal swab while STR analysis confirmed the presence of all 23 Y chromosome loci in the blood, buccal swab and urine. ddPCR revealed *AMELY/AMELX* signal ratios of 0.22 (blood), 0.40 (buccal swab), 0.10 (urine). In the positive control patient, with karyotype 45,X[20]/46,Xididc(Yp)[8], the *AMELY/AMELX* ratios were 1.18 (blood), 0.60 (buccal swab), 0.89 (urine). **Conclusion:** Both ddPCR and STR analysis proved effective in detecting the occult Y chromosome and allowed for gonadoblastoma risk restratification. Furthermore, ddPCR allowed for detection of varying levels of mosaicism across tissues. Standard karyotyping (30 metaphases) failed to detect Y chromosome in the patient, even though, according to ddPCR, it was present in every fifth cell.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3059F: Integrating algorithms to optimize clinical identification of structural variation using whole-genome sequencing

Authors:

L. Pizzo[1,2], T. Lewis[2], A. Bolia[2], C. Krueger[2], B. O'Fallon[2], A. Kellogg[2], E. Fredrickson[2], E. Andersen[1,2], H. Best[1,2]; [1]Univ. of Utah, Sch. of Med., Salt Lake City, UT, [2]ARUP Lab., Salt Lake City, UT

Abstract:

Recent cost reductions and increased utility of whole-genome sequencing (WGS) have led to its widespread adoption in clinical laboratories for identifying structural variants (SV) like deletions, duplications, and to a lesser extent, balanced rearrangements (e.g., inversions, insertions, translocations). However, there is limited consensus as to which SV detection algorithm is most sensitive in a clinical setting. Cytogenetically visible SVs and small CNVs are highly prevalent in

newborns with congenital anomalies, highlighting the relevance of their identification using clinical WGS in neonatal and pediatric patients. Here, we compared the efficacy of different algorithms in identifying a diverse group of clinically relevant SVs using WGS. We performed PCR-free short read WGS at an average depth of 45X on sixteen samples carrying 27 SVs previously identified by our cytogenetics and genomic microarray clinical service. Tested SVs represented a range of sizes (6kb to aneuploidies), mosaic and non-mosaic findings, as well as structural changes and genomic regions of varying complexity. DRAGEN exhibited the best performance, identifying 26/27 (96%) positive events, followed by the ensemble method integrating six SV algorithms, Parliament2 (24/27, 89%), GATK (23/27, 85%), and MANTA (8/27, 30%). DRAGEN, GATK, and Parliament2 successfully detected recurrent CNVs (including 8p23.1 deletion, 16p11.2 duplication, and DiGeorge syndrome 22q11.2 deletion); however, only DRAGEN identified all tested aneuploidies (trisomy 21, 47,XYY, and 47,XXY). Small variants were more challenging to detect, including a 6kb 11p11.2 deletion, not identified by any algorithm, and a 25 kb deletion encompassing 3' exons of *EPCAM* and exons 1-6 of *MSH2*, accurately identified solely by DRAGEN. None of the callers precisely estimated mosaic copy-number, presenting a significant limitation. We evaluated the orientation of duplications detected in these samples using read pair features as structural information aids in the interpretation of partial gene duplications. 25/28 duplications showed direct orientation. Examples of inverted duplications included a 49 kb Xq28 duplication that was part of a dup-trip-dup event, and a terminal duplication in a recombinant chromosome 9, solely identified by Parliament2. Our results underscore the importance of combining multiple algorithms in clinical WGS and highlight that SV patterns indicative of chromosomal rearrangements should be visualized cytogenetically for accurate interpretation, direct visual inspection of abnormal read pair orientation/distance, and to offer appropriate guidance for follow-up testing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3060F: Karyotype findings in recurrent pregnancy loss: a path to better outcomes

Authors:

M. Kanjee, Z. Bulduk, C. Yuce Kahraman, A. Tatar; Atatürk Univ., Faculty of Med., Dept. of Med. Genetics, Erzurum, Turkey

Abstract:

Background/Objectives: According to the American Society for Reproductive Medicine (ASRM), recurrent pregnancy loss (RPL) is defined as two or more spontaneous pregnancy losses. Several

studies conducted in various countries reported that chromosomal abnormalities (CA) were detected in 4-10% of couples with RPL. Conventional karyotype analysis is a valuable tool for detecting such CA in one or both partners with RPL. In this study, we aim to identify the prevalence and types of CA in couples with RPL who underwent karyotype analysis in our clinic and to evaluate their implications for genetic counseling and clinical management. **Methods:** The karyotype findings of 258 couples with a history of RPL who underwent karyotyping using the G-banding technique at our clinic between January 2023 and April 2024 were recruited and evaluated retrospectively. The results were reported according to the International System for Human Cytogenomic Nomenclature (ISCN). Data were analyzed to determine the frequency and types of chromosomal abnormalities detected by karyotyping. **Results:** A total of 520 peripheral blood samples were analyzed. The median ages of males and females were 28 (18-45) and 31 (18-60), respectively. Chromosomal abnormalities were identified in 4.7% (12/258) of couples, with 8 (8/12; 66.7%) being female and 4 (4/12; 33.3%) being male. Among those 12 patients, balanced translocations were the most common CA, accounting for 91.7% (11/12) of the detected abnormalities: 8 Reciprocal and 3 Robertsonian translocations. Only one patient carried a numerical abnormality; a mosaic Turner syndrome (1/12; 8.3%). **Conclusion:** Our findings reemphasize the importance of karyotype analysis in detecting such abnormalities since it is vital for affected couples to receive comprehensive genetic counseling, including the consideration of preimplantation genetic testing (PGT) and/or prenatal genetic testing if a CA has been detected. This comprehensive approach is crucial to ensure better psychological and emotional outcomes by providing clarity and reducing uncertainty. Although our study only addresses the karyotype results of these patients, it's important to note that additional diagnostic tests such as microarray, fluorescence in situ hybridization (FISH), and other molecular assays were recommended for some patients. Our study underscores the indispensable role of karyotype analysis in the investigation of RPL, given the low but non-negligible prevalence of CA in these patients. Moreover, these findings provide valuable insights for genetic counseling and emphasize the role of karyotype analysis in guiding reproductive planning and management.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3062F: Long-read whole genome sequencing increases diagnostic yield for short-read negative pediatric sensorineural hearing loss patients ⭐

Authors:

S. Redfield[1], W. Shao[1], C. French[1], C. Davidson[2], A. Halpern[2], A. Pastolero[1], A. Carroll[2], R. Bombardi[2], S. Rockowitz[1,3], M. Kenna[1,4], E. Shearer[1,4]; [1]Boston Children's Hosp., Boston, MA, [2]Illumina, Inc, San Diego, CA, [3]Div. of Genetics, Boston Children's Hosp., Boston, MA, [4]Dept. of Otolaryngology-Head and Neck Surgery, Harvard Med. Sch., Boston, MA

Abstract:

Identification of an etiologic diagnosis for pediatric sensorineural hearing loss (SNHL) has crucial implications for clinical care and counseling. Current genetic diagnostic yields for pediatric SNHL using gene panels, exome sequencing, and short read genome sequencing (srGS) are approximately 40% across audiologically and demographically diverse cohorts. These short-read sequencing methodologies limit the ability to detect important variants in inaccessible genomic regions, such as areas of high homology or highly repetitive regions. We piloted the use of long-read genome sequencing (lrGS) via Illumina Complete Long Read sequencing (ICLR; Illumina, San Diego, CA) to assess for a genetic diagnosis for pediatric SNHL. DNA was derived from buccal or whole blood samples for pediatric probands with previously negative exome or srGS. Mapping and variant calling, including copy number variants, was performed with the DRAGEN Bio-IT Platform (Illumina). Primary variant analysis focused on 366 known and candidate hearing loss genes. Variants suspected to be causative were clinically confirmed using an orthogonal method in a CLIA-certified environment (Prevention Genetics, Marshfield, WI). In a pilot study, 10 probands underwent genomic sequencing via ICLR. Median depth of coverage was 37x and median read length was 3.5 kb. LrGS via ICLR identified causative variants in 3 of10 probands tested (30%). LrGS identified a heterozygous loss of *STRC* (previously identified) in trans with a pathogenic single nucleotide variant in a child with bilateral mild to moderate SNHL. A child with unilateral moderate SNHL and blond hair was identified to have a copy-number neutral inversion upstream of *MITF* predicted to impact alternatively spliced transcripts. A third proband was identified to have a small copy-number-loss in the *OTOA* gene in trans to a predicted deleterious missense variant. Given the success of the pilot study, LrGS data for sixty additional pediatric SNHL patients is underway, and data will be delivered in approximately July, 2024 and available for presentation. ICLR had a 30% diagnostic yield in this pilot cohort of 10 probands with previously nondiagnostic short read sequencing. Diagnoses in the *STRC* and *OTOA* genes, which have a highly homologous pseudogene and lie within a segmental duplication, respectively, highlight the strength of lrGS at detecting variants in inaccessible genomic regions. This pilot study emphasizes the promise of long-read sequencing to improve genetic diagnosis beyond current benchmarks in the setting of SNHL.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3063F: Microduplication in chromosomal region 2p16.1p15: Further characterization of the phenotype

Authors:

J. Kapalanga[1], T. Parikh[2], C. Lefebvre[1], R. Kamran[3]; [1]Western Univ./GBHS, London, ON, Canada, [2]Western Univ., London, ON, Canada, [3]McMaster Univ., Hamilton, ON, Canada

Abstract:

Over half a dozen cases of chromosome 2p16.1p15 microduplication are reported between 2015 and 2023. This charter group with the 2p16.1p15 chromosomal abnormality share a certain phenotype characterized by developmental delay in motor skills, in speech and language, intellectual deficit, and facial dimorphisms. In addition, subsets of patients show three or four additional clinical features that include head shape/head size anomalies, limb anomalies, cardiac defects, ocular defects, and neurological disorders, in different combinations. So far, the majority of the reported cases are male; only one reported case is female. The size of the microduplication varies from 1.66Mb to 8.49 Mb. OMIM genes located in this genomic region include *BCL11A, FAM161A, EHBP1, XPO1, COMMD1, CCT4, PEX13, VRK2, FANCL, PAPOLG, REL, PUS10, KIAA1841, USP34, B3GNT2, TMEM17*, and *OCT4*. Microduplication occurs de novo or is inherited from a parent. Regardless of the size of the microduplication, all affected cases presented with developmental delay including gross and fine motor delay, language delay, intellectual deficit, and facial dimorphisms. Here we report a case of a male patient who presented at birth with respiratory distress necessitating resuscitation. During infancy, the patient was hypotonic, and developed torticollis and plagiocephaly. Between infancy and age 4 years the patient was noted to have dysmorphic facial features including a prominent occiput, down slanting palpebral fissures, epicanthic folds, clinodactyly of the toes and hands, first and second toe diastasis. The patient was developmentally delayed in motor skills, and speech and language. An MRI showed gray matter heterotopia. A microarray done at 22 months of age showed a copy number variant with a gain of 8.49 Mb in chromosomal region 2p6.3p16.1 involving OMIM morbid genes FAM161A, EHBP1, *CCDC88A, PNPT1, EFEMP1, PEX13, REL, FANCL PEX13 and BCL11A.* The chromosomal abnormality was of maternal origin. Mom also had mild to moderate intellectual deficit and a history of developmental delay during childhood. The child's facial dysmorphism was reflected in mom. The findings in this case provide further characterization of the phenotype spectrum and the genes involved.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3064F: Molecular diagnostic findings of 166 patients with Pachyonychia Congenita ⭐

Authors:

S. Rathod[1], C. Neckelmann[1], Z. Chen[1], D. Hansen[2]; [1]Fulgent Genetics, El Monte, CA, [2]Intl. Pachyonychia Congenita Res. Registry, Holladay, UT

Abstract:

Pachyonychia congenita is a dermatologic disease characterized by primary findings of painful palmoplantar keratoderma and hypertrophic dystrophy in one or more nails. The phenotypic spectrum also includes oral leukokeratosis, steatocystomas, pilosebaceous cysts, and follicular hyperkeratosis on friction areas. The severity of these findings may vary among members of a single family and between different families with the same pathogenic variants. Within the clinical setting, the diagnosis of pachyonychia congenita is further confounded by the underlying molecular complexities associated with the *KRT6A*, *KRT6B*, and *KRT6C* genes. Variants within these 3 genes comprise a significant portion of pachyonychia congenita cases, and these genes all share a high degree of sequence homology. As such, pathogenic/likely pathogenic variants within these genes may not be properly detected by short read NGS analysis methods unless paired with multi-mapping, long-range PCR, and/or long read sequencing methods; workflows without these orthogonal methods may be prone to false negatives should variants occur below thresholds of detection in unique vs non-unique reads. In collaboration with the Pachyonychia Congenita Project, we describe the findings from a custom gene panel targeting 62 known pachyonychia congenita genes. Using a specialized NGS workflow with long-range PCR confirmation, 110 patients were observed with positive diagnostic variants, 34 patients were observed with unresolved variants of uncertain significance, and 22 patients were observed with negative findings. Identifying the underlying molecular diagnosis of pachyonychia congenita allows for appropriate disease management and empowers families by clarifying reproductive risks. These findings also inform pretest counseling for pachyonychia congenita, and most importantly, enable patient enrollment in ongoing clinical trials aimed at discovering new treatments for this condition.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3065F: Multicolor metaphase FISH analyses using selected BAC clones after CNV detection in a patient revealed that two types of chromosomal abnormalities involving chromosomes 16: a balanced translocation [t(16q;20q)pat], and a submicroscopic unbalanced translocation [der(22)t(16p;22q)dn].

Authors:

K. Wakui[1,2], K. Takano[2,1], R. Takeda[3,4], T. Kosho[1,2], Y. Fukushima[1,2]; [1]Dept. of Med. Genetics, Shinshu Univ. Sch. of Med., Matsumoto, Japan, [2]Ctr. for Med. Genetics, Shinshu Univ. Hosp., Matsumoto, Japan, [3]Dept. of Med. Genetics, Nagano Children's Hosp., Azumino, Japan, [4]Life Sci. Res. Ctr., Nagano Children's Hosp., Azumino, Japan

Abstract:

Background: Chromosomal microarray analysis (CMA) is useful for detecting CNVs in patients with birth defects of unknown cause. The limitations of not providing information on chromosomal rearrangements associated with CNVs are well known but must be re-acknowledged.
Materials and Methods: A 2-year-old girl with developmental delay, short stature, and characteristic facial features was referred to our hospital for diagnosis and genetic counseling. The patient had been identified as having a paternally inherited balanced translocation by G-banding [t(16q;20q)pat] at the previous hospital. First, we performed CMA on the patient. After detection of the CNVs in the patient, multicolor metaphase FISH analyses using selected BAC clones were performed on the patient and both parents to confirm the chromosomal rearrangements associated with the CNVs.
Results and Discussion: The CNVs detected in the patient suggested a terminal gain of approx. 2 Mb at 16p, and suggested a terminal loss of approx. 1.6 Mb at 22q (incl. *SHANK3*) [arr[hg19] 16p13.3(102839_2146889)x3, 22q13.33(49562171_51178149)x1]. The clinical significance of 16pter triplosensitivity is currently uncertain. However, the patient's symptoms were thought to be consistent with haploinsufficiency of *SHANK3*. To confirm whether the 16p copy number gain and the 22q copy number loss detected by CMA were derived from unbalanced translocations or not, several BAC clones were selected as probes within and outside each CNV region, and multicolor probe mixes were prepared to confirm the chromosomal rearrangements. Since the two CNVs identified in the patient were found to be due to rearrangement of an unbalanced translocation. Then we also checked whether either of the parents was a balanced carrier. As neither of the parents had the submicroscopic balanced translocation t(16p;22q), it was confirmed that the unbalanced translocation in the patient was de novo [ish der(22)t(16p;22q)dn].

Conclusion: It is important to interpret detected CNVs while checking the whole chromosome ideogram, in order to assume chromosomal rearrangements related to CNVs and not overlook the possibility of unbalanced translocations. Multicolor metaphase FISH analyses using multiple appropriate probe mixes are essential to confirm chromosomal rearrangements associated with CNVs. When a patient's CNVs form an unbalanced translocation, it may be de novo or may be due to a balanced translocation in either parent, then metaphase FISH analyses of both parents are essential for genetic counseling. Basic knowledge deriving from classical cytogenetics and molecular cytogenetics is essential for proper interpretation of CMA results.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3066F: Multiple bioinformatic tool strategy for implementation of non-coding variant analysis in the genome diagnostics

Authors:

K. Abbott, S. Jansen, B. Charbon, L. Johansson, D. Hendriksen, B. Sikkema-Raddatz, J. Jongbloed; Univ. Med. Ctr. Groningen, Groningen, Netherlands

Abstract:

Diagnostic rates for solving rare genetic diseases have plateaued at around 30-40% using current whole exome sequencing techniques. As whole genome sequencing replaces WES for clinical diagnostics, the non-coding regions can also be explored. Studies showing how variants in regulatory elements such as promoters and enhancers, UTRs and intronic regions impact disease are on the rise, but identifying these high-impact variants remains a challenge, especially in a diagnostic setting. While many bioinformatic tools and algorithms can be used to chip away at the mountain of data, they return too many candidate variants to make them useful on their own. Therefore, the aim here is to create a filter strategy combining existing tools to significantly reduce the number of non-coding variants remaining for detailed interpretation and follow-up, focusing on variants in the regulatory elements and on splicing candidates.
WGS data from one positive control case with a known deep-intronic splicing variant, as well as from ten individual unsolved cases were examined. Consent was obtained for all individuals for personalized extended examination of their NGS data. All sequence variants passed pipeline quality measures and had a gnomAD population frequency less than 0.01%. Variants in promoter/enhancer, intronic and UTR regions meeting the significance thresholds for at least one of the FATHMM-MKL, ReMM and ncER algorithms were retained. VIP (in-house) and GREEN-VARAN (G-V) (open-source) tools were tested separately. Additionally, VIP uses SpliceAI

and UTRAnnotator algorithms to identify splice and 5'UTR variants, and G-V uses an internal high-impact constraint score. Finally, filtering using a panel of 5045 disease associated genes was performed, a tool consensus step was applied, and a final curation of variants using clinical phenotypes was done.
Using VIP resulted in 1321±323 variants of interest (VOI) and using the G-V tool resulted in 138±46 VOIs. Application of gene panel filtering reduced the number of VOIs to 21-51 for VIP, and 4-14 for G-V. The tool consensus approach (allowing for additional splice and 5'UTR variants from VIP only) reduced the number of VOIs to 5-17. Lastly, clinical curation led to 0-3 variants per individual. Notably, the positive control intron variant was only prioritized using VIP (SpliceAI). Follow-up includes parental inheritance control and functional testing: RNAseq to confirm spicing of VOI in *TLK2*, and methylation pattern testing for *ZEB2* enhancer VOI.
Our results demonstrate that applying a diagnostic-aimed filter strategy incorporating existing bioinformatic tools makes diagnostic non-coding variant analysis feasible.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3067F: PCH Gene Kids: Comprehensive WGS-based healthcare for the Intermountain West

Authors:

M. Yandell[1], B. Moore[1], E. Clark[2], S. Andrews[2], B. J. Shayota[3], R. Palmquist[2], S. Malone Jenkins[4], B. Peterson[5], E. F. Juarez[6], E. Javier Hernandez[5], E. Frise[7], C. Hobbs[8], M. N. Bainbridge[6], M. Wright[6], K. Ellsworth[6], M. DiStefano[9], S. Hofherr[9], H. L. Rehm[9], S. Kingsmore[6], M. G. Reese[7], M. Tristani-Firouzi[10], J. L. Bonkowsky[2]; [1]Dept. of Human Genetics, Utah Ctr. of Genetic Discovery, Univ. of Utah, Salt Lake City, UT, [2]Div. of Pediatric Neurology, Primary Children's Ctr. for Personalized Med., Intermountain Hlth.Care. Dept. of Pediatrics, Univ. Hosp., Univ. of Utah, Salt Lake City, UT, [3]Div. of Med. Genetics, Dept. of Pediatrics, Univ. of Utah, Salt Lake City, UT, [4]Div. of Neonatology, Dept. of Pediatrics, Univ. of Utah, Salt Lake City, UT, [5]Dept. of BioMed. Informatics, Univ. of Utah, Salt Lake City, UT, [6]Rady Children's Inst. for Genomic Med., San Diego, CA, [7]Fabric Genomics, Oakland, CA, [8]Dept. of Pediatrics, UCSD, San Deigo, CA, [9]Broad Clinical Lab., Broad Inst. of MIT and Harvard, Cambridge, MA, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [10]Nora Eccles Harrison Cardiovascular Res. and Training Inst., and Div. of Pediatric Cardiology, Salt Lake City, Utah, 84112, USA Div. of Neonatology, Dept. of Pediatrics, Univ. of Utah, Salt Lake City, UT

Abstract:

PCH Gene Kids (Primary Children's Hospital Genomic Initiative for Children) is a new clinical-research initiative of the Primary Children's Hospital (PCH) Center for Personalized Medicine (CPM) and the University of Utah (UU). PCH cares for over 1.7 million children across the Intermountain West, the largest pediatric catchment area in the nation. PCH Gene Kids' goal is to leverage tools and services developed by Rady Children's Institute for Genomic Medicine (RCIGM), Broad Clinical Laboratories, and Fabric Genomics to enable comprehensive WGS-based healthcare for its entire pediatric patient population**.** In its pilot phase (beginning July 1, 2024), PCH Gene Kids anticipates sequencing 1,500+ in- and out-patients per year, including ICU and undiagnosed infants, children, and adolescents. A key aspect is reaching children in rural, frontier, and underserved populations. Achieving this goal will require an integrated infrastructure for genomic and EHR-based disease discovery, diagnosis, and treatment. At these scales, AI must necessarily play an essential role. Patient recruitment will be accomplished using the Mendelian Phenotype Search Engine (MPSE), which can continuously surveil an entire health system, automatically identifying patients for sequencing using clinical notes. Ultrarapid and rapid sequencing and analyses will be provided by RCIGM and Broad laboratories, respectively, using GEM, an AI developed by Fabric Genomics and the UU that assists clinical geneticists with WGS diagnosis. Critically, GEM can directly ingest the outputs of MPSE, providing integrated means for communicating EHR contents to the downstream AIs, clinical geneticists, and physicians overseeing diagnostic reporting. Fabric Genomics will coordinate these analyses, including both periodic and on-demand physician-initiated re-analyses. Finally, Transformer, an extension of the GEM platform for automatic (re)interpretation of WGS results will be used to triage cases. Our ultimate goal is to provide a model for delivery of pediatric genomic medicine. We utilize telegenetics to reach and facilitate patient enrollment. Preliminary benchmarks using a multi-institutional dataset of over 1000 WGS cases indicate that our AI-based workflow can identify and dramatically speed nomination of patients for WGS while maintaining diagnostic rates. Diagnostic turnaround times are likewise accelerated by the AI, with clinical review of a median of 3 diagnostic gene-candidates per proband achieving > 90% True Positive rates. Given the scale of PCH Gene Kids, we anticipate reporting analysis results for a sizable prospective cohort by the meeting.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3068F: Research genomics strategies to end the diagnostic odyssey for challenging rare disease patients ⭐

Authors:

D. Koboldt[1], S. Ramadesikan[2], M. Marhabaie[1], U. Gurusamy[2], I. Showpnil[2], E. Varga[2], L. Dublin-Ryan[1], A. Daley[1], C. Colwell[1], S. Goldson[1], A. Antoniou[1], J. Hunter, M[1], M. Leung[1], R. Stottmann[3], B. Chaudhari[2], S. Hickey[4], P. White[5], R. Wilson[6]; [1]Nationwide Childrens Hosp., Columbus, OH, [2]Nationwide Children s Hosp., Columbus, OH, [3]Nationwide Childrens Hosp./ Ohio State Univ., Columbus, OH, [4]Nationwide Children s Hosp, Columbus, OH, [5]Nationwide Children's Hosp., Columbus, OH, [6]Nationwide Children's Hosp. Inst. for Genomic Med., Columbus, OH

Abstract:

Rare genetic disorders are individually uncommon in the general population, but collectively affect 25 million patients in the United States and 300 million worldwide. Despite decades of discovery and technological advances in the field of genetics, ~70% of patients remain undiagnosed after exome or genome sequencing. Such patients continue on a diagnostic odyssey that can last years, creating a significant burden on patients, families, and healthcare systems. For the last eight years, we have enrolled more than 1,100 individuals from 350 families with undiagnosed rare diseases into an IRB-approved, multi-disciplinary translational research study. We apply an evolving and diverse set of approaches to make diagnostic findings, resolve variants of uncertain significance, and identify novel candidate genes. A pilot study within this project of 150 negative research-consented exomes performed >1 year ago demonstrated that reanalysis alone uncovered clinically reportable variants in ~18% of cases, primarily due to new gene-disease associations. Most of our families undergo additional genomic studies on a research basis. The primary assay is genome sequencing (212/250 families), often in combination with RNA-seq (n=33). However, we also apply deep exome sequencing (n=16), RNA-seq only (n=14), or long-read single molecule sequencing (n=8). These additional studies yielded diagnostic findings in 31.6% of cases and compelling candidate variants in another 13.6%. The majority of the diagnostic variants (74%) were single nucleotide variants (SNVs) or small insertions/deletions. Many of these were *de novo* or dominantly inherited variants in emerging disease genes such as ATP2B1, AUTS2, CDX2, EEF1A2, GRIA2, PPP1R10, and SETD1B. Around 25% of diagnostic variants in splice-disrupting variants or structural variants. These are the challenging variants often missed by standard genetic testing modalities. Our multi-omics approach uncovered missing “second hits” due to deep intronic variants (TTN, ATP7B, IGHMBP2), synonymous splicing variants (YIF1B, SEPSECS), and small structural variants (FXN, DNAJC19). Notably, 52% (34/65) of our findings were made possible by international collaborations facilitated through GeneMatcher and Matchmaker Exchange, highlighting the power of genomic data sharing to accelerate novel gene discovery and elucidate gene-disease relationships. In summary, our work highlights the power of research genomics to discover new disease genes and end the odyssey for families with rare undiagnosed conditions.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3069F: Searching for answers: Continued utility of dynamic re-analyses in rare disease diagnostics

Authors:

A. Cohen[1,2,3], J. J. Johnston[1], M. I. Gibson[1], J. Alaimo[1,2,3], J. Herriges[1,2,3], C. Saunders[1,2,3], E. G. Farrow[1,2,3], I. Thiffault[1,2,3], T. Pastinen[1,3,4]; [1]Genomic Med. Ctr., Children's Mercy Kansas City, Kansas City, MO, [2]Dept. of Pathology and Lab. Med., Children's Mercy Kansas City, Kansas City, MO, [3]UKMC Sch. of Med., Univ. of Missouri Kansas City, Kansas City, MO, [4]Dept. of Pediatrics, Children's Mercy Kansas City, Kansas City, MO

Abstract:

Improved cost-efficiency of short-read (SR) sequencing shifts current challenges in rare disease diagnostics to test accessibility, technological blindspots, and systems for reanalyses given new genomic knowledge and evolving phenotypes. The “Genomic Answers for Kids” (GA4K) Program at Children’s Mercy Kansas City (CMKC) has so far recruited 6,422 probands with suspected rare genetic disorders and identified 1,761 diagnoses. Our repository includes data from various technologies including SR exomes and genomes, as well as the largest dataset of rare disease HiFi genomes (PacBio). Emedgene (Illumina) is utilized for AI-assisted variant annotation and prioritization. Data extracted in January 2023 included 691,562 “candidate” variants from 4,651 affected individuals. Restricting to private variants reduced this almost 4-fold to 174,503, demonstrating the power of having robust internal datasets. Given that gene discoveries still represent a major source of novel diagnoses, we queried new entries in OMIM (June 23-May 24) and found 193 new disease-gene associations for 177 genes, with 123 of these being novel to OMIM. Inheritance patterns were as follows: 36 autosomal dominant (AD), 77 autosomal recessive (AR), 3 AD/AR, and 7 X-linked. Analysis focused on AD genes with medium-to-strong clinical validity and expected loss-of-function (LOF) mechanism per manual review (n=20) resulted in 318 unique single nucleotide variants (SNVs) and indels, of which 47 were absent from gnomAD v3. An additional 31 variants were excluded based on low quality call, presence in gnomAD v4, and variant properties (deep intronic, synonymous). Of the 16 variants left to review, 3 were diagnostic: two had been suspected (*HNRNPC, PSMB9*) while one was an unexpected second diagnosis (*SRRM2*). To enable reanalyses and data mining across 15,000 rare disease cases with clinical or research sequencing done at CMKC, we recently exported SNVs and structural variants (SVs) to XetaBase (Zetta Genomics), creating an annotated variant warehouse of over 250 million variants with linked phenotype and pedigree information. We queried Xetabase for LOF variants in the same 20 AD genes among patients tested clinically at

CMKC and found an additional 4 diagnoses (*CAPRIN1*, *SRRM2* x3): one had been reported as a gene of uncertain significance (2019) while the other three as diagnostic based on strong evidence from the literature even prior to OMIM inclusion. We anticipate similar gains in diagnoses across other novel genes and further gains by inclusion of SVs. Our work demonstrates the clinical utility of dynamic genomic and phenotypic data tracking in a health system allowing for systematic reanalyses.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3071F: STRchive: a dynamic resource detailing population-level and locus-specific insights at tandem repeat disease loci

Authors:

L. Hiatt[1], B. Weisburd[2], E. Dolzhenko[3], G. VanNoy[2], N. Kurtas[2], H. Rehm[2], A. Quinlan[1], H. Dashnow[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]Broad Inst., Boston, MA, [3]Pacific BioSci.s of California, Menlo Park, CA

Abstract:

Approximately 3% of the human genome consists of repetitive elements called tandem repeats (TRs), which include short tandem repeats (STRs) of 1-6bp motifs and variable number tandem repeats (VNTRs) of 7+bp motifs. TR variants contribute to several dozen mono- and polygenic diseases but remain understudied and “enigmatic,” particularly relative to single nucleotide variants. It remains comparatively challenging to interpret the clinical significance of TR variants. Although existing resources provide portions of necessary data for interpretation at disease-associated loci, it is currently difficult or impossible to efficiently invoke the additional details critical to proper interpretation, such as motif pathogenicity, disease penetrance, and age of onset distributions. It is also often unclear how to apply population information to analyses.
We present STRchive (S-T-archive, http://strchive.org/), a dynamic resource consolidating information on TR disease loci in humans from research literature, up-to-date clinical resources, and large-scale genomic databases, with the goal of streamlining TR variant interpretation at disease-associated loci. We apply STRchive —including pathogenic thresholds, motif classification, and clinical phenotypes—to a gnomAD cohort of ~18.5k individuals genotyped at 60 disease-associated loci.
Through detailed literature curation, we demonstrate that the majority of TR diseases affect children despite being thought of as adult diseases. Additionally, we show that pathogenic genotypes can be found within gnomAD which do not necessarily overlap with known disease prevalence, and leverage STRchive to interpret locus-specific findings therein. We apply a

diagnostic blueprint empowered by STRchive to relevant clinical vignettes, highlighting possible pitfalls in TR variant interpretation. As a living resource, STRchive is maintained by experts, takes community contributions, and will evolve as understanding of TR diseases progresses.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3072F: Structural Variation Detection from Short-Read Whole Genome Sequencing Data: An Approach that Increases Diagnostic Yield for Rare Disease

Authors:

A. Kandhadai[1], A. B. Lay[1], S. S. Mokashi[1], M. Lally[1], K. M. Butler[1], T. S. Fee[1], B. A. Hilton[1], B. R. DuPont[1], R. J. Louie[2], J. R. Jones[3], M. J. Friez[1]; [1]Greenwood Genetic Ctr., Greenwood, SC, [2]Greenwood Genetic Ctr., Portland, OR, [3]Greenwood Genetic Ctr., Greenville, SC

Abstract:

Whole genome sequencing (WGS) is rapidly becoming the first-tier test of choice for rare disease diagnostics, with its ability to detect a wider spectrum of variants compared to exome sequencing and microarray testing. Most clinical WGS is performed on Illumina short-read sequencing platforms. Efficient curation and prioritization across all the variants called from WGS data remains a challenge. Single nucleotide variants (SNVs) are easier to call and curate but also more numerous, and hence the challenge for reviewers is mainly with respect to annotation and prioritization. On the other hand, copy number variants (CNVs) and structural variants (SVs) are less common, but challenging for accurate detection and curation. Here we document how structural variant analysis on short-read WGS data, using the algorithm Manta and further filtering heuristics, identified variants of interest in undiagnosed patients.
Clinical WGS testing at the author's institution is currently sequenced on Illumina short-read libraries. Tools for analysis include Illumina's DRAGEN Bio-IT Platform and Emedgene software, and Bionano's VIA software. This analysis suite is configured to prioritize SNVs with protein-coding consequences, as well as CNVs larger than 3 kb with protein-coding consequences in known disease genes. However, some causative variants for rare disease fall outside this space and go undetected. Our investigation of undiagnosed patients began with the set of structural variant calls (SVs) for each sample made by the variant caller Manta, which is included in the DRAGEN secondary analysis toolkit. Recurrent calls among patients, which may represent either common variation or technical noise, were filtered out. The remaining SV calls were passed through a simple in-house annotation pipeline. SVs with protein-coding consequences, as well as de novo SVs where parental data was available, were prioritized for review. We identified

new candidate variants in 7 out of 92 patient datasets examined (7.6%), of which 6 were deletions and the remaining one was an inversion. 4 of the 7 SVs were de novo. 5 of the 6 deletions were larger than the CNV threshold of 3 kb in our validated pipeline, but were not prioritized due to inadequate disease association. One deletion was classified as Likely Pathogenic and the other SVs are suspected to be clinically relevant despite their Variant of Uncertain Significance (VUS) classification.
Thus, analysis of structural variation in short-read WGS data has the potential to increase diagnostic yield in testing for rare disease, complementing analysis of sequence variants and copy number variants.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3073F: The fourth Korean patient to receive a gene therapy for *RPE65*-related LCA caused by a paternal UPiD

Authors:

H. Lee[1], D. Moon[1], C. Yoon[2], U. Park[2], K. Park[2], R. Khang[1], G. Seo[1], E. Lee[2]; [1]3billion, Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract:

Leber congenital amaurosis (LCA) is a rare genetic eye disease which presents from birth with severe vision problems, nystagmus, and an extinguished electroretinogram. With 25 genes identified, it is the most severe form of inherited retinal dystrophies. The *RPE65* gene, which encodes retinoid isomerohydrolase, an enzyme that plays a critical role in the regeneration of visual pigment in photoreceptor cells, is a causative gene for autosomal recessive LCA. *RPE65* has received increased attention with the advent of the first-in-class recombinant adeno-associated virus gene therapy (voretigene neparvovec [VN]), which works by delivering a normal copy of the *RPE65* gene.
Here, we report a 33-year-old Korean male patient diagnosed with *RPE65*-related LCA and received gene therapy as the fourth patient treated in Korea. At the age of 18 years when he was first referred, he was clinically diagnosed with LCA based on his typical fundus findings of salt-and-pepper retinal dystrophy with fine pigment clumping, diffuse extinguished signal on fundus autofluorescence, and attenuated amplitude of electroretinogram. The patient initially received a proband exome sequencing (ES) at the age of 33 years. A homozygous frameshift variant, 1-68438247-AT-A (GRCh37); NM_000329.3:c.1067del (p.Asn356MetfsTer17) in the *RPE65* gene was identified with the entire chromosome 1 being called as a copy number neutral region of homozygosity, leading us to suspect uniparental isodisomy (UPiD). ES on both

parents confirmed not only the father being a heterozygous carrier and the mother not carrying any pathogenic variant but also the UPiD being of paternal origin. Within 5 months after the molecular diagnosis, the patient was treated with subretinal VN therapy and is being followed for prognosis.
This case is particularly unique in the aspect that the *RPE65*-related LCA was caused by UPiD of chromosome 1, but also because, to our knowledge, the index patient was the first UPiD case to receive VN treatment. It is estimated that less than 5% of rare diseases have an approved treatment, and in many cases, identifying the molecular basis is the first step in finding an appropriate treatment. Therefore, it is critical for the ophthalmologist to identify patients suspected of having *RPE65*-related LCA and perform a genetic test as early as possible. Although single gene/panel tests may be an option, in this case, performing ES as a first-tier test was favorable because there was another genetic event, the UPiD, that needed to be detected in addition to the disease-causing variant.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3074F: The Impact of Trisomy 21 on Dental Development - A Focus on Tooth Agenesis

Authors:

E. Severin[1], G. Moldoveanu[2], A. Moldoveanu[1]; [1]Carol Davila Univ. of Med. and Pharmacy, Bucharest, Romania, [2]C.I. Parhon" Natl. Inst. of Endocrinology, Bucharest, Romania

Abstract:

The presence of an extra chromosome 21 or a part of its long arm, including the 21q22 band, can occur through various genetic mechanisms such as full trisomy 21, mosaic trisomy 21, or translocation trisomy 21, leading to Down syndrome (DS). The severity of clinical manifestations in DS varies among individuals, making it difficult to predict the extent of impact on a fetus or newborn as they grow into adulthood. Dental anomalies are a common health problem for individuals with DS. The risk of dental anomalies, such as tooth agenesis or misaligned teeth, is higher in individuals with DS due to the genetic and physical differences associated with the condition. The severity of clinical consequences and the impact on oral health-related quality of life increase with the number of missing teeth. Our interdisciplinary study, involving orthodontics and medical genetics, aims to:

Predict the severity of oro-dental phenotype in individuals with DS based on cytogenetic analysis.Evaluate a range of oro-dental characteristics in individuals with DS.Assess the orthodontic treatment needs and complexity.Correlate these characteristics with different

chromosomal abnormalities (full trisomy 21, Robertsonian translocation, or mosaicism).Evaluate the need for integrated complex treatment.

Materials and Methods: Ninety-seven individuals with DS, aged 5 to 34 years, underwent chromosomal analysis to confirm the clinical diagnosis of DS. All cases were diagnosed cytogenetically after birth, and karyotyping of parents was required for children with Robertsonian translocation trisomy 21. To characterize the oro-dental phenotypic spectrum, extra- and intra-oral examinations and radiological evaluations were performed.**Results:** Individuals with full trisomy 21, mosaic trisomy 21, and translocation trisomy 21 were identified based on cytogenetic analysis. Most oro-dental features, including oral, dental, periodontal, and occlusion anomalies, showed variability in both frequency and expression. The most common dental anomaly was hypodontia, followed by microdontia. Delayed eruption of dentition was observed in most cases. Dental aesthetics were significantly affected in individuals with DS due to tooth malposition and skeletal discrepancies.**Conclusion:** The study underscores the complexity and variability of oro-dental phenotypes in individuals with DS, highlighting the need for personalised and multidisciplinary approaches to their dental care. The variation in oro-dental phenotype is not specifically associated with the different chromosomal mechanisms of trisomy 21.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3076F: Unraveling undiagnosed rare disease families by HiFi long-read genome sequencing

Authors:

W. Steyaert[1], L. Sagath[2], G. Demidov[3], N. Zomer[2], M. Steehouwer[2], L. O'gorman[2], R. Derks[2], J. Weiss[2], A. den Ouden[2], S. van den Heuvel[2], H. Swinkels[2], K. Neveling[4], S. Ossowski[3], L. Vissers[2], C. Gilissen[5], A. Hoischen[6], Solve RD long-read-sequencing working group, Solve-RD DITF-ITHACA, Solve-RD DITF-Euro-NMD, Solve-RD DITF-RND, Solve-RD DITF-EpiCARE, SOLVE-RD DATF, Solve-RD consortium; [1]Ghent Univ., Ghent, Belgium, [2]Radboudumc, Nijmegen, Netherlands, [3]Univ.sklinikum Tübingen, Tübingen, Germany, [4]Radboud Univ. Med. Ctr., Nijmegen, Gelderland, Netherlands, [5]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [6]Radboud Univ. Med. Ctr., Nijmegen, Netherlands

Abstract:

Solve-RD is a pan-European rare disease (RD) research program aiming to identify disease-causing genetic variants in previously undiagnosed RD families. We utilized 10-fold coverage high-fidelity long-read sequencing (HiFi LRS) to detect causative structural variants (SVs), single

nucleotide variants (SNVs), insertion-deletions (InDels), and short tandem repeat (STR) expansions in extensively studied RD families without a clear molecular diagnosis. Our cohort includes 293 individuals from 114 genetically undiagnosed RD families selected by European Reference Network (ERN) experts. Of these, 21 families were affected by so-called 'unsolvable' syndromes for which genetic causes remain unknown, and 93 families had at least one individual affected by a rare neurological, neuromuscular, or epilepsy disorder without a genetic diagnosis despite extensive prior testing.
Clinical interpretation and orthogonal validation of variants in known disease genes yielded thirteen novel genetic diagnoses due to de novo and rare inherited SNVs, InDels, SVs, and STR expansions (affecting *TUBA1A*, *DMD*, *NHS*, *NOP56*, *DAB1*, *RFC1*, *SPAST*, *TTN*, *MYOT*, and *REEP1*). In an additional four families, we identified candidate disease-causing SVs affecting several genes, including an *MCF2/FGF13* fusion and a *PSMA3* deletion. However, no common genetic cause was identified in any of the 'unsolvable' syndromes. Taken together, we found (likely) disease-causing genetic variants in 13% of previously unsolved families and additional candidate disease-causing SVs in another 4.3% of these families.
Our findings underscore the potential of HiFi LRS to uncover genetic causes in cases where traditional short-read sequencing may fail, particularly for complex SVs and repeat expansions. Additionally, our work highlights the importance of integrating advanced genomic technologies with clinical expertise and data sharing across international networks to accelerate the diagnosis and understanding of rare diseases.
In conclusion, our results demonstrate the added value of HiFi long-read genome sequencing in undiagnosed rare diseases, paving the way for improved diagnostic yield and better patient outcomes. Future studies should focus on expanding these efforts to larger cohorts and diverse populations to fully realize the potential of this technology in the field of rare disease research.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3077F: Utility of liquid biopsy in cutaneous mosaicims and hemihypertophy

Authors:

M. Witzel[1], M. Fujera[1], A. Hallermayr[1], T. Kessler[1], C. Gebhard[1], A. Abicht[1], E. Holinski-Feder[1], T. Neuhann[2]; [1]MGZ Medizinisch Genetisches Zentrum, Munich, Germany, [2]MGZ Medizinisch Genetisches Zentrum München, München, Germany

Abstract:

Background: Diagnostics of skin mosaicism and hemihypertrophy remains a challenge in the pediatric and genetic practice. Tissue biopsies require invasive procedures with limited application in children, CNS lesion, deep vascular anomalies, or lesions with unclear proliferational potential (e.g. Proteus syndrom), especially in an outpatient setting. Liquid biospy (LB) analyses cell free DNA (cfDNA) thereby revealing both postzygotic mosaic variants and germline variants. In addition, cell type-specific lethality decreases cell viability thereby potentially increasing the abundance of pathogenic variants alleles in the fraction of cfDNA, again underling the utiltiy of LB.
Methods: Duplex sequencing technology for detection of low frequency variants in cfDNA (liquid biopsy (LB)) was applied in children with mosaic conditions. The multi-gene panel included AKT3, BRAF, CCND2, EPHB4, FGFR1, GNA11, GNAQ, GNAS, HRAS, KRAS, MAP2K1, MTOR, NRAS, PIK3CA, PIK3R2, RASA1, SMO and others (>20 genes).
Results: We identified relevant pathogenic variants with suballelic VAF in 7 patients: patient (P1) with the clinical diagnosis Klippel-Trenaunay syndrome (AKT1 gene, NM_001382430.1:c.49G>A, p.(Glu17Lys), VAF 0.6%, class 5), patient (P2) with cavernoma of basal ganglia and prominent vessels of the eye sclera (PIK3CA gene, NM_006218.4:c.277C>T, p.(Arg93Trp), VAF 0.14%, class 4), patient (P3) with hemihypertrophy, Naevus flammaeus and syndaktyly (PIK3CA gene, NM_006218.4:c.316G>T, p.(Gly106Cys), VAF 2%, class 4), patient (P4) with hemihypertrophy, hyperpigmentation (KRAS gene, NM_004985.5, c.35G>A p.(Gly12Asp), VAF 1%, class 5), patient (P5) with diffuse capillary malformation with overgrowth (PIK3CA gene, NM_006218.4:c.353G>A p.(Gly118Asp), VAF 5.1%, class 4/5), patient (P6) with premature puberty, asymmetrical breast development, subcutaneous lipoma, diffuse hyperpigmentation and café-au-lait spots (NM_006218.4:c.1093G>A p.(Glu365Lys), VAF 0.12%, class 4/5) and patient (P7) with encephalocraniocutaneous lipomatosis (PIK3CA gene, NM_006218.4:c.277C>T,p.(Arg93Trp), 0.37%VAF, class 4/5).
Additional NGS sequencing of blood and skin fibroblast had remained unconclusive in all patients.
Conclusion: LB increases diagnostic yield in patients with mosaic diseases. Therefore, LB has profound implications for targeted therapy, cancer surveillance and genetic counseling.

Session Title: Molecular and Cytogenetic Diagnostics Poster Friday Session

Board 3078F: X chromosome CNVs: Candidate gene search by reanalysis of 18,666 chromosomal microarrays at the Greenwood Genetic Center

Authors:

N. Sahajpal[1], B. Hilton[1], T. Fee[1], B. Dupont[2], M. Friez[1], R. Stevenson[1]; [1]Greenwood Genetic Ctr., Greenwood, SC, [2]Greenwood Gen Ctr, Greenwood, SC

Abstract:

Genomic data reanalysis is a common practice in clinical diagnostic laboratories, performed either on a case-by-case basis or periodically in a defined manner, based on institutional preferences/policies. This process includes re-annotation and reanalysis of the genomic data (mostly exome/genome sequencing), and has been efficient in identifying causative variants in a percentage of previously undiagnosed cases. However, this approach is less commonly used to identify new candidate genes or to establish new genotype-phenotype associations. Herein, we reanalyzed copy number changes on the X chromosome from 18,666 chromosomal microarrays performed on individuals affected with rare disorders at the Greenwood Genetic Center (GGC) during 2010-2022. After excluding X chromosome aneuploidies, 4% of cases (749/18,666) carried a CNV on the X chromosome, of which 516 were duplications and 233 were deletions. Of these 749 CNVs, 216 duplications and 82 deletions overlapped with at least one XLID gene (X chromosome re-annotated with 162 XLID genes known at the time of reanalysis). Sixteen duplications included only one XLID gene, 10 of which were found to have no reported cases of duplications in the literature. The phenotypic consequences of these genes when duplicated were found to be more severe (*ZNF711*), similar (*GK*) or less severe (*MSL3, ATP6AP2, SHROOM4, LAS1L, RIPPLY1, PAK3, SLC25A5,* and *IDS*) compared to loss-of-function variants in these genes. To identify new gene-disease associations, at least two cases with overlapping CNVs that did not include any disease associated gene were shortlisted for detailed genotype-phenotype assessment. This identified 21 genes/regions with duplications and 14 with deletions. Upon manual review, two new XLID genes [*SH3KBP1* and *IL1RAPL2*; both deletions], two strong candidate neurodevelopmental delay genes [*TENM1;* intragenic duplication and *IGSF1;* whole gene duplication] and 3 potential candidate XLID genes (*JPX;* deletion in females, *LDOC1;* deletion, and *TMEM185A;* deletion] were identified. This study describes the phenotypic consequences of duplication involving 10 XLID genes, and identifies 2 new XLID genes, and 5 candidate XLID genes. This study highlights the utility of data reanalysis to identify new gene-disease associations.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6002F: A high-throughput screen to enable identification of pleiotropic transcriptional programs in asthma

Authors:

I. Salamone, L. Zhang, X. Zhong, C. Ober, X. He, M. Nobrega; Univ. of Chicago, Chicago, IL

Abstract:

Asthma is a chronic, heterogeneous respiratory syndrome and is one of the most prevalent diseases of childhood worldwide. Genome-wide association studies (GWAS) have identified more than 170 loci that impact asthma pathogenesis. Because genetic variants associated with asthma risk are overwhelmingly within non-coding regions, it is difficult to pinpoint which gene(s) are regulatory targets of distant non-coding variants and in which tissue(s) regulatory dysfunction contributes to disease pathogenesis. The answer may not be as simple as identifying one variant acting on one gene in one cell type; recent studies suggest that regulatory elements (REs) conferring genetic risk act through different mechanisms simultaneously in different tissues. Using statistical fine-mapping analysis, we have identified 160 candidate REs at asthma-associated loci that may be responsible for changes in respiratory epithelial and adaptive immune cells, two of the preeminent tissues involved in asthma pathogenesis. To understand how these REs, some of them hundreds of kilobases away from the nearest gene, are involved in asthma pathogenesis, we developed a CRISPR-based screen to investigate how gene regulation modulates key lung cell functions. We have generated a list of >300 candidate target genes and will determine how they modulate asthma pathogenesis in bronchial epithelial and T cells by investigating the effect of their perturbation on phenotypes such as barrier maintenance and cytokine signaling, respectively. We have found that perturbation of canonical immune and inflammation-related genes in bronchial epithelial cells results in changes in epithelial barrier integrity and wound healing, suggesting that genes conferring genetic risk for asthma may act pleiotropically in multiple tissues. Overall, these results will enable us to not only identify novel causal genes and the tissues in which they are acting, but also tease apart how genetic variation, transcriptional regulation, and cellular context intersect to modulate genetic risk and asthma pathogenesis.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6003F: A humanized yeast model to test neuropathy-associated *AARS1* variants for dominant-negative effects and activation of the integrated stress response

Authors:

S. Laureano Figueroa, A. Antonellis; Univ. of Michigan, Ann Arbor, MI

Abstract:

Charcot-Marie-Tooth (CMT) disease is the most common inherited peripheral neuropathy, affecting approximately 1 in 2,500 individuals. Typically inherited as a dominant trait, CMT exhibits significant phenotypic and genetic heterogeneity, with over 130 implicated genes. Among these, several aminoacyl-tRNA synthetases (ARS) have been identified as relevant to CMT. ARS are enzymes responsible for charging tRNA with cognate amino acids, a process essential for protein synthesis. Currently, seven ARS have been reported in dominant CMT, including alanyl-tRNA synthetase (*AARS1*), which ligates alanine to tRNA in the cytoplasm. While dominant-negative effects and activation of the integrated stress response (ISR) have been implicated in ARS-associate neuropathy, a link between these mechanisms has not been identified. A common pathogenic *AARS1* variant (R329H) has been identified in multiple families with dominant axonal CMT. This variant has a dominant-negative effect in a yeast growth assays that allows expression of both wild-type and mutant human *AARS1*. Located in *AARS1* exon 8, the R329H variant arises from 5-methylcytosine (5MC) deamination of a CpG dinucleotide on the non-coding DNA strand. Exon 8 is highly mutable, necessitating close evaluation for mutations in patients with dominant axonal CMT. We aim to determine the role of R329 in disease by testing the functional consequences of all possible amino-acid changes at this codon. The pathogenicity of these variants will be assessed using the above yeast assay and we expect to identify several loss-of-function variants with variable dominant-negative effects. Importantly, these studies may reveal separation of function mutations (e.g., loss-of-function mutations that do not enact a dominant-negative effect) that will be valuable for further refining disease mechanisms. We will also modify the above yeast growth assay to test pathogenic *AARS1* variants for activation of the ISR. To this end, we will use CRISPR-Cas9 to insert the GFP coding sequence into the *PCL5* gene, a critical component of the yeast ISR. This will both sensitize the yeast cells to ISR activation and provide a reporter for ISR activation. We will then test pathogenic *AARS1* variants with dominant-negative effects (e.g., R329H) for activation of the ISR. If successful, this model will be valuable to evaluate the pathogenicity of newly identified *AARS1* variants in patients with CMT disease.In sum, these studies will improve our understanding of the allelic heterogeneity and molecular mechanisms of *AARS1*-associated peripheral neuropathy.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6006F: A transcriptome-wide association study of amyotrophic lateral sclerosis

Authors:

X. Zhou, H. Zeng, J. Warren, T. Soare, C. O'Dushlaine; insitro, South San Francisco, CA

Abstract:

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease that has a lifetime risk of 1 in 350 worldwide. Early ALS genetics research was driven by the discoveries in the familial form of the disease where disease is caused by single gene mutations. The remaining ALS liability is likely a combination of genetic and environmental factors similar to other complex traits. Recent genome-wide association studies (GWAS) have identified over 15 independent loci, but translation of GWAS findings into therapeutics has been limited. A transcriptome wide association study (TWAS) uses genetically predicted expression levels to probe for association of expression level with complex traits. It can identify drug targets and screenable phenotypes and represents a general strategy to integrate information of high-content molecular endophenotypes in a densely phenotyped cohort with large-scale GWASs. We implemented and applied a TWAS strategy on ALS that leverages public available molecular QTL data sets from central nervous systems related cell types, and also analyzed individual level transcriptomics and whole genome sequencing data from about 1300 post-mortem brain tissues of over 400 patient donors collected by the New York Genome Center (NYGC) ALS consortium. Our analyses uncovered known ALS genes (including *SCFD1*, *MOBP*) and suggested new candidates (including *RESP18*, *SLC9A8*) implicating protein processing in endoplasmic reticulum and cytoskeletal functions. Deeper review of selected loci revealed the presence of regulatory pleiotropy, tissue or condition specific regulatory effects. Comparative TWAS analysis using post-mortem tissues from affected and unaffected donors also revealed the limitation of post-mortem brain tissues to capture disease-specific cell types.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6007F: A unified framework for cell-type-specific eQTLs prioritization by integrating bulk and scRNA-seq data

Authors:

J. Xiao[1], X. Yu[1,2]; [1]Shenzhen Res. Inst. of Big Data, Shenzhen, China, [2]The Chinese Univ. of Hong Kong, Shenzhen (CUHK-Shenzhen), Shenzhen, China

Abstract:

Genome-wide association studies (GWASs) have identified numerous genetic variants associated with complex traits, yet the biological interpretation remains challenging, especially for variants in non-coding regions. Expression quantitative trait loci (eQTLs) studies have linked these variations to gene expression, aiding in identifying genes involved in disease mechanisms. Traditional eQTL analyses using bulk RNA sequencing (bulk RNA-seq) provide tissue-level insights but suffer from signal loss and distortion due to unaddressed cellular heterogeneity. Recently, single-cell RNA sequencing (scRNA-seq) has provided higher resolution enabling cell-type-specific eQTL (ct-eQTL) analyses. However, these studies are limited by their smaller sample sizes and technical constraints. In this paper, we present a novel statistical framework, IBSEP, which integrates bulk RNA-seq and scRNA-seq data for enhanced cteQTLs prioritization. Our method employs a hierarchical linear model to combine summary statistics from both data types, overcoming the limitations while leveraging the advantages associated with each technique. Through extensive simulations and real-data analyses, including peripheral blood mononuclear cells and brain cortex datasets, IBSEP demonstrated superior performance in identifying ct-eQTLs compared to existing methods. Our approach unveils new transcriptional regulatory mechanisms specific to cell types, offering deeper insights into the genetic basis of complex diseases at a cellular resolution.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6008F: Assessment of Longitudinal Trajectory Proteomic QTLs in the Multi-Ethnic Study for Atherosclerosis (MESA)

Authors:

M. Gillman[1], F. Aguet[2], K. Ardlie[2], J. Barber[3], T. Blackwell[4], Z-Z. Chen[3], C. Clish[2], D. Cruz[3], P. Durda[5], C. Johnson[6], A. Manichaikul[7], S. Onengut[7], P. Rao[3], U. Tahir[8], K. Taylor[9], R. Tracy[5], J. Wilson[3], A. Wood[10], R. Gerszten[3], S. Rich[7], J. Rotter[9], L. Raffield[1]; [1]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [2]Broad Inst., Cambridge, MA, [3]Beth Israel Deaconess Med. Ctr., Boston, MA, [4]Univ. of Michigan, Ann Arbor, MI, [5]Univ. of Vermont, Burlington, VT, [6]Univ. of Washington, Seattle, WA, [7]Univ. of Virginia, Charlottesville, VA, [8]Beth Israel Deaconess Med. Ctr., Cambridge, MA, [9]Lundquist Inst., Torrance, CA, [10]Baylor Coll. of Med., Houston, TX

Abstract:

Introduction:
Circulating plasma protein levels can provide health insights. Identifying genetic factors associated with these levels as protein quantitative trait loci (pQTLs) can provide a direct link to potential functional mechanisms or biological pathways related to heath. To date, almost all pQTL analyses have focused on a single timepoint, however, protein levels change over the lifetime, with interindividual variability in this rate. While prior work has identified pQTLs with different effect sizes in different ages, few studies utilize proteomics data collected from the same individual across multiple decades, enabling exploration of genetic associations with interindividual protein level changes.
Methods:
Using longitudinal proteomics data (Olink 3k) in the Multi-Ethnic Study for Atherosclerosis (MESA), we calculated a protein trajectory (slope) for each individual (n = 1,967) and protein (n = 2,941) across 3 visits spanning 14 to 18 years. We conducted a genome-wide analysis for each trajectory, adjusting for baseline age, sex, ancestry principal components (PCs), and baseline cross-sectional protein value (in order to identify pQTLs significantly associated with protein abundance beyond a single time point). SuSiE credible sets were identified using individual-level LD. A UK Biobank subset (n = 948) with longitudinal proteomics data (three visits across ~13 years, Olink 1.5k) was used for replication, adjusting for baseline age, sex, assessment center, genotype chip, ancestry PCs, and baseline cross-sectional protein value.
Results:
We identified 393 trajectory cis-pQTLs ($p < 5e-8$) and 26 trajectory trans-pQTLs ($p < 1.7e-11$) in MESA (distance-based clumping, 1Mb). Of the proteins also assayed in UK Biobank, the majority of MESA trajectory pQTLs overlap with trajectory pQTLs in UK Biobank (72.3% of cis-pQTLs and 83.3% of trans-pQTLs). Replicated pQTLs include a cis-pQTL for the protein IL18R1 on chromosome 2 (2q12.1) and a trans-pQTL for IL3RA at the pleiotropic *ABO* locus on chromosome 9.
Conclusion:
In conclusion, when adjusting for the cross-sectional measure, we identify several trajectory pQTLs in MESA that replicate in UK Biobank. Most of the identified credible sets overlap with cross-sectional credible sets identified in UK Biobank (in > 10x the sample size), suggesting that many of the same pQTLs that influence the cross-sectional protein level in this large dataset also influence (independent of the cross-sectional measure) the trajectory of a protein in the MESA dataset. Our findings build on understanding of interindividual proteome changes over time and suggest potential genetic mechanisms influencing that change.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6009F: Benchmarking and characterizing *FMR1* mosaicism in the human brain

Authors:

A. Dischler[1], C. Dias[2], C. Lau[1]; [1]Univ. of Colorado Anschutz, Aurora, CO, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Fragile X-associated tremor/ataxia syndrome (FXTAS) is an adult-onset neurodegenerative disorder with clinical features including ataxic gait, parkinsonism, executive dysfunction and cognitive decline, neuropathy, and mood dysregulation. FXTAS affects carriers of the Fragile X *premutation*, which is a 55-200 CGG trinucleotide repeat expansion located in the 5' UTR of the Fragile X Messenger Ribonucleoprotein 1 gene (*FMR1*). Individuals with the *FMR1* full mutation (≥200 CGGs) develop Fragile X Syndrome (FXS), a neurodevelopmental disorder. Interestingly, in general, individuals with FXS do not go on to develop FXTAS and individuals with FXTAS did not have FXS, implicating distinct neurological mechanisms in both disorders. The repetitive nature of the CGG sequence leads to unstable expansion/contraction of the repeat region across the germline and in somatic tissue. Prior studies of the premutation suggest that somatic instability is incredibly common, but these studies tend to focus on blood. Surprisingly, little is known about patterns of CGG repeat instability in the human brain or if size mosaicism of the repeat impacts clinical severity, as has been implicated in other expansion disorders like Huntington's. Understanding this is crucial, given that repeats of different lengths have distinct clinical phenotypes. We hypothesize that somatic mosaicism of the CGG repeat in post-mortem brain corresponds with areas of the brain and cell types known to be impacted in FXTAS. We performed triplet repeat primed PCR/capillary electrophoresis (TP-PCR/CE) to amplify the CGG repeat of *FMR1* across several brain regions from individuals with FMR1 expansions and controls. We benchmarked injection parameters, experimental details, and data analysis methods to better interpret the TP-PCR/CE findings. We identified variable patterns of *FMR1* CGG somatic instability across the human brain and demonstrate the importance of establishing robust experimental and analytic approaches when using TP-PCR/CE. Collectively, this work will provide insight into the mechanisms, clinical impact, and therapeutic potential of *FMR1* CGG repeat mosaicism in the human brain

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6010F: Biomarker discovery for Hypophosphatasia

Authors:

C-F. Lin[1], J. Kubica[2,3], B. Busby[1]; [1]DNAnexus, Mountain View, CA, [2]Univ. Grenoble Alpes, CNRS, UMR 5525, TIMC / MAGe, Grenoble, France, [3]Lab. of Functional and Structural Genomics, Ctr. of New Technologies, Univ. of Warsaw, Warsaw, Poland

Abstract:

Hypophosphatasia (HPP) is a rare disease that results from defective bone and teeth mineralization. The severity of the disease has a broad spectrum, ranging from lethal in infants to mild in adults. Currently, diagnosis of HPP consists of clinical feature analysis, lab work, radiology, and genetic analysis of the alkaline phosphatase gene (ALPL). However, it is often delayed because some symptoms, e.g. low activity of serum and bone alkaline phosphatase overlap with other diseases. The presence of genetic variants in ALPL, however, is not a requirement for a HPP diagnosis. As of today, there are ~1200 variants documented for ALPL in ClinVar. The vast majority of the variants were evaluated based on their impact on the protein-encoding function. The inheritance of reported pathogenic variants encompasses both autosomal recessive and autosomal dominant. Some patients appear to exhibit dominant negative effects, suggesting the involvement of Nonsense Mediated Decay (NMD) escape. More than half of the gene body is occupied by an intron (> 40k bp), which shows signals of epigenetic modifications and protein binding activities. The allele frequencies of this intron ascertained from the WGS data of large scale population genetic initiatives (e.g. gnomAD, TOPMED, UK Biobank), alongside multiomic data (epigenetic, transcriptomic, proteomic), provide an *in silico* framework to assess impact of non-coding variants. We set out to discover additional variants that may affect the presentation of HPP, in particular, focusing on the non-coding variants of ALPL in the context of RNA secondary structure as well as cis- and trans-regulatory activities. In order to identify potential novel HPP-causing genes, we expand the analysis to all genes thought to be involved in the bone mineralization process. This framework could be applied to other rare diseases.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6011F: Borzoi-guided fine mapping improves variant and gene prioritization in GWAS

Authors:

D. Srivastava, H. Yuan, A. Korsakova, L. Ruiz, Q. Wang, D. Kelley; Calico Life Sci., South San Francisco, CA

Abstract:

Genome-wide association studies (GWAS) have identified thousands of trait-associated loci, but identifying causal variants and characterizing their functional consequences remains challenging. Fine mapping is a commonly used approach that identifies causal variants at trait-associated loci. Functionally-informed fine mapping (FIFM) is an augmentation to fine mapping that estimates the prior likelihood of a variant being causal based on functional annotations, such as overlap with a baseline 187 annotations including both coding and noncoding cis-regulatory elements. Despite improving the accuracy of fine-mapping methods, FIFM with the baseline annotation set is unable to resolve a large fraction of GWAS loci. A limitation of using overlap-based annotations to prioritize variants is that only a fraction of variants in regulatory regions have functional consequences and overlap alone with a regulatory element does not necessarily imply a functional consequence.
Sequence-based deep learning models accurately predict functional epigenomic profiles across cell-types. Borzoi is a state-of-the-art long-context transformer model which predicts RNA-seq in addition to epigenomic tracks such as ATAC-seq and ChIP-seq. Borzoi estimates variant effects by replacing the reference allele with an alternative allele and measuring the difference in predicted epigenomic outcomes. Here, we demonstrate that FIFM with Borzoi variant effect predictions (VEPs) improves GWAS fine mapping. We score ~18 million common variants and perform Borzoi-guided fine mapping for 15 UK Biobank traits. We combine Borzoi VEPs with additional functional annotations via PolyFun to derive prior causal probabilities for each variant. Incorporating Borzoi annotations results in an increased number of fine-mapped variants compared to both standard fine mapping (14 out of 15 traits) and FIFM with baseline annotations (11 out of 15 traits). Compared to FIFM with baseline annotations, we fine map an additional 116 (6.8%) variants at PIP > 0.95 and 310 (11.5%) variants at PIP > 0.8 across the 15 traits. Borzoi-guided fine mapping results in more resolved credible sets. Borzoi-guided fine mapping implicates sets of novel genes that are trait relevant. We find that genes nominated by Borzoi-derived credible sets are more likely to be implicated as causal genes by PoPs, an orthogonal similarity-based gene prioritization approach. Finally, we transfer Borzoi to disease-relevant single-cell ATAC-seq atlases and leverage cell-type specific VEPs to improve fine mapping for disease GWAS and generate compelling hypotheses for the functional consequences of the fine-mapped variants.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6013F: Cell type-specific effects of common germline variation on gene expression and pancreatic cancer risk

Authors:

G. Prasad[1,2], T. Truong Vo[1], J. Zhong[2], J. W. Hoskins[2], L. T. Amundadottir[2], H. Arda[1]; [1]Lab. of Receptor Biology and Gene Expression, Ctr. for Cancer Res., Natl. Cancer Inst., NIH, Bethesda, MD, [2]Lab. of Translational Genomics, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Rockville, MD

Abstract:

Pancreatic ductal adenocarcinoma (PDAC) is an aggressive, often lethal cancer originating from the exocrine compartment of the pancreas that is composed of acinar and duct cells. Identifying genes affecting PDAC risk in these cells is crucial for early detection, prevention, and treatment. While high-risk inherited PDAC variants are typically found in protein-coding regions, common variants identified through Genome Wide Association Studies (GWAS) often reside in non-coding regions of the genome, making it difficult to determine their target genes. Previous expression QTL studies (eQTL) have used bulk tissue samples that may obscure cell-specific genetic effects. To better understand these effects in distinct exocrine pancreas cell populations, we employed fluorescence-activated cell sorting to isolate purified populations of pancreatic acinar (n=96) and duct (n=94) cell populations from 107 healthy donors. Subsequently, cell type-resolved expression quantitative trait loci (eQTL) and isoform quantitative trait loci (isoQTL) analyses were conducted to identify the target genes at common germline genetic signals that may influence gene function by modulating either gene expression or relative isoform levels, respectively. Using RNA-Seq data and imputed genotypes, we tested associations between 6 million imputed SNP genotypes and the expression of 19,185 genes in acinar and 22,115 genes in duct cell populations. We identified 66,471 *cis*-eQTLs (2,288 eGenes) in acinar cells and 114,723 *cis*-eQTLs (4,010 eGenes) in duct cells (FDR<0.1), with about 20% of *cis*-eQTLs appearing specific to each cell type. Compared to our similarly sized (n=95) previously published study on histologically normal bulk pancreas samples, this study identified 5 to 10 times more eGenes, underscoring a key benefit of cell-resolved datasets. Nearly 73% of acinar eQTLs and 72% of duct eQTLs replicated in previously published largest normal pancreas eQTL dataset (GTEx v8). Therefore, 27% of acinar eQTLs and 28% of duct eQTLs in this study represent newly identified exocrine pancreas eQTLs. Additionally, isoQTL analysis identified 7,396 isoQTLs (424 isoforms) in acinar cells and 13,261 isoQTLs (715 isoforms) in duct cells. Chromatin accessibility QTL (caQTL) and DNA methylation QTL (meQTL) analyses on paired ATAC-seq and meDNA-seq data, respectively, is ongoing, which may aid in fine-mapping functional variants and implicate mechanisms by which they influence their target genes. Our findings reveal novel genes and regulatory mechanisms specific to exocrine pancreas cell types, offering valuable insights into PDAC risk.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6014F: Cell-type specific and temporally dynamic genetic regulatory effects in response to environmental perturbations

Authors:

W. Lin[1], M. Li[2], A. Battle[2], Y. Gilad[1]; [1]Univ. of Chicago, Chicago, IL, [2]Johns Hopkins Univ., Baltimore, MD

Abstract:

Understanding the molecular effects of genetic variation across different cellular contexts is crucial for elucidating the mechanisms underlying disease-associated loci. Cellular contexts encompass not just spatial components (cell types) but also temporal components (developmental stages), and external components (environmental exposures). Despite advancements in genomics, the intricate interplay between genetic variants and their regulatory impact on gene expression within a multi-dimensional cellular context remains underexplored. To explore genetic effects on gene regulation across a broader array of cellular contexts, we leveraged heterogeneous differentiating cultures (HDCs). HDCs are 3D-aggregates derived from human iPSCs and comprise asynchronously developing cells from all three germ layers. We generated control HDCs and HDCs exposed to nicotine, caffeine, and ethanol separately from 54 individuals. Using scRNA-seq data from over 1.8 million cells, we identified 58 distinct cell types and characterized the effects of each exposure on gene expression, gene regulatory networks, cell type abundance, cell state transition, and cell-cell communication within each cell type. Across all three perturbations, we discovered 3566 genes with at least one cis-eQTL (i.e., cis-eGenes) and 419 response eGenes that are largely cell-type specific. To account for the heterogeneity of response, we estimated a continuous perturbation score for each cell and identified additional reQTLs that only emerge when the response is strong. We also discovered six cellular developmental trajectories from all germ layers. Across the trajectories, we characterized continuous and transient genetic effects in response to each exposure and identified hundreds of temporally dynamic reQTLs. Our results show dynamic gene-environment (GxE) interactions that emerge temporally and transiently in early developing cell types and along specific developmental trajectories.
We anticipate that these context-specific gene regulatory effects will help to explain disease-associated loci that have thus far failed to display regulatory potential in steady-state tissues and advance our understanding of GxE interactions in modulating disease risk that cannot be characterized in *vivo* in humans.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6015F: Characterization of inter-individual and cell-type-specific responses to treatment with *ALK* inhibitors

Authors:

E. Woo, K. Rhodes, K. Barr, Y. Gilad; Univ. of Chicago, Chicago, IL

Abstract:

Genetic background contributes to individual susceptibility for drug-induced severe adverse events (SAEs). Identification of genetic factors associated with SAEs may be useful for predicting which patients are at increased risk. We hypothesize that inter-individual variability in response to treatment is mediated in part by genetic variants that have drug-dependent effects on gene expression and interact either directly or indirectly with treatment (i.e. response eQTLs), often in the context of a specific tissue or cell type. Mapping response eQTLs to anticancer drug treatments can help provide intuition on the mechanisms and genes underlying SAEs. We investigate this hypothesis for anaplastic lymphoma kinase (*ALK*) inhibitors used to treat *ALK*-positive non-small cell lung cancer (NSCLC). *ALK* inhibitors have increased efficacy compared to standard chemotherapy but are also associated with a variety of SAEs affecting the eyes, heart, liver, brain, and other tissues. Given the impact of SAEs on multiple tissues, we leveraged iPSC-derived heterogeneous differentiating cultures (HDCs) from which we can identify and distinguish over 70 cell types, including those representative of tissues affected by *ALK* inhibitor toxicity. We exposed HDCs of 49 individuals to treatment with 5 *ALK* inhibitors (crizotinib, alectinib, ceritinib, brigatinib, lorlatinib). We performed single-cell RNA-sequencing to measure transcriptional responses from over 1.9 million cells, with the goal of characterizing inter-individual and cell-type-specific responses to treatment. Preliminary analyses identified thousands of genes that are differentially expressed in treated compared to untreated HDCs, including some with cell-type specific patterns. Ongoing analyses will identify drug response eQTLs that may mediate tissue-specific responses to *ALK* inhibitors and can be further evaluated for their capacity to develop predictive classifiers for SAEs in NSCLC patients. This work contributes to a better understanding of individual genetic risk for SAEs, which could eventually help clinicians make more informed decisions when choosing an *ALK* inhibitor for a patient.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6016F: Comparative, cross-population analysis of the genetic regulation of plasma proteins in the Million Veteran Program ⭐

Authors:

D. Panyard[1], C. Yandava[2], P. Schreiner[3], D. Dochtermann[2], B. Gorman[2], M. Snyder[4], S. Pyarajan[5], P. Tsao[6]; [1]Stanford Univ., Stanford, CA, [2]VA Boston Hlth.care System, Boston, MA, [3]Booz Allen Hamilton, Lexington, MA, [4]Stanford Dept. of Genetics, Palo Alto, CA, [5]VA, Jamaica Plain, MA, [6]VAPAHCS/Stanford Univ., Palo Alto, CA

Abstract:

Background: The use of proteomics data in large-scale biobanks and population studies has increased dramatically in recent years. However, the field faces several challenges, including the lack of population diversity among proteomics cohorts and poor correlation between the leading proteomics platforms for many proteins. To address these challenges, we analyzed genetic associations with proteomics data (pQTLs) from the Million Veteran Program (MVP) in a diverse population across two proteomics platforms. **Methods**: We analyzed data on 862 unrelated veterans from the MVP cohort, comprising 616 individuals from the European (EUR) Harmonized Ancestry and Race/Ethnicity (HARE) group and 246 from the African (AFR) HARE group. For each individual, relative protein levels from both the Olink Explore (1,472 assays) and SomaLogic SomaScan (7,289 assays) platforms were measured. We conducted a population-stratified genome-wide association study (GWAS) for each protein, controlling for age, sex, duration of sample storage, and the top genetic principal components. Only common (MAF > 0.05) variants were included, and a Bonferroni-corrected significance threshold of 5.7×10^{-12} was used. **Results**: Among EUR individuals, 221 (15.0%) proteins from the Olink platform had a statistically significant pQTL compared to 750 (10.3%) proteins from the SomaLogic platform. Among AFR individuals, 79 (5.4%) of Olink proteins had a pQTL compared to 977 (13.4%) of SomaLogic proteins. Among Olink proteins with a pQTL, few (EUR: 4.1%; AFR: 2.5%) had multiple independent loci, while a greater proportion of SomaLogic proteins (EUR: 8.3%; AFR: 20.8%) had multiple loci, particularly for the AFR cohort. Trans-pQTLs were far more common among the SomaLogic data, with 510 (68.0%) and 883 (90.3%) of proteins in EUR and AFR, respectively, having a trans-pQTL compared to only 62 (28.1%; EUR) and 36 (45.6%; AFR) of Olink proteins with a trans-pQTL. Cis-pQTLs from the SomaLogic data were much more likely to be replicated in the Olink platform (67.1% replication) than SomaLogic trans-pQTLs (3.5% replication). Furthermore, SomaLogic proteins with only trans-pQTLs were more likely to be proteins poorly correlated with the Olink measurements. **Conclusion**: Numerous pQTLs were identified using both Olink and SomaLogic platforms for both EUR and AFR individuals. Olink proteins were more likely to have cis-pQTLs while SomaLogic proteins were more likely to have trans-pQTLs. The trans-pQTLs from the SomaLogic proteins were less likely to be identified in the Olink data for those same proteins, and SomaLogic trans-pQTLs were more likely to be for proteins with poor interplatform correlation.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6017F: Context-dependent proportional amplification of genetic and environmental effects on gene expression

Authors:

C. Fadil[1,2], B. Strober[3], A. Price[3]; [1]PhD Program in Biological and BioMed. Sci., Harvard Univ., Boston, MA, [2]Div. of Med. Sci., Harvard Univ., Boston, MA, [3]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Genetic effects on gene expression are widely believed to be context-dependent (Kim-Hellmuth et al. 2020 Science), likely limiting the proportion of disease heritability mediated by gene expression in assayed contexts (Lappalainen et al. 2024 Cell). Context may also impact environmental variance, e.g. via a “proportional amplification” model in which context impacts genetic and environmental variance proportionately (Durvasula et al. medRxiv). Here, we explore the impact of age and other contexts on genetic and environmental variance of gene expression.

We analyzed GTEx whole blood RNA-seq data (N=670), stratifying samples into young (median age<45) and old age groups. We considered a proportional amplification model in which genetic effects and environmental variance vary proportionately with age, as well as a standard model in which genetic effects and environmental variance are independent of age. When compared to a null model with no genetic effects, the proportional amplification model (11,878 eGenes) and the standard model (11,877 eGenes) identified similar numbers of eGenes; the proportional amplification model attained a better model fit for 50% of eGenes identified by both models. Interestingly, 62 eGenes were identified only by the proportional amplification model (of which only 1 was identified using a standard GxAge interaction test that assumes that environmental variance is independent of age).

We investigated a larger set of contexts, including 60 PEER factors, imputed cell type composition for 9 cell types, and technical variables such as post-mortem interval and circumstance of death that are known to impact gene expression. We determined that the full set of contexts analyzed explained 32% of total variance in gene expression, averaged across genes. The largest contributions were explained by PEER factors (30%), neutrophil abundance (1.2%), and circumstances of death (1.2%), with age accounting for 0.3%. We will assess the extent of proportional amplification with respect to these contexts, and quantify the increase in power to map eQTL main effects attained using sample-dependent weights that account for differences in predicted variance.

In conclusion, we show that multiple contexts can impact genetic and environmental variance in gene expression, and that modeling their effects can enhance eQTL discovery.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6018F: Coronary artery disease SNPs regulate transcript diversity through splicing quantitative trait loci in endothelial cells

Authors:

A. Golebiewski, L. Stolze, C. Romanoski; Univ. of Arizona, Tucson, AZ

Abstract:

Endothelial cells (ECs) line blood vessels and maintain blood-tissue barriers and vascular homeostasis. Chronic inflammatory signaling can induce endothelial dysfunction, a hallmark of many complex diseases including coronary artery disease (CAD). We and others have characterized the transcriptomic and epigenomic EC inflammatory response, but few studies have described the effects on splicing. Thus, we sought to quantify the effect of inflammatory signaling on RNA splicing in ECs, and to identify genetic signals associated with alternative splicing. We identified 1224 alternatively spliced transcripts (deltaPSI > 0.05) between Il1B and control treatments. 29% of differential splicing events are also associated with at least one splicing quantitative trait locus (sQTL, FDR < 0.05). We cross-referenced sQTLs with summary statistics from the most recent GWAS for CAD and found that sQTLs are significantly enriched for CAD loci above random expected rates of overlap. sQTLs are significantly enriched for EC molecular QTLs for markers of active transcription including histone 3 lysine 27 acetylation (H3K27ac), chromatin accessibility (ATAC), and binding of the transcription factors *Ets-related gene (ERG)* and *nuclear factor kappa B (NFkB)*. This indicates that alternative splicing depends on multi-faceted epigenetic regulatory mechanisms that affect both the epigenome and the transcriptome. For example, *rs12548549* is an EC molQTL for H3K27ac (p=0.0003, b=0.41), and a sQTL for *PTK2* (p<0.05, b=0.05) and is marginally associated with CAD (p=9.378e-06, b=0.017). *PTK2* encodes the focal adhesion kinase protein (FAK) which regulates key EC functions including nitric oxide synthesis and angiogenesis. The alternative splicing event regulated by *rs12548549* results in changes in the 5' UTR leading to changes in translation initiation. Another example is *rs11259179*, which is an sQTL for *FAM107B* (p=3e-04, b=0.12) and molQTL for H3K27ac (p=3.6e-06, b=0.42). This SNP is in an intronic site and associates with alternative promoter and first exon usage between Il1B and control treatments (deltaPSI=0.09, FDR=0.001). We found that alternative promoter and first exon usage represent 30% of the differential splicing events between Il1B and control treatments. Furthermore, sQTL SNPs

identified in this analysis are enriched in promoter regions for the target gene suggesting they affect assembly of transcriptional machinery. Overall the sQTLs identified in this analysis provide potential functional insight as to how CAD-associated SNPs impact disease progression or severity through modulation of both the epigenetic landscape and transcript diversity.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6019F: De novo assembled genomes from Middle Eastern trios unveil high-resolution diversity and effect of inbreeding

Authors:

Y. Mokrab[1,2,3], M. Ghorbani[1], S. Moosa[1], R. Farhad[1], Z. Siddig[1], H. Naeem[1], W. Harvey[4], F. Mastrorosa[4], K. Munson[4], E. Aliyev[1], I. Diboun[1], R. Abouelhassan[1], M. Tauro[1], S. Hassan[1], R. Mathew[1], M. Hashmi[1], F. Vempalli[1], A. Khouly[1], Qatar Genome Project Research Consortium, I. Azwani[1], S. Tomei[5], K. Fakhro[5,2], A. Satti[1], R. Benini[1], A. Rhie[6], E. Eichler[7,8]; [1]Sidra Med., Doha, Qatar, [2]Weill Cornell Med., Doha, Qatar, [3]Qatar Univ., Doha, Qatar, [4]Univ. of Washington Sch. of Med., Seattle, WA, [5]Sidra Med., Doha, Qatar, Qatar, [6]NIH, Bethesda, MD, [7]Univ of Washington, Seattle, WA, [8]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA

Abstract:

Substantial advances in long-read sequencing have enabled routine de novo assembly of human genomes, opening a new era of “precision genomics” set to transform biomedicine. Hundreds of near-complete genomes have so far been published, yet there is still much to be done to represent understudied populations and show impact on disease gene discovery. Here, we de novo assemble highly accurate (Max QV=59.3) near-complete (> 99% including 14 telomere-to-telomere chromosomes; Max contig N50=98.2 Mb) and well-phased genomes from peripheral blood DNA of six family trios (n=18) having children with neurodevelopmental conditions, from diverse Middle Eastern (ME) ancestries encompassing Sudan, Jordan, Syria, Qatar, and Afghanistan. The genomes showed varying levels of Arab, Eurasian and African admixture, revealing 75 novel HLA/KIR alleles and 46.1 Mb of novel sequence. Enriched for runs of homozygosity (ROH), they showed strong signature of background and recent inbreeding, with ROH covering as much as third of Chr 6 and 12 in the healthy Jordanian father. In the probands, ROH harbored deletions impacting several genes providing strong candidates for resolving previously undiagnosed conditions. Patterns of SNVs and SVs emphasized ancestry bias in both GRCh38 and CHM13 and showed up to 10% fewer calls made for ME samples against our assemblies, stressing the need for population-specific references to enhance the discovery of both founder and disease mutations.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6020F: Deep Cellular Phenotyping of genetic variants in Hunter Syndrome

Authors:

A. Viswanathan[1], S. Elia[2], S. Q. Le[3], B. Doray[3], S. Hurt[3], W. Buchser[4], P. Dickson[5]; [1]Washington Univ. in St. Louis, St. Louis, MO, [2]Washington Univ. in St. Louis, Saint Louis, MO, [3]Washington Univ. Sch. of Med., 1 Brookings Dr, MO, [4]Washington Univ. in St Louis, St Louis, MO, [5]Washington Univ. Sch. of Med., Saint Louis, MO

Abstract:

Mucopolysaccharidosis II (MPS II, Hunter syndrome) is a lysosomal disorder caused by hemizygous or biallelic loss of function variants in the IDS gene encoding iduronate 2-sulfatase. Numerous variants of unknown significance (VUS) in the IDS gene have been identified that cannot be easily classified as pathogenic or benign. To aid in VUS classification, we used a functional genomics platform that combines high-content imaging with machine learning on microrafts to define a cellular phenotype of human A549 cells that were wild-type or engineered to express known pathogenic (G224R and 228fsX), or a known benign (V223I) IDS variant. IDS activity assay of isogenic cultures of these lines showed little to no activity for the pathogenic variants, whereas 7 units/mg protein activity was observed for V223I, and 28 units/mg protein activity for WT, as expected. Cells were imaged using a fluorescent confocal microscope and intracellular features were extracted from nuclei and lysosomes on a per-cell basis. Combining lysosomal measurements for intensity, puncta morphology and texture generated a 'pathogenic score', which separated isogenic cultures of the benign from the two pathogenic variants, with 70% of G224R variant cells, 68% of 228fsX variant cells, and 30% of V223I (benign) variant cells scoring as pathogenic-like. To test the robustness of the IDS cellular phenotype, we performed therapeutic "rescue" with recombinant human IDS (rhIDS) applied to cells. G2424R and 228fsX variant cells showed phenotypic rescue, with 50 or 59% of cells scoring as pathogenic-like 144h after rhIDS treatment, respectively. RNAseq experiments will evaluate transcriptome changes in IDS-treated and untreated variant cell lines. IDS uptake experiments will further elucidate the extent of biochemical recovery in treated cells. Deep cellular phenotyping offers a novel means of evaluating the function of patient-specific IDS variants to improve the diagnostic yield of molecular testing for Hunter syndrome.

Board 6021F: Defining tissue- and isoform-specific function of *OCRL* using patient-derived tissues

Authors:

G. Iannello[1], E. Semenova[1], A. Guo[1], C. Sena[1], B. Corneo[1], C. Leduc[2], G. Stratigopoulos[3], V. Thaker[1]; [1]Columbia Univ. Irving Med. Ctr., New York, NY, [2]Columbia Univ, New York, NY, [3]Columbia Univ., New York, NY

Abstract:

Background: The *OCRL* gene encodes an inositol polyphosphate-5-phosphatase, which plays a critical role in membrane trafficking, endocytosis, and primary cilia assembly. Mutations in *OCRL* are associated with Lowe syndrome and Dent disease, involving kidney, brain, and eye. By interacting with clathrin and Rab GTPases, OCRL can impact cell differentiation and function. *OCRL* exists in two isoforms - brain-specific isoform A (901 aa) and a shorter isoform (893 aa) expressed in the remaining tissues. Our study of tissue and isoform-specific human induced pluripotent stem cells (iPSCs) aims to define the function of isoform A. **Results:** We identified a novel hemizygous X-linked frameshift *OCRL* variant (c.2136_2139del+1insCT p. L713Rfs*28) in a 5-year-old boy with extreme obesity (BMI 27 kg/m2, Z= +5), short stature (Z=-2.2), neurobehavioral dysfunction, central hypoventilation, and ganglioneuroblastoma. Tissue expression profiling of patient-derived fibroblasts, iPSC, and iPSC-derived neurons showed lack of expression of isoform A compared to control. The expression of isoform A was restored in isogenic CRISPR-corrected iPSCs. We found that differentiated patient iPSC had impaired neuronal specification, with lower expression of NeuroD1, NRP1, MAP-2, PSD95, b3 Tubulin, GAP43 (P_{adjust} < 0.05 for all), as well as impaired axonogenesis. Higher expression of the neuronal stem cell markers OTX-2, SOX-2, PAX6, CDH2 was observed in the proband (P_{adjust} < 0.05 for all). We found decreased expression of oligodendrocyte, cells essential for myelination and neuronal function, and astrocytes markers, cells needed for synaptic formation and neuronal support. Higher expression of microglia markers was also found, yet with lower HLA-DR expression compared to the isogenic line, suggesting a diminished immune response capability in the patient microglia. These markers align with the downregulated molecular pathways of loss of axonal projection and guidance in the patient line. Pathways including components of the SNARE complex (synaptobrevin, syntaxin and SNAP-25) are differentially expressed in patient and isogenic lines, suggesting a brain-specific role of OCRL in clathrin-mediated endocytosis and vesicle-mediated transport. **Conclusions**: Our findings highlight a novel role of the OCRL isoform A in neuronal differentiation and specification. The tissue- and isoform- specific mutation impairs the development of neurons, and supportive structures. It also alters the cellular environment by impairing vesicle processing, critical for neuronal function.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6022F: Diagnosis of *CHKB* associated muscular dystrophy, congenital, megaconial type after variant phasing and RT-PCR

Authors:

J. Farris[1], M. Wilke[2], K. Muthusamy[3], C. Lamb[3], M. O'Connell[1], C. Schmitz[1], C. Dergam-Larson[1], L. Lambert[1], E. Klee[1]; [1]Mayo Clinic, Rochester, MN, [2]Washington Univ., St Louis, MO, [3]Mayo Clinic, Jacksonville, FL

Abstract:

Introduction: *CHKB encodes* a choline kinase that plays a crucial role in the synthesis of the membrane phospholipid phosphatidylcholine. Biallelic pathogenic variants in CHKB are associated with the phenotype of muscular dystrophy, congenital, megaconial type (CHKB-MD). CHKB-MD is associated with histologic findings in the muscle biopsy known as megaconial congenital muscular dystrophy (MCMD). We describe here a proband presenting MCMD due to two *CHKB* variants. The patient was an 18-year-old female with progressive muscular dystrophy, developmental delay, autism, and hidradenitis suppurativa. Muscle biopsy was highly suggestive of MCMD. Clinical neuromuscular disorder panel sequencing revealed two variants in *CHKB*:c.446del (Pro149Glnfs*5), which was classified as pathogenic, and a maternally inherited c.940C>T (Arg314Cys) which was classified as a variant of uncertain significance (VUS). The pathogenic variant was not identified in either tested parent, thus the variants could not be phased. The c.940C>T VUS had a Splice-AI prediction of 0.44 (range 0-1) exon 9 splice acceptor loss that contributed to the clinical suspicion of this variant.

Methods:

Variant phasing: cDNA was synthesized from patient's isolated RNA after which the region of interest was amplified by PCR. The amplicon containing both variants was visualized by gel electrophoresis, purified, and the product was TOPO cloned and transformed in competent E. coli cells. Sequences from multiple colonies were aligned and compared.

Splicing analysis: The patient's RNA sample was utilized as a template for first-strand cDNA synthesis. Colonies obtained from the cloning process were isolated and the resulting product was sequenced using M13f and M13R primers.

Results:

With cDNA sequencing, we were able to determine that the variants are *in trans*, which is consistent with the recessive disease mechanism of *CHKB* muscular dystrophy. RT-PCR demonstrated an abnormal splicing effect of the c.940C>T VUS. We observed 12 wildtype reads from the other allele, 6 reads with exon 8 skipping and splicing into a cryptic acceptor in exon 9,

and 7 reads of exon 8 splicing into a cryptic exon 9 acceptor. Both uses of the cryptic acceptor result in an out-of-frame product. Based on this, the first variant was classified as likely pathogenic based on the ACMG/AMP criteria PVS1 and PM2_supp and we reclassify the VUS as pathogenic based on the criteria PVS1(RNA), PM2_supp, and PM3.
Conclusion: Our case demonstrate how translational research was able to reclassify a variant of uncertain significance and to determine the inheritance pattern in a NGS study with ambiguous phasing.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6023F: Discovering disease variants and mechanistic networks for small-cohort GWAS with geometric deep learning ⭐

Authors:

K. Huang[1], T. Zhang[1], S. Koc[2], A. Pettet[2], C. Ruiz[1], H. Ren[1], K. Aiello[2], L. Howe[2], K. Branson[2], M. Zhang[3], J. Leskovec[1]; [1]Stanford Univ., Stanford, CA, [2]GlaxoSmithKline, San Francisco, CA, [3]Carnegie Mellon Univ., Pittsburgh, PA

Abstract:

Genome-wide association studies (GWAS) have identified thousands of non-coding variants associated with disease. However, mechanistic interpretations of these variants remain challenging, and, for many uncommon diseases, the small GWAS sample sizes limit detection power and hamper treatment development. Here, we introduce KGWAS, a deep learning framework that integrates functional information to increase GWAS power and links disease variants to a mechanistic network consisting of genes and gene programs.
KGWAS leverages a curated functional knowledge graph with 18 million links from 95 relation types across SNPs, genes, and gene programs (e.g., fine-mapped eQTLs for SNP-gene links and PPI network for gene-gene links). For a given disease, KGWAS trains a graph neural network to predict GWAS chi-square statistics from SNP functional embeddings while accounting for linkage disequilibrium. Furthermore, KGWAS recomputes marginal SNP-disease association p-values and infers putatively causal links between SNPs, genes, and gene programs using graph explainable AI methods. KGWAS produced well-calibrated p-values and accurate link prioritizations in simulations with real genotypes, and outperformed baselines in replication experiments between the subsampled (N=1K) and full GWAS (N=389K) cohorts across 21 independent UK Biobank traits (avg. 89.9%, 58.3% higher than GWAS, FINDOR, resp.).
We applied KGWAS to 554 small-cohort UK Biobank diseases (each with <5K cases). First, KGWAS discovered substantially more independent disease loci than GWAS at P=5e-8 (576 vs.

392). For instance, KGWAS discovered a novel locus, rs2155219, for the rare disease myasthenia gravis (358 cases), supported by functional data. Second, KGWAS prioritized links overlapped significantly with the gold standard links, such as expert-curated SNP-gene links for type II diabetes, breast cancer, asthma, and CAD (OR=11.2, 17.3, 8.0, 6.0 and P=1.5e-12, 2.5e-7, 3.1e-3, 7.9e-4 for the four diseases, resp.) and gene-gene links from K562 perturb-seq for mean corpuscular hemoglobin (OR=7.4, P=7.1e-128). Third, KGWAS linked disease loci to biologically plausible genes, e.g., a novel locus rs6690215 for Alzheimer's disease linked to drug targets (e.g., ESR2) and immune genes (e.g., CD55, CR1, VCP) based on Promoter Capture Hi-C interactions. Fourth, KGWAS significantly improved disease-gene prioritization using MAGMA (29.8% more replications than GWAS) and disease-cell association using scDRS (44.3% more associations than GWAS). Overall, KGWAS is a powerful method for discovering novel disease-critical variants, genes, cells, and networks, especially for small-cohort GWAS.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6024F: *Drosophila* model of *de novo PHACTR1* variant demonstrates alterations to tubulin rich structures and synaptic function.

Authors:

J. Andrews[1], P. Hall[1], S. Jangam[1], L. C. Briere[2], R. Townsley[1], Undiagnosed Disease Network, S. Yamamoto[1], D. Sweetser[2], M. Wangler[3]; [1]Baylor Coll. of Med., Houston, TX, [2]Massachusetts Gen Hosp, Boston, MA, [3]Molecular and Human Genetics, Baylor Coll. of Me, Houston, TX

Abstract:

Phosphatase and Actin Regulator 1 (*PHACTR1*) is a member of a family of Protein Phosphatase 1 (*PP1*) binding proteins which define the substrate specificity of the Phactr1/PP1 holoenzyme. Multiple *PHACTR1* variants have been previously associated with different human pathologies, including developmental and epileptic encephalopathy, West syndrome, myocardial infarction, melanoma growth, coronary artery disease, and cervical artery dissection. Here we report the identification of a human proband with a *de novo* duplication of exon 5 and parts of the surrounding introns. The proband presented with global developmental delay, hypotonia, absent speech, seizures, behavioral problems, mild prominence of ventricles and sulci, and slight dysmorphism of the left hippocampus. *Drosophila CG32264*, which is orthologous to human *PHACTR1*, is highly expressed in the nervous system of the adult fly. We have generated a *Drosophila* model of three *PHACTR1* variants, including our proband's variant and two missense mutations previously identified as causal for West Syndrome. Overexpression studies

have demonstrated significant differences in survival between our variant and the West Syndrome variants when ubiquitously overexpressed via Actin-, Tubulin-, or Pannier-Gal4 drivers. Furthermore, ubiquitous overexpression of the proband's variant was sufficient to induce forking or abnormal bends in bristles. These changes within the actin-rich bristle structures were not observed in animals overexpressing the reference human cDNA, and knockdown of *CG32264* via ubiquitous Gal4 drivers also demonstrated no effect on fly survival or morphology. An electroretinogram of the adult fly eye demonstrated significant changes in the on/off transients in animals expressing our *PHACTR1* variant compared to both *PHACRT1* reference- and *CG32264*-expressing animals. We also performed a series of flip-out experiments using a Heat-shock Gal4 line to acutely express reference and variant *PHACTR1* in a subset of cells within the salivary glands. We identified a significant change in the expression of tubulin in the cells expressing our variant, while fly salivary glands expressing either the human reference, fly *CG32264* or either West syndrome-associated variant showed no significant effect. While the mechanism of this effect has not yet been elucidated, our results clearly indicate that our *Drosophila* model is capable of identifying important changes resulting from *PHACTR1* variants and that duplications of the 5th exon putatively have a mechanism of pathogenesis separate from variants previously associated with West syndrome.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6025F: Dysregulated GluA2-Y876 PhosphorylationContributes toLossof Synaptic Upscaling,Impaired Sociability and Increased Repetitive Behaviors in Grip1 Mutant Mice

Authors:

T. Wang[1], M. Han[1], H. Tan[1], J. Kim[1,2], R. Mejias[1,3], S-I. Chiu[1,4], R. Huganir[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Korea Brain Res. Inst., Daegu, Korea, Republic of, [3]Univ. of Seville, Seville, Spain, [4]Inst. of Cellular and Organismic Biology, Academia Sinica, Taipei, Taiwan

Abstract:

Synaptic upscaling is a post-synaptic homeostatic plasticity characterized by increasing levels of surface AMPA receptors in response to reduced neural activity. Loss of synaptic upscaling has been reported in autism models of Fragile X syndrome, Mecp2, Ube3A and Shank3 deficiency, but the underlying mechanisms remain poorly understood. Glutamate receptor interacting protein 1 (GRIP1) binds the c-terminal domains of AMPA receptor 2 (GluA2) through its PDZ domains 4-6 where rare functional variants were identified in autism. We characterized mice

carrying one of these variants, *GRIP1*-I586L (murine I507L), which results in increased binding with GluA2 and accelerated GluA2 recycling in neurons. *Grip1*-I507L mice show significantly impaired sociability and increased repetitive behaviors. Electrophysiology studies identified an increase in neuronal excitability and excitatory-to-inhibiory ratio in the layer 2/3 of the medial prefrontal cortex. *Grip1*-I507L cortical neurons show loss of synaptic upscaling in response to tetrodotoxin (TTX). We next studied site-specific GluA2 phosphorylation which is known to regulate AMPARs trafficking and synaptic upscaling. *Grip1*-I507L neurons show increased basal phosphorylation of GluA2 at Y876 (GluA2-pY876) and its protein tyrosine kinase, Fyn, which is consistent with the increased binding of mutant Grip1 with GluA2. Importantly, *Grip1*-I507L neurons show no further induction at GluA2-Y876 phosphorylation in response to TTX, which correlates with reduced AMPAR accumulation at post-synaptic density and loss of synaptic upscaling. These results support a novel mechanism that dysregulated GluA2-pY876 contributes to loss of synaptic upscaling in neurons from *Grip1*-I507L mice manifesting autism-related social impairment and increased repetitive behaviors.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6026F: Enhanced mapping of gene environment interactions for vitamin D through variability quantitative trait loci

Authors:

T. Lu[1], W. Zhang[2], L. Sun[1], A. Paterson[3]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Montreal Heart Inst., Montreal, QC, Canada, [3]Hosp for Sick Children, Toronto, ON, Canada

Abstract:

Vitamin D deficiency is a common global health issue, impacting various populations and contributing to a range of health problems. The metabolism and action of vitamin D are influenced by both genetic and environmental factors, such as sunlight exposure, diet, and lifestyle. Understanding gene-environment (GxE) interactions is crucial for personalized nutrition and public health strategies for vitamin D-deficient individuals. However, existing GxE studies for vitamin D had limited statistical power and only identified interactions between several genetic variants and season of measurement.
Since GxE interactions are enriched in variability quantitative trait loci (vQTLs), we first conducted ancestry-specific vQTL discovery for serum 25-hydroxy-vitamin D using a newly developed, most powerful quantile integral linear model in the UK Biobank individuals of European (N = 313,514), African (N = 7,800), East Asian (N = 2,146), and South Asian (N = 8,771) ancestries, respectively. We identified 24 independent vQTLs (p-value < $5x10^{-8}$) in the European

ancestry population. These vQTLs were verified using multiple sensitivity analyses and were replicated in the All of Us Research Program (N = 7,258 European ancestry individuals). No vQTL was identified in non-European ancestry populations due to limited sample sizes.
Next, we tested interactions between the 24 vQTL lead variants and 20 environmental factors. We detected 55 GxE interactions involving 17 loci with a false discovery rate < 0.05. We confirmed known interactions with season of measurement underlying canonical genes, such *GC* and *CYP2A1*. Importantly, for the first time, we identified strong GxE interactions in several loci with modifiable risk factors, including body mass index, average time spent outdoors, fish intake frequency, alcohol intake frequency, smoking status, and biomarkers of liver and kidney functions. The magnitudes of these GxE interactions did not attenuate upon adjusting for season of measurement and other covariates. We also identified gene-sex interaction in a vQTL mapping to *DHCR7*, which regulates vitamin D biosynthesis. Integrating gene expression profiles, we illustrated the sex-differentiated genetic effects on vitamin D likely act through sex-biased expression of a protein isoform of DHCR7 in the skin due to alternative splicing.
In summary, this work has greatly expanded known GxE interactions affecting serum vitamin D. These findings may help to tailor interventions, such as targeted supplementation programs for at-risk individuals, leading to better prevention and management of diseases associated with vitamin D deficiency and promoting equity in healthcare.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6027F: eQTL Discovery and Analysis in Individual Pancreatic Cell Types from the Human Pancreas Analysis Program ⭐

Authors:

M. Weidekamp[1], R. M. Elgamal[2], K. Lorenz[3], K. Kaestner[3], K. Gaulton[4], S. F. A. Grant[5], B. Voight[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of California San Diego, San Diego, CA, [3]Univ Pennsylvania, Philadelphia, PA, [4]UCSD, San Diego, CA, [5]Children s Hosp. of Philadelphia/UPenn, Philadelphia, PA

Abstract:

The pancreas is composed of multiple, distinct cell types with etiologically relevant functions for diabetes. Pancreatic genomic studies have collected data largely from bulk tissue, limiting the elucidation of how each pancreatic cell type contributes to disease. Single-cell technologies coupled to tissue collection efforts have expanded for the pancreas, facilitating novel characterization of genetics, expression, and regulation of these cell types at a population scale

and across disease states. A leading effort generating this data, the Human Pancreas Analysis Program, aims to comprehensively study the human pancreas in diabetes and integrate data with the clinical history from de-identified donors. Data from this program provides an opportunity to investigate how each of the pancreas' cell types uniquely contributes to metabolic traits. Here, we report an initial discovery effort to identify genetic variation associated with gene transcript abundance - i.e., expression quantitative trait loci (eQTL) - using 333,399 cells collected from single-cell RNA sequencing of 75 pancreatic donors. Clustering analysis of this data identified pancreatic islets, pancreatic exocrine cells, and immune cell types. We initially focused on populations of alpha, beta and delta, gamma, acinar, and ductal cells, which had sufficient cell numbers across donors. Using standard analysis pipelines, we observed 1,805, 1,868, 1,050, 671, 2,143, and 1,536 significant eGenes in alpha, beta, delta, gamma, acinar, and ductal cells (permutation $P<0.05$ and q-value <0.05), respectively. Formal colocalization with GWAS for relevant traits such as type 2 diabetes, type 1 diabetes, and BMI revealed over 150 significant associations. Top associations, such as FKBP1B and UBE2E2, yielded evidence supporting their role in BMI and T2D respectively. Other identified genes, such as ENSG00000285693 for T2D, warrant further characterization to fully define their role. Future meta-analyses and comparison to datasets for open chromatin regions and chromatin capture contacts should provide additional support for the biological relevance of the eQTLs identified.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6028F: E-Selectin as a Possible Biomarker for Vaso-Occlusive Crisis in Sickle Cell Disease

Authors:

P. Gupta, R. Kumar; ICMR-Natl. Inst. of Res. in Tribal Hlth., Jabalpur, India

Abstract:

Introduction: Sickle cell disease (SCD) is a hereditary condition characterized by frequent episodes of vaso-occlusive crises (VOC), which lead to intense pain and damage to organs. E-selectin, a cell adhesion molecules (CAMs) expressed in endothelial cells may have role in pathophysiology of vaso-occlusive crises through a series of event. Further, the SNPs present in SELE gene may regulate the expression of Selectin E. Therefore, the aim of this study was to explore the possible role of Selectin E in VOC initiation and to see any possible genetic association of SELE gene SNPs with E selectin levels during crisis and at steady state. Methods: A total of 90 SCD patients who admitted in a Government Medical College, Jabalpur during September 2022 to February 2024 were recruited in the present study. Serum levels of E-

selectin were measured at three different time points (at the time of admission (90), at discharge (81) and during steady state (48) of each patient). The exons 4 of SELE gene of all samples were sequenced to find out SNP rs5362. The statistical analysis was performed to see the changes in E selectin levels at different time point. Possible association with SNPs of SELE gene with E selectin levels during crisis and steady state was also analyzed. Results: A total 90 SCD patients were recruited in the present study and mean age of the patients was 18.2 ± 8.1 years. The serum levels E-selectin were significantly higher during crisis (median =54.38, IQR=38.31 ng/mL) than at the discharge (median=80.21, IQR=21.37 ng/mL) and steady state levels (median=62.68, IQR=25.74 ng/mL) (p value<0.0001). A total of eight SNPs of SELE gene were observed in recruited patients. The SNP rs5362 were found to be associated with E selectin levels at crisis state. Conclusions: E Selectin may be utilized as a predictive biomarker for VOC in SCD patients.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6029F: First Exon Splice Donor Variants: a Meta-Analysis

Authors:

J. Murdoch, C. L. Kurtz, G. A. Maston; Quest Diagnostics, Secaucus, NJ

Abstract:

Background: Variants disrupting splice donors following the first exon pose additional complications relative to splice donor variants in other exons in the context of variant classification. The first exon is more likely to be noncoding than elsewhere in the transcript, and the usual considerations of splice variant scoring and classification may not always apply; thus, less is understood about how first exon canonical splice donor variants should be interpreted.
Objective: In this meta-analysis of published literature and public databases, data from splice assays were evaluated to better understand the impact of first exon canonical splice donor variants and help guide the interpretation of such variants.
Methods: Published literature and ClinVar database were surveyed for data on first-exon donor variants identified from a review of publicly available data for 592 genes included on Quest Diagnostics clinical genetic test panels and from HGMD data mining of splice variants. Of the 912 reported splicing variants thus identified, 667 had disrupted canonical splice donor sites at the +1 and/or +2 position; other variants had disrupted canonical splice donor sites at either the near-intronic+3 to +6 region or the final exon codon.
Results: Of the 667 canonical splice variants in 323 genes, 46 variants in 42 genes had available experimental data evaluating the impact. We found that most variants for which published data

existed were reported to result in little to no expression. When available, transcript data for these variants indicated either the utilization of a nearby cryptic splice site in the adjacent coding or non-coding sequence, or otherwise abolition of splicing at exon 1 and full retention of intron 1.
Conclusion: In either case, transcript degradation could be presumed to occur given the lack of protein expression, suggesting that variants disrupting exon 1 canonical splice donor sites can be treated as presumptive loss-of-function (with certain caveats, e.g., prediction of cryptic splice sites that, if used, could reconstitute in-frame splicing), and scored accordingly in line with the ClinGen published expansion of the PVS1 ACMG/AMP criterion.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6030F: Folate as a Potential Treatment for Lethal Ventricular Arrhythmias in TANGO2-Deficiency Disorder

Authors:

W. Xu[1], Y. Cao[1], S. Stephens[2], M. J. Arredondo[2], Y. Chen[1], W. Perez[3], L. Sun[1], A. C. Yu[1], J. J. Kim[1], S. Lalani[4], N. Li[1], F. T. Horrigan[1], F. Altamirano[3], X. Wehrens[1], C. Y. Miyake[5], L. Zhang[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp. and Baylor Coll. of Med., Houston, TX, [3]Houston Methodist Res. Inst., Houston, TX, [4]Baylor Coll. Med., Houston, TX, [5]Texas Children's Hosp. and Baylor Coll. of Med., Texas Children's Hospital and Baylor College of Me, TX

Abstract:

TANGO2-deficiency disorder (TDD) is an autosomal-recessive genetic disease caused by biallelic loss-of-function variants in the *TANGO2* gene. TDD-associated cardiac arrhythmias are recalcitrant to standard antiarrhythmic medications and constitute the leading cause of death. Disease modelling for TDD has been primarily carried out using human dermal fibroblast and, more recently, in *Drosophila* by multiple research groups. No human cardiomyocyte system has been reported, which greatly hinders the investigation and understanding of TDD-associated arrhythmias. Here, we established novel patient-derived induced pluripotent stem cell differentiated cardiomyocyte (iPSC-CM) models that recapitulate key electrophysiological abnormalities in TDD. These electrophysiological abnormalities were rescued in iPSC-CMs with either adenoviral expression of WT-TANGO2 or correction of the pathogenic variant using CRISPR editing. Our natural history study in TDD patients suggests that the intake of multivitamin/B-complex greatly diminished the risk of cardiac crises in TDD patients. In agreement with the clinical findings, we demonstrated that high-dose folate (vitamin B9) virtually abolishes arrhythmias in TDD iPSC-CMs and folate's effect was blocked by the

dihydrofolate reductase inhibitor methotrexate, supporting the need for intracellular folate to mediate antiarrhythmic effects. In summary, data from TDD iPSC-CM models together with clinical observations support the use of B-vitamins to mitigate cardiac crises in TDD patients, providing potentially life-saving treatment strategies during life-threatening events.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6031F: Founder Effect Maintains Rare Genetic Disease Variants in Pakistani Punjabi Population

Authors:

I. Tabassum[1], **M. S. Akhtar**[2]; [1]Natl. Ctr. of Excellence in Molecular Biology, Lahore, Pakistan, [2]Univ. of California, Davis, Davis, CA

Abstract:

Rare genetic diseases (RGD) are rare because of their low prevalence, but cumulatively affect more than five percent of the common masses. More than 9,000 RGDs have been documented to date. However, the rarity of these diseases poses challenges in terms of diagnosis and treatment. Treatment of most RGDs is only based on survival therapies. Such therapies are widely available in economically developed countries, but patients in underdeveloped countries strive for them. Pakistan is one of the underdeveloped countries ranked among the top fifth-largest countries in terms of population where incidence of RGDs is high due to higher rate of consanguineous marriages. In an effort to bridge the knowledge gap in genetics of RGDs, we focused on the genetic variation landscape within the Punjabi Population in Pakistan. We used snowball sampling to recruit families with established RGD diagnosis. We identified 168 families (122 Punjabi and 46 non-Punjabi) carrying rare RGD alleles and other pathogenic variants of unknown significance for 70 RGDs. The most common RGDs were progressive familial intrahepatic cholestasis (PFIC), Nieman Pick disease (NP) and Mucopolysaccharidosis (MPS). Majority of families were carrying RGD variants for metabolic and neurological diseases of autosomal recessive inheritance (88.8%). Seven of these families were diagnosed by whole genome sequencing (7.7%), 40 families by whole exome sequencing (44.4%), 10 families by targeted sequencing analysis (11.11%). All these families were carrying short variants (point mutations and small indels) except 33 (36.6%) families with copy number variants. Among whole population, 1.11% were pathogenic variants, 4.44% likely pathogenic, 23.33% missense pathogenic 23.33%, 18.88% missense likely pathogenic, 12.22% nonsense pathogenic, 1.11% nonsense likely pathogenic variants, 4.44% frameshift pathogenic variants, 3.33% frameshift likely pathogenic, 24.44% variants of unknown significance, 3.33% splicing pathogenic variants,

1.11% splicing likely pathogenic, 1.11% loss of function variant (1.11%) and 1.11% de novo in-frame pathogenic variant. By matching normal controls to families with rare diseases, we learned that 93% of RGD carrying families are consanguineous. We also found that many of the variants were novel and had never been reported in other populations, presumably pooled only in the Punjabi population. Pooling of these highly pathogenic variants and higher consanguinity rate maintain RGD variants within carrier families in a founder effect manner.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6032F: Functional effects of*BRCA1*and*BRCA2*mutations in iPSC derived breast and ovarian cancer organoids

Authors:

S. Gayther[1], A. Okimioto[1], S. Dhungana[1], M. Jones[2], F. Medeiros[3], B. J. Rimel[3], K. Lawrenson[1], N. Yucer[1]; [1]The Joe R. and Teresa Lozano Long Sch. of Med., San Antonio, TX, [2]Cedar Sinai Med. Ctr., Los Angeles, CA, [3]Cedars Sinai Med. Ctr., Los Angeles, CA

Abstract:

Long term survival in patients diagnosed with breast cancer (BC) and ovarian cancer (OC) caused by underlying genetic risk factors like *BRCA1* and *BRCA2* mutations are poor; and so it remains a priority to establish improved approaches to risk prediction, screening and early detection for prevention. Risk stratification in the population can identify individuals at greatest risk of BC and OC who can be targeted for early detection and clinical intervention. This includes high risk gene mutation carriers (*BRCA1*, *BRCA2*) and carriers of more moderate penetrance gene mutations (e.g., *BRIP1, PALB2, RAD51C RAD51D, BARD1* and *ATM*). Combined, these genes probably account for ~30% of the heritability of ovarian cancers breast cancers respectively. Basal mammary epithelial cells and secretory epithelial cells in the fallopian tube (FT) are the likely cells of origin for triple negative BC (TNGC) and high grade serous OC (HGSOC). Modeling early-stage cancer could identify clinically relevant biomarkers associated with early stage disease. Induced Pluripotent Stem Cell (iPSC)-derived human organoids can replicate physiologically-relevant processes of normal tissues and cancer progression. We have established iPSC-derived mammary and FT organoids that recapitulate the cellular differentiation and architecture of primary human mammary and FT tissues in vivo. iPSC derived mammary and FT organoids from BRCA1/2^{mut} carriers have structural abnormalities similar to pre-neoplastic precursor lesions associated with BC and OC development. We hypothesize that an individual's genetic risk of BC and OC differentially affects the function of precursor cells, their neoplastic development into preneoplastic lesions for these cancers, and then their progression to specific cancer subtypes,

TNBC and HGSOC. We have made iPSCs from women at risk of BC and OC because they carry a pathogenic BRCA1/2^{mut}, who subsequently underwent breast and/or ovarian cancer risk reducing surgery. We find that iPSC mammary and fallopian tube models from heterozygous and homozygous BRCA1/2^{mut} carriers show phenotypic characteristics of early stage BC and OC development compared to BRCA1/2 wildtype subjects. Spatial molecular profiling analysis is currently being used to identify early stage clinical biomarkers for TNBC and HGSOC to establish a functional rationale for the differences in these iPSC derived early stage models of TNBC and HGSOC. The ultimate goal is to identify precision based approaches to screen for and prevent BRCA1 & BRCA2 associated BC and OC in the population.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6033F: Genetic association of *APOL1 and GSTM1* with diabetic kidney disease (DKD) in Ghanaian and Nigerian populations ⭐

Authors:

M. Gakpey[1], B. Dzudzor[2], N. K. Agyemang[1], M-L. Sakyi[3], M. A. Apaflo[1], D. Adu[4]; [1]Noguchi Mem. Inst. for Med. Res., Univ. of Ghana, Legon, Ghana, [2]Dept. of Med. Biochemistry, Univ. of Ghana Med. Sch., Legon, Ghana, [3]West African Ctr. for Cell Biology of Infectious Pathogens, Univ. of Ghana, Legon, Ghana, [4]Dept. of Med. and Therapeutics, Univ. of Ghana Med. Sch., Legon, Ghana

Abstract:

The prevalence of several forms of chronic kidney disease is much higher in Africans and African Americans than in other races. Much of this disparity has been attributed to several risk factors including genetic risk factors. Two genetic variants of the apolipoprotein L1 (*APOL1*) gene termed G1 and G2, found primarily in individuals with recent African ancestry have been associated with focal segmental glomerulosclerosis, hypertension-related end-stage renal disease, and sickle cell nephropathy in Africans and African Americans, however, their effect in diabetic kidney diseases is not clear. Also, the absence of the glutathione S-transferase mu-1 (*GSTM1)* gene (null allele) has been associated with a higher risk of diabetes and associated complications, and its combined effects with *APOL1* high-risk alleles may increase the risk of diabetic kidney disease (DKD). Hence this study aimed to determine the association of *APOLI* and *GSTM1* polymorphisms in patients with DKD. A total of 652 unrelated individuals comprising of 331 diabetic kidney disease cases and 321 diabetic controls were selected from the H3Africa Kidney disease research study. Genomic DNA was extracted from blood samples of the study participants and genotyping for the *APOL1* G1 (rs73885319, rs60910145) and G2

(rs71785313) alleles was done using the TaqMan single nucleotide polymorphism (SNP) assay. *GSTM1* null allele was identified using the TaqMan Copy number variation assay. Statistical analysis was performed to determine allele distribution using the chi-squared test and association studies were done using logistic regression models The minor allele frequencies (MAF) of the G1, G2 and *GSTM1* alleles in DKD were found to be 0.36, 0.15 and 0.49, respectively. This was similar in Ghana and Nigeria. However, in comparing cases and controls, a significant difference was observed in the G2 MAF (15.3% cases vs. 7.9% controls, p=0.000). An increased risk of developing DKD was observed in individuals carrying 2 *APOL1* risk alleles (OR=1.57 (0.035 - 1.031), p=0.035). There was no interaction between *APOL1* and *GSTM1* variants in increasing the risk of kidney disease in diabetics in this study. In summary, the G2 allele of *APOL1* may be associated with DKD, and having two *APOL1* risk alleles increases the risk of DKD by 57%. Also, *GSTM1* and *APOL1* did not interact to increase the risk of DKD. This study demonstrated that the G2 allele of *APOL1* may be associated with DKD.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6034F: Genetic Predictors that Modify the Association Between Hippocampal Volume and Memory Performance

Authors:

A. Belachew[1], S. Biber[2], W. Kukull[2], S. Mukherjee[2], P. Crane[2], J. B. Mez[3], S-E. Choi[2], B. Klinedinst[2], M. L. Lee[2], P. Scollard[2], S. Risacher[4], M. Yu[4], C. Davatzikos[5], H. Shou[5], G. Erus[5], A. Naj[5], Y. Cui[5], L-S. Wang[5], J. Haines[6], M. Pericak-Vance[7], R. Mayeux[8], A. Durant[1], A. Regelson[9], T-C. Wang[10], D. Archer[9], L. Dumitrescu[9], T. Hohman[9]; [1]Vanderbilt Memory and Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Dept. of Med., Univ. of Washington, Seattle, WA, [3]Dept. of Neurology, Boston Univ. Sch. of Med., Boston, MA, [4]Sch. of Med., Indiana Univ., Indianapolis, IN, [5]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [6]Sch. of Med., Case Western Reserve Univ., Cleveland, OH, [7]Miller Sch. of Med., Univ. of Miami, Coral Gables, FL, [8]Dept. of Neurology, Columbia Univ., New York, NY, [9]Vanderbilt Memory and Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, [10]Vanderbilt Univ., Nashville, TN

Abstract:

Late-life memory impairment is a common symptom of multiple age-related hippocampal neurodegenerative diseases (e.g., Alzheimer's disease (AD), hippocampal sclerosis, and Limbic-predominant age-related TDP-43 encephalopathy (LATE-NC)). It can contribute to late-life memory decline, making hippocampal volume measured on MRI one of the strongest structural

neuroimaging predictors of memory performance. However, substantial heterogeneity exists in the association between hippocampal volume loss and memory performance, highlighting an exciting opportunity to identify genetic factors driving this heterogeneity.
The present study investigates genetic variants that modify the associations between hippocampal volume and memory performance in 2,371 non-Hispanic White (NHW) and 372 non-Hispanic Black (NHB) participants from the National Alzheimer's Coordinating Center (NACC; mean age =72.28, 56.14% female), a cohort of aging and AD. A genome-wide interaction study between common genetic variants and hippocampal volume on baseline memory performance was performed using linear regression, covarying for age, sex, education, intracranial volume, and the first ten ancestral principal components to control unmeasured population stratification. The GWAS was run separately in each ancestry group.
Among NHW, a genome-wide interaction signal with hippocampal volume was found on chromosome 8 (rs6980724, p= 4.976×10^{-8}, β = -1.3×10^{-4}, MAF=0.3448, and rs7014815, p= 3.875×10^{-8}, β = -1.3×10^{-4}, MAF=0.3433), which are eQTLs for KHDRBS3. Minor allele carriers of each variant showed a stronger association between hippocampal volume and memory performance and lower expression of KHDRBS3, suggesting low levels of this gene may increase susceptibility to the clinical manifestation of hippocampal degeneration. KHDRBS3 binds RNA, facilitates RNA splicing, and is associated with signal transduction. Additionally, we identified 2 genome-wide significant loci on chromosome 5 (rs2410722, p=4.97×10^{-8}, β = 3.45×10^{-4}, MAF=0.203) and chromosome 12 (rs17395410, p=2.879×10^{-8}, β =4.1×10^{-4}, MAF=0.122) among NHB participants, although no significant eQTLs were found for these SNPs. Future work will expand to include a longitudinal change in memory performance, additional cohorts to increase our sample size, and seek independent replication.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6035F: Genetic Regulation of Vitamin B12 Concentrations in Plasma and Human Milk Across 4 Study Sites: Data from the MILQ Study

Authors:

C. Arrington[1], M. Islam[2], G. Kac[3], D. Mucci[3], K. F. Michaelsen[4], C. Mølgaard[4], S. E. Moore[5], L. H. Allen[1], B. J. Bennett[1]; [1]USDA-ARS-WHNRC, Davis, CA, [2]Intl. Ctr. for Diarrhoeal Disease Res., Dhaka, Bangladesh, [3]Univ.e Federal Do Rio De Janeiro, Rio De Janeiro, Brazil, [4]Univ. of Copenhagen, Copenhagen, Denmark, [5]King's Coll., London, United Kingdom

Abstract:

The Mothers, Infants and Lactation Quality (MILQ) study was conducted in Bangladesh, Brazil, Denmark, and The Gambia. The study aimed to develop reference values for nutrients and other constituents of human milk. This multi-center cohort study was registered at clinicaltrials.gov as NCT03254329. The current analysis aims to determine the consistency of genotype-phenotype associations between the FUT2 genotype and plasma and human milk concentrations of vitamin B12. The study enrolled 960 mothers, of which 956 have genetic data measured from the buffy coat using the Illumina global screening array. Plasma and human milk samples were analyzed and quantified for vitamin B12 by competitive protein binding assay yielding a sample size of n=906 for plasma analysis and n=899 for human milk analysis. Linear models with ANOVA were used to estimate the association of FUT2 with vitamin B12 after adjustment for covariates including participant site and visit as well as maternal pre-pregnancy parity, BMI, and age.rs492602, a single nucleotide polymorphism in FUT2, has been associated with plasma vitamin B12. Across the 4 sites, the alternate allele (G) frequency of rs492602 is 0.38, the highest in The Gambia (0.55) and the lowest in Bangladesh (0.21). The overall mean (SD) of plasma B12 was 427.8 (194.5) pmol/L, while milk B12 was 273.6 (204.2) pmol/L. There was a significant additive association between rs492602 and plasma B12 concentrations across sites at each visit. For example, the overall association of rs492602 was significant by ANOVA (p = $2.2x10^{-10)}$ with 2 alternate alleles significantly associated with higher levels of plasma B12 (transformed difference GA vs AA: 0.042, $p = 1.9x10^{-10}$; GGvAA: 0.22, $p < 2.2x10^{-16}$; GGvAG: 0.18, $p = 2.0x10^{-9}$). While each site exhibited a significant contribution of rs492602 to plasma B12, the most significant association was in Bangladesh ($p = 1.910^{-6}$). There is minimal evidence of an association between rs492602 and milk B12 concentrations across sites; however, there is a marginally significant association (ANOVA p = 0.077) and women with G alleles have significantly higher milk vitamin B12 (transformed difference GA vs AA: 0.00013, p = 0.0038; GGvAA: 0.00026, $p = 2.3x10^{-7}$; GGvAG: 0.00013, p = 0.0067).These data from four contrasting settings confirm previous findings of an association between rs492602 and plasma vitamin B12 concentrations. However, our data show weak evidence in human milk suggesting this SNP, rs492602, may not contribute the same genetic regulation of vitamin B12 in human milk. These results suggest that there may be different genetic controls over the composition of vitamin B12 in human milk compared to plasma.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6036F: Genetically regulated, cell-type-specific transcriptional responses to anticancer drugs in heterogeneously differentiating cultures of cardiac cells

Authors:

E. McIntire, K. Barr, O. Allen, N. Gonzales, J. Burnett, Y. Gilad; Univ. of Chicago, Chicago, IL

Abstract:

Doxorubicin (DOX) and 5-fluorouracil (5-FU) are commonly prescribed to treat a variety of cancers. Though efficacious, up to 48% of patients treated with DOX and 30% of patients treated with 5-FU experience adverse cardiac side effects, suggesting that risk for drug-induced cardiotoxicity (CT) may be genetically regulated. Indeed, we previously found that genetic variants that regulate cardiomyocyte transcriptional responses to DOX *in vitro* can be used to predict individual risk for DOX-induced CT in cancer patients. We reasoned that by characterizing transcriptional responses to DOX in other cell types relevant for CT, we could identify additional drug-responsive gene regulatory loci (i.e., response eQTLs) to improve risk prediction. We included 5-FU in our study as it has been less explored in a gene regulatory context compared to DOX, despite also being widely prescribed and associated with cardiotoxic effects. By comparing the transcriptional effects of these two distinct compounds (an anthracycline and an antimetabolite, both of which are associated with diverse CT effects), we sought to identify general and drug-specific mechanisms of CT.

To efficiently assay multiple CT-relevant cell types across many individuals, we modified our cardiomyocyte differentiation protocol to produce heterogeneous differentiating cultures of cardiac cell types (cardiac HDCs). Cardiac HDCs are three-dimensional, rhythmically contracting cell aggregates that, after just 2 weeks in culture, include cardiomyocytes, cardiac fibroblasts, myofibroblasts, neural crest cells, epithelium, neural progenitors, neurons, gut endoderm, epicardium, endothelium, hepatocytes, and hematopoietic cells. We independently treated cardiac HDCs from 45 individuals with 1 µM DOX and 5 µM 5-FU for 24 hours, followed by single-cell RNA sequencing to capture transcriptional responses. We targeted 12,000 cells per individual (4,000 cells per condition, plus a vehicle control) for over half a million cells (more than 20 billion reads in total). Our preliminary analyses reveal thousands of drug-responsive genes that are differentially expressed between treatment conditions, many of which show cell-type-specific expression patterns. These data allow us to identify and characterize drug response eQTLs, which facilitate the exploration of mechanisms of DOX- and 5-FU-induced CT and the development of a retrospective genetic predictor of CT in cancer patients. Ultimately, our work may enable clinicians to provide personalized cancer treatments based on patient genotype.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6037F: Genome-Wide Association Study of Individual Vitamin D Metabolite Levels and Ratios Replicates Known Variants on Chromosome 6 and Identifies Novel Variants in the Diabetes Autoimmunity Study in the Young (DAISY)

Authors:

P. Carry[1], H. Smith[1], R. Johnson[1], T. Buckner[2], L. Vanderlinden[1], O. Fiehn[3], S. Slack[4], B. Defelice[5], S. Rich[6], S. Onengut-Gumuscu[6], K. Waugh[1], K. Kechris[1], M. Rewers[1], J. Norris[7]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Northern Colorado, Greeley, CO, [3]Univ. of California Davis, Aurora, CO, [4]Univ. of Colorado, Aurora, CO, [5]Chan Zuckerberg Biohub Network, San Francisco, CA, [6]Univ. of Virginia, Charlottesville, VA, [7]Colorado Sch Pub Heal, Aurora, CO

Abstract:

Introduction: Vitamin D (VitD), typically assessed based on 25OHD3, has pleiotropic effects. GWAS have provided insight into the genetic basis of 25OHD3; however, 25OHD3 is an upstream metabolite that may not fully represent the biologically active metabolite, 1α,25(OH)2D3. Genetic regulation of downstream metabolites is not well understood. We evaluated individual VitD metabolites and ratios of upstream metabolites to 1α,25(OH)2D3 to test the hypothesis that, by capturing variability in the conversion of upstream to downstream metabolites, ratios will provide insight into VitD pathway regulation. **Methods**: We conducted a GWAS of plasma VitD metabolites quantified with a custom LC-MS/MS panel. We quantified 1α,25(OH)2D3, 24,25(OH)2D3, 25(OH)D3, 25(OH)D2, 3-epi-25(OH)D3, and D3 and evaluated ratios of upstream metabolites to 1α,25(OH)2D3. Given longitudinal data, random intercepts from linear mixed models adjusted for age, sex, season, and number of visits were used to represent average metabolite levels. Samples were genotyped using the TEDDY-T1D Exome array with imputation using the Trans-Omics for Precision Medicine (TOPMed) multi-ancestry reference panel. Statistical analyses were conducted using PLINK (QC) and REGENIE (association analysis under unbalanced case-control ratios). Sex, first 5 ancestral PCs, and the high-risk genotype (HLA *DRB1*03:01*/*DRB1*04:01*) were adjusted for in genetic models. Genome-wide significance was defined by $P < 5.0 \times 10^{-8}$. **Results**: We replicated genome-wide significance between 25OHD3 and six previously identified SNPs (PMID: 32059762) near *GC* on chromosome 6 (rs844806, rs705120, rs705121, rs843005, rs222035, and rs222040). SNPs at two novel loci were associated with the ratio of 3-epi-25(OH)D3 to 1α,25(OH)2D3: rs149231007 (13q14.3, intronic in *GUCY1B2*, $P=4.3 \times 10^{-8}$) and rs182781381 (2q32.2, intergenic between *NAB1* and *GLS*, $P=3.89 \times 10^{-}$

[8]). **Conclusion**: We confirmed known SNPs near *GC* are correlated with 25OHD levels in children and adolescents. 25(OH)D3 is converted to downstream metabolites, 3-epi-25(OH)D3, 24,25(OH)2D3, or 1α,25(OH)2D3, via cell-type specific pathways. We also identified variants that contribute to increased levels of 3-epi-25(OH)D3 relative to 1α,25(OH)2D3. Increased levels of this ratio may reflect preferential conversion of 25OHD3 to the less active 3-epi-25(OH)D3 and thus, may identify a subgroup of patients at risk for VitD deficiency not detected by conventional 25(OH)D3 assays. Additional research is need to replicate the role of *GUCY1B2* and genes in the 2q32.2 region (identified by rs182781381) in VitD pathway regulation and characterize their functional impact.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6038F: Genome-wide CRISPR screen identifies genes that affect hepatic steatosis and implicate new targets for MASLD

Authors:

P. Ponnandy[1], C. Raut[2], R. Hussain[1], A. Irshaid[1], B. Halligan[1], E. Speliotes[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan at Ann Arbor, Ann Arbor, MI

Abstract:

Background: MASLD has been estimated to affect 30% of the adult population worldwide. MASLD has few effective targeted therapies in part due to a poor understanding of its etiology. MASLD is partially heritable, and we have identified 17 loci associated with nonalcoholic fatty liver disease showing that genetic changes can affect MASLD. Here we carry out a genome-wide and sub-genomic loss of function CRISPR screen on human hepatoma cells to identify genes that regulate hepatic steatosis to possibly validate previous GWAS implicated genes as well as identify new genes that can affect hepatic steatosis. Methods: We transfected the GeCKOv2 library into human Huh7 cells. After selection and passage of cells for two weeks we used Lipitox to stain cells for lipids and separated the pooled cells into a top 10%, bottom 10% and unsorted groups. We used next generation sequencing to identify guides enriched in the top and bottom 10% using MAGeCK. We took the top 300 and bottom 300 guides and followed them up in directed screens in Huh7 and HepG2 cells using a similar enrichment approach. Results: We identified 29 genes that increase, and another 20 genes decrease lipid accumulation in liver cells. We identified PNPLA2 which releases triglycerides from lipid droplets as increasing steatosis and DGAT2 which promotes de novo lipogenesis as decreasing steatosis in our loss of function screen. We also identify a previous GWAS hit TM6SF2 which promotes loading of lipids into VLDL as increasing steatosis in our screen. Conclusions: A whole genome CRISPR screen

with follow up has identified positive and negative drivers of hepatic steatosis. These include genes that affect release of lipids from lipid droplets, load lipids onto VLDL, and participate in de novo lipogenesis. Since hits such as DGAT2 are targets for therapeutic intervention of MASLD, many of our new hit offer possible new therapeutic targets for the treatment of this disease.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6039F: Genome-wide glomerular allele-specific expression in human proteinuric kidney tissue

Authors:

A. Onuchic-Whitford[1,2,3], J. Sung[1,2], E. Sakkas[2], M. McNulty[2], A. Greenberg[2], J. Yoon[4,5], S. Badina[1,2], NEPTUNE (Nephrotic Syndrome Study Network), M. Sampson[2,3], D. Lee[3,2]; [1]Brigham and Women's Hosp., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [5]Gangnam Severance Hosp., Seoul, Korea, Republic of

Abstract:

Background: Regulation of gene expression plays a key role in disease pathophysiology and contributes to penetrance of pathogenic variants. Allele-specific expression (ASE), where each copy of a gene contributes differentially to total gene expression, is a robust measure of cis-regulatory effects. While this phenomenon is well described in other organs, ASE analysis in kidneys is limited, mainly to normal tissue. Here, we successfully generated a genome-wide, high-quality ASE dataset in glomerular tissue from individuals with proteinuric kidney disease in NEPTUNE, using a rigorous computational pipeline. **Methods:** We leveraged whole-genome sequencing (WGS) paired with kidney biopsy-derived glomerular RNA-seq from 240 NEPTUNE participants to robustly quantify ASE events across the genome. WGS was population-phased using WhatsHap and SHAPEIT4 with a large reference panel of phased haplotypes (~500K individuals) from UK Biobank. RNA alignment was done with STAR/WASP to address read mapping bias; duplicate reads were removed. phASER was used to integrate phased WGS and RNA-seq to generate SNP- and gene-level allelic expression data for all individuals. We applied the same pipeline to 88 individuals from GTEx with kidney cortex samples and paired WGS/RNA data. **Results:** Using an RNA-seq read count >=20 cutoff, we calculated allelic expression for a mean of 7,173 genes per individual. Applying the binomial test with correction for multiple comparisons, we identified genes with significant ASE (FDR<0.05). 19,175 genes had ASE observed in ≥1 person and 14,148 of these genes were protein-coding. Individuals had a mean of 1,433 ASE genes and 55 genes with monoallelic expression. *PLA2R1* ranked 8th (genome-

wide) as most common ASE gene among NEPTUNE participants, with *PLCG2* ranking 12th and *NPHS2* ranking 13th. While these key glomerular genes were also significantly expressed in GTEx kidney samples, they had substantially lower ASE among individuals, ranking as the 954th (*PLA2R1*), 1719th (*PLCG2*) and 206th (*NPHS2*) most common ASE genes. **Conclusion:** We generated a genome-wide dataset of ASE in proteinuric glomerular tissue through a rigorous multi-step process integrating several computational tools. Our initial analysis reveals overrepresentation of ASE in known proteinuric kidney disease-related genes among NEPTUNE participants.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6040F: Genome-wide quantitative trait locus mapping of serum proteins using whole-genome sequencing in a Japanese population

Authors:

Y. He, M. Koido, C. Tanikawa, A. Kanai, Y. Suzuki, Y. Oda, A. Nagai, T. Morisaki, Y. Okada, N. Kumasaka, K. Matsuda, Y. Kamatani; The Univ. of Tokyo, Tokyo, Japan

Abstract:

Circulating proteins, as mediators and biomarkers of complex traits, play important roles in a wide range of biological processes. The investigation of such protein levels provides valuable insights into genetic regulation and disease-associated genes. Using Olink Explore 3072 platform, we conducted proteogenomic profiling of 2,925 proteins in serum samples from two whole genome sequencing (WGS) cohorts in BioBank Japan first cohort (BBJ), totaling 2,886 Japanese individuals. Genome-wide protein quantitative trait locus (pQTL) mapping was performed in each cohort using a whole genome regression model, which was then combined by a fixed-effect meta-analysis. We identified a total of 1,360 associations at the study-wide significance threshold ($p < 1.7e-11$) for 1,120 proteins. At genome-wide significance threshold ($p < 5e-8$), 3,078 associations were identified for 1,838 proteins, of which 1,163 were cis (within 500kb from gene body) associations, and 847 were rare variants (minor allele frequency < 0.01). Out of the 3,078 associations, we identified 574 common or low frequency (minor allele frequency >= 0.01) pQTLs and 754 rare associations that were not reported in large-scale studies conducted in Europeans (using imputation or whole exome sequencing) or East Asians (imputation). Besides common variants at *CFH*, *ABO*, *ALDH2*, and *FUT2*, we additionally identified rare variants (For example, a novel East-Asian specific rare missense pQTL at *TMEM201*) showing high pleiotropy (Number of associated proteins > 10). The comparison of pQTL effect sizes between Europeans and East Asians also revealed uncorrelated population-

specific associations in cis regions for proteins including *FCGR2B*, *TREH*, *GIPC2*, and *RNASE3*, suggesting the distinct regulation of these proteins across ancestries. Mendelian randomization across 47 complex diseases in BBJ suggested causal relationships for 144 protein-disease pairs (false discovery rate < 0.05), of which 23 were further supported by colocalization analyses (posterior inclusion probability for the hypothesis that both traits are associated and share a single causal variant > 0.8). We identified significant causal associations of ALDH2 with multiple complex diseases including cardiovascular diseases, stroke, chronic obstructive pulmonary disease, and liver cirrhosis, marking its extensive pleiotropic effect on complex diseases in East Asians. In summary, this study identifies both common and rare pQTLs in East Asians, provides insights into the cross-ancestry difference in protein level regulation, and prioritizes protein targets for drug development for complex diseases.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6041F: Haploinsufficiency versus Gain of Function in AHDC1 and the Xia-Gibbs Syndrome

Authors:

J. Hu[1], A. Hansen[2], K. Ward[2], M. Murugan[1], M. Dawood[1,3,4], J. E. Posey[3], J. R. Lupski[3,1,5,6], M. Wangler[3], R. Gibbs, A.[1,3]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]Geneial, Houston, TX, [3]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [4]4Med. Scientist Training Program, Baylor Coll. of Med., Houston, TX, [5]Texas Children's Hosp., Houston, TX, [6]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX

Abstract:

Xia-Gibbs Syndrome (XGS; MIM# 615829) is a rare neurodevelopment disorder (NDD) caused by pathogenic variants in the single coding exon AT-Hook DNA-Binding Motif-Containing 1 (*AHDC1*) gene. Hallmark XGS features include hypotonia, delayed motor milestones, global developmental delay, and intellectual disability (PMID: 34902026). Since the first characterization of XGS in 2014 (PMID: 24791903) by our group, more than 500 individuals have been diagnosed worldwide. A key question is whether the pathogenicity is via haploinsufficiency or a gain of function/dominant negative mechanism. To explore this, we have compiled matched genotype to phenotype data from 188 XGS individuals in a HIPAA-compliant Registry - now supported by a cloud-based tool developed by Geneial - that enables phenotype/genotype correlations and supports multi-omic studies. The Registry has allowed us to document XGS clinical heterogeneity and identify core phenotypes (PMID: 29696776, 33644933). Further, we have shown that both truncating and missense variants cause XGS

(PMID: 33644933, 34950897). We have shown the mutation position is correlated with variation in the XGS phenotypic spectrum. Further, missense variant pathogenicity is highly dependent on position relative to functional domains (PMID: 34950897). We experimentally confirmed that truncating alleles in *AHDC1* escape nonsense-mediated decay, consistent with it being a single exon gene (PMID: 36054313). In one individual with a large deletion (350kb) including some of *AHDC1*, the gene's expression from the unaffected haplotype was elevated by a compensatory regulatory mechanism (PMID: 36054313). All these observations support the model of a gain of function pathogenesis. However, others have reported pathogenic large deletions and *AHDC1* ablation has been shown to modulate ectoderm/mesoderm differentiation in cellular models (PMID: 35585237). Therefore, the combined data support variable expressivity of *AHDC1* via multiple mechanisms of pathogenicity with an early developmental impact from haploinsufficiency but the predominant disease mechanism resulting from a gain of function, truncated *AHDC1* allele. Saturated understanding of all possible variant alleles in *AHDC1* would provide a complete understanding of relevant molecular consequences manifesting clinical disease. Resolution of the exact contribution of each driver of pathogenicity is essential for guiding therapeutics.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6042F: Harnessing virtual Capture C to identify GWAS target genes for the African ancestry-specific variant at *GRIN3B* in Alzheimer's disease

Authors:

L. Wang[1], W. Xu[2], F. Rajabli[1], A. Ramirez[1], S. Moura[1], K. Celis[3], M. Gearing[4], D. Bennett[5], S. Weintraub[6], C. Geula[6], T. Schuck[7], D. Davis[1], W. K. Scott[1], J. Haines[8], G. S. Byrd[9], C. Reitz[10], D. Dykxhoorn[1], J. Young[1], A. Griswold[1], F. Jin[11], J. Vance[1], M. Pericak-Vance[12]; [1]Univ. of Miami, Miami, FL, [2]Case Western Resreve Univ, Cleveland, OH, [3]Univ Miami, Miami, FL, [4]Emory Univ., Atlanta, GA, [5]Rush Univ., Chicago, IL, [6]Northwestern, Chicago, IL, [7]Univ. of Pennsylvania, Philadelphia, PA, [8]Case Western Reserve Univ, Cleveland, OH, [9]Wake Forest Univ., Winston-Salem, NC, [10]Columbia Univ, New York, NY, [11]Case Western Reserve Univ., Cleveland, OH, [12]Univ. of Miami Miller Sch. of Med, Miami, FL

Abstract:

Introduction: Genome-wide association studies (GWAS) have made great contributions to our understanding of Alzheimer's disease (AD). Identifying the causal variants and target genes of GWAS findings, however, remains to be a challenge in the post-GWAS era. Lately, a multi-

ancestry GWAS analysis done by AD Genetics Consortium (ADGC) and The Collaborative for AD Research (CADRE) reported a genome-wide significant association at rs115882880 (OR=1.5, P=2.45x10-10) in African American (Rajabli F et al. medRxiv 2023 Jul 8). Herein, we applied functional genomic study to identify the causal variants and target genes of the African ancestry-specific GWAS finding. **Methods:** Variants that are linkage disequilibrium (LD, R2 >0.6) with rs115882880 in the African population from the 1000 Genome Project were extracted using the LDproxy Tool. Variants were prioritized based on 1) colocalization with open chromatin region (OCR, annotated by snATAC-seq in frontal cortex); 2) transcription factor binding (JASPAR 2024 TFBS collection). To map target gene(s) of candidate causal variants, promoter-enhancer interactions were annotated using virtual Capture C analysis of the Hi-C data in frontal cortex and induced pluripotent stem cells (iPSCs)-derived microglia and spheroids. **Results:** Fine-mapping revealed rs75122091 (R2=0.85 with rs115882880, 23 bp away from rs115882880) as a candidate causal variant. It co-localizes with an OCR in brain cells and disrupts a Wt1 binding site, suggesting a regulatory potential of the variant. Both rs75122091 and rs115882880 reside in the first intron of GRIN3B on chromosome 19. The intronic region containing both variants interacts with several promoters in the vicinity, including ARID3A, R3HDM4, CFD, and MIDN, but not GRIN3B itself or ABCA7 nearby. The chromatin interactions were detected in frontal cortex and confirmed in iPSC-derived cells, suggesting this GWAS locus is involved in regulating expression of multiple genes nearby. **Conclusions:** Leveraging on the omics data to map 3D genome regulatory architecture, our study supports a function role of an African ancestry-specific variant in AD and implicate several target genes. This demonstrates the utility of virtual Capture C in identifying the causal genes of GWAS loci.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6043F: High-throughput genotype-phenotype association data aid variant classification in a diverse cohort

Authors:

C. Zeng[1], L. Bastarache[2], E. Venner[3], T. Ferrara[4], T. Tran[5], J. Yang[5], J. Denny[1]; [1]NIH, Bethesda, MD, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]HGSC - Baylor Coll. of Med., Houston, TX, [4]NIH, Bethesda, MD, [5]NHGRI, Bethesda, MD

Abstract:

Most rare variants are classified as variants of uncertain significance (VUS), hindering clinical applications. Large cohorts like the UK Biobank and the *All of Us* Research Program have uncovered >1 billion variants connected to phenotypic data. We evaluated phenotype risk score

(PheRS) in *All of Us* to aid in classifying rare variants (allele frequency <1%) for 81 genes in the ACMG secondary finding gene list (v3.2). PheRS aggregates multiple electronic health record (EHR) phenotypes related to a given gene into a weighted score specific to each gene tested. We tested the ability of PheRS to "interpret" pathogenicity of unknown genetic variants. Using 206,173 participants with both EHRs and whole genome sequencing data in *All of Us*, including 45,772 of predominantly African ancestry and 36,152 of Admixed Americans, we identified 2.67 million rare variants in genes in the ACMG secondary finding gene list (v3.2). We curated a gold standard classification of variants using classifications data from multiple resources, including 586 pathogenic (P), 151 likely pathogenic (LP), 44,388 VUS, 25,426 likely benign (LB), and 9,795 benign (B) variants. We next evaluated the association of each variant with its corresponding PheRS. Based on the estimated effect sizes and the estimated penetrance of the genes, we created cutoff PheRS values to mark variants with strong evidence of pathogenicity ("PheRS-path") or benign ("PheRS-benign") and categorized the other PheRS scores as "PheRS-unknown." Using these cutoffs, the PheRS sensitivity was 20.0% and positive predictive value (PPV) was 100% for gold-standard pathogenic variants. The sensitivity and PPV for gold-standard benign variants was 1.4% and 100%, respectively. Applying PheRS to all variants, we provided evidence for potential pathogenicity (PheRS-path) for 1,330 VUS and 827 LB variants (~3% of all LB variants). Many of the non-P/LP variants classified as PheRS-path were mainly found in participants of predominantly African or Admixed American genetic ancestry. The estimated percentage of participants carrying PheRS-path variants were increased from those with known P/LP variants in every ancestral group, with the greatest increases for participants of predominantly African ancestry (from 1.3% to 4.7%) and Admixed Americans (from 3.5% to 5.2%). For example, NM_020975.6*(RET)*:c.2497C>T, which is VUS, had a z-score of 3.5 (classified as PheRS-path) and mainly found in those with African ancestry. This study illustrates the utility of large, diverse cohorts and approaches like PheRS to facilitating high-throughput variant classification across diverse populations.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6044F: Identification of variants associated with kidney dysfunction in individuals of African ancestry with sickle cell disease

Authors:

F. Osman[1], K. Esoh[1], V. Ngo Bitoungui Essomba[2], J. Casella[1], A. Wonkam[1]; [1]Johns Hopkins Sch. of Med., Baltimore, MD, [2]Univ. of Dschang, Dschang, Cameroon

Abstract:

Background: Sickle cell disease (SCD) is the most-common monogenic recessive disease in humans, annually affecting almost 300,000 newborns worldwide. In SCD, the abnormal, sickled Hemoglobin (*HbS*) tends to polymerize in red blood cells under hypoxic conditions, leading to damage across multiple organs including the kidneys. Sickle Cell Nephropathy (SCN) occurs in 5-18% of SCD patients. 30% of adult patients develop chronic renal failure, contributing to a large proportion of morbidity and mortality in SCD. A systematic identification of genetic factors associated with variations in kidney dysfunctions in SCD patients could promote effective management. In this study, we conducted the first genome-wide analysis of kidney function in patients living with SCD from Africa (Cameroon). **Methods**: We used the H3Africa SNP genotyping array (Illumina) to genotype DNA samples from our cohort. Chromosomes were phased and imputed via TOPMed, 1000 Genomes, and African Genome Resource (Sanger) imputation panels. Standard quality control measures were applied to samples and SNPs prior to association analyses using PLINK2. A generalized linear model was then used to perform a genome wide association study (GWAS) of estimated glomerular filtration rate (eGFR) and urine protein/creatinine ratio (uPCR) levels in patients with SCD from Cameroon (n = 571), who have not been previously exposed to hydroxyurea treatment, followed by replication with data from an independent African American (n = 484) with similar quality control measures. **Results:** Median eGFR (CKD-EPI 2021) in our cohort was 150.7 mL/min/1.73m^2. Median uPCR was 170.45 mg/g. Kidney functions were strongly correlated with patient age, body-mass-index, blood pressure, and sex. We identified two novel loci including *CCNT4* ($p = 5 \times 10^{-8}$) and *FAR1* ($p = 1.47 \times 10^{-6}$), associated with eGFR, which were replicated in an independent analysis of the African American cohort of individuals with SCD ($p < 1.5 \times 10^{-5}$) Additionally, we replicated two other loci in *LRP1B* ($p = 1.95 \times 10^{-5}$) and *VWF* ($p = 1.56 \times 10^{-5}$), that were recently associated with eGFR and proteinuria in an adult African Americans group living with SCD. **Conclusion:** This first GWAS in an African cohort Sickle Cell Nephropathy using an array enriched with variants from African genomes enhanced discovering novel loci (*CNT4* and *FAR1*) associated with estimated glomerular filtration rate and urine protein/creatinine ratio. These findings will be further investigated with functional studies in cell/animal models.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6045F: Integrated computational and experimental biophysics reveal shared mechanisms for genetic variants across cancer and RASopathies

Authors:

J. Wagenknecht, D. Jensen, H. Neshatul, J. Licklider, B. Volkman, B. Smith, R. Urrutia, M. Zimmermann; Med. Coll. of Wisconsin, Milwaukee, WI

Abstract:

The Kirsten rat sarcoma virus (KRAS) GTPase is extensively studied due to its significant role in cancer, where activating variants induce cellular proliferation and transformation. Yet, KRAS variants are implicated in numerous RASopathies, defined as rare diseases caused by dysregulation of the RAS/MAPK pathway. Most studies concentrate on the effects of KRAS-G12D because it is the most recurrent mutation and a common benchmark for the field. However, understudied variants in KRAS may have similar or different mechanisms of pathogenicity compared to G12D, as previous work by our lab and others has demonstrated KRAS mutation's heterogeneity. In this study we aim to identify variants which mimic specific effects of G12D, as well as to distinguish their biomolecular features. To this end, we conducted two types of molecular dynamics (MD) simulations, gaussian accelerated and unbiased, of WT KRAS and 86 distinct missense genetic variants which are observed in both cancer and RASopathies. We analyzed these trajectories, leveraging established biophysical and biochemical characteristics of and benchmarks for RAS-family proteins. Specifically, we were able to scale these heterogeneous mutations in terms of their propensity to change Switch I or Switch II conformations, destabilize the nucleotide pocket, and lose specific ligand-mediated contacts. To further expand and support our analysis, we performed ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) Nuclear Magnetic Resonance (NMR) Spectroscopy and thermal stability assays. Our analysis revealed heterogeneity across the mutations. A subset of mutations act by biasing the protein structure, making mutated GTPases that are less sensitive to nucleotide state; of those variants, some mimic the biophysical characteristics and structural conformations of GTP-bound KRAS even when bound to GDP, whereas the others adopt GDP-bound KRAS characteristics and conformations even when GTP-bound. Notably, results obtained from NMR and MD were highly congruent. Additional subsets of mutations are identified that we anticipate act by distinct molecular mechanisms. Our data shows that each subset has different conformational biases that may bind differently to small molecules, affording the potential for predicting which G12-based treatments are more likely applicable to non-hotspot mutations. Further, our exploration of the shared and distinct attributes of a broad set of KRAS mutations reveals common mechanisms of pathogenicity between somatic and germline mutations alike.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6046F: Investigating the functional relationships between mitochondrial DNA variation and nitric oxide regulation through multi-omics integration

Authors:

K. McKessy[1], W. Shi[1], N. Lake[2], D. Arking[1]; [1]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Yale Univ., New Haven, CT

Abstract:

Previous work has documented the functional roles of heritable genetic polymorphisms in concert with somatic mutations in the etiology and progression of metabolic, neurodegenerative, and autoimmune disorders. A smaller subset of genetic changes in the mitochondrial DNA (mtDNA) genome are functionally linked to the manifestation of genetic disorders originating from markedly compromised mitochondrial function. Our goal is to identify functional relationships between mtDNA variation and proteins and metabolites, to elucidate the impact of mtDNA on mitochondrial function and human health.
Here, we assess both inherited variation-in the form of mtDNA haplotype- and somatic variation via mtDNA heteroplasmy. We defined a variant allele fraction of 5% as heteroplasmic, and functionally characterized mtDNA single nucleotide variants using a constraint-based score, mitochondrial local constraint score sum (MSS).
This study integrates proteomic (n=50,000), metabolomic (n=300,000), and genomic (n=500,000) data from the UK Biobank to investigate how mtDNA variation affects mitochondrial function and nitric oxide (NO) metabolism. Our findings reveal that mtDNA haplotype is significantly associated with key NO-related proteins such as NOS1 (p=3.7e-34), while MSS correlates with DDAH1 (p=7.2e-6), a crucial regulator of NO production. Additionally, omega-3 fatty acids, known for their cardiovascular benefits, show significant associations with mtDNA haplotype (p=8e-4) and regulate NOS3 (endothelial nitric oxide synthase), highlighting their potential role in modulating mitochondrial function through interactions with nitric oxide pathways. Pathway enrichment analyses reveal enrichment among association results for NO metabolism pathways, including NOS3 activation (p=8e-4) and guanylate cyclase stimulation (p=4.7e-4), as well as broader metabolic processes tied to mitochondrial energy production. Our ongoing Mendelian Randomization studies examine the causal links between NO-related proteins, metabolites, and health-oriented phenotypes such as cardiovascular diseases and mitochondrial disorders. This comprehensive multi-omics analysis provides new insights into the role of mtDNA variations in mitochondrial function and NO regulation, and may identify potential therapeutic targets for NO-related diseases.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6047F: Lentivirus-based MPRA allows fine-mapping of transcription factor binding sites within the promoter of a cardiomyopathy-associated gene. ⭐

Authors:

A. Renberg, O. Merritt, S. Coppersmith, A. Tovar, S. Parker, J. Kitzman, A. Helms; Univ. of Michigan, Ann Arbor, MI

Abstract:

Hypertrophic cardiomyopathy (HCM) is a primary heart muscle disorder characterized by increased heart wall thickness that affects approximately 1/500 people and is largely genetic in origin. Autosomal dominant coding variants in sarcomere genes are collectively the most common cause of HCM, but the rate of genetic test positivity is only ~50%. Haploinsufficiency due to loss of function coding variants in the sarcomere gene, Myosin Binding Protein C3 (*MYBPC3*) is the most common cause of HCM. However, effects of loss of function noncoding single nucleotide variants (ncSNVs) have not been tested since the noncoding regulatory space has not been defined. Even in patients with identified coding variants, incomplete penetrance and highly variable expressivity suggest unaccounted-for genetic contributors. We hypothesize that ncSNVs disrupting transcription factor binding sites (TFBSs) in *MYBPC3*'s CREs, consistent with transcriptional loss of function, are additional contributors to HCM. To investigate the effect of ncSNVs on *MYBPC3* expression, we performed a lentivirus-based massively parallel reporter assay (LentiMPRA) using a library containing all possible single nucleotide variants in the *MYBPC3* promoter. The library was transduced into induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs). RNA and DNA were extracted, and unique 5' UTR barcodes were counted through high-throughput sequencing of RNA and DNA as a readout of transcription normalized to genomic integrations of each variant. This approach captured 1,173 variants (over 245 base pairs of sequence space) represented by 64,210 unique barcodes (55±33 barcode coverage per variant). A map of variant effect with single base-pair resolution was generated for the *MYBPC3* promoter, revealing that variants with large effects clustered at predicted TFBSs. ncSNVs within JASPAR-predicted TFBSs were more likely to decrease expression than those outside of predicted TFBSs (mean log-fold change: -0.29 vs -0.04, $p<0.0001$ on Mann Whitney test). We mapped ncSNVs from the MPRA to ncSNVs from the general population (gnomAD database), identifying 6 ncSNVs that had at least 20% reduction of *MYBPC3* expression that were also present in gnomAD. These findings suggest that ncSNVs have the potential to contribute to HCM through loss of function effects. Further mapping of the CREs of cardiomyopathy-associated genes is warranted to improve genetic diagnosis.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6048F: Leveraging bulk to interpret adipose single nucleus RNA-seq data: adipocyte subtypes reveal adipocyte genes for insulin sensitivity

Authors:

Z. Wang[1,2], S. T. Lee[1], A. Kar[1,2], M. Alvarez[1], H. M. Stringham[3], K. Gelev[1], M. Laakso[4,5], L. Scott[3], K. Mohlke[6], P. Pajukanta[1,2,7]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [3]Dept. of Biostatistics and Ctr. for Statistical Genetics, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [4]Inst. of Clinical Med., Internal Med., Univ. of Eastern Finland, Kuopio, Finland, [5]Dept. of Med., Kuopio Univ. Hosp., Kuopio, Finland, [6]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [7]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract:

Cellular insulin sensitivity (IS) protects against obesity and type 2 diabetes (T2D). However, genes linked to IS in subcutaneous adipose tissue (SAT) cell-types remain elusive. We hypothesize that SAT bulk RNA-seq data can be leveraged to identify SAT cell-type level IS genes in single nucleus RNA-sequencing (snRNA-seq) data. We performed snRNA-seq, followed by a multi-step quality control pipeline, in SAT biopsies from 120 Finnish METSIM men with detailed IS-related phenotypes (mean age=55.4 ±4.8, BMI=26.5±3.7, Matsuda index (MI)=7.98±3.8). To identify IS-associated subpopulations of cells across the obtained 14 SAT cell-types, we correlated non-overlapping METSIM bulk SAT RNA-seq data (n=335) with cell-level gene expression from the snRNA-seq data using Scissor. Comparing to insulin resistant (IR) cells (comprising ~60% immune cells), we found that the IS cells (~90% adipocytes) are depleted for expression across T2D-associated GWAS genes using scDRS (p=0.002). The IS adipocyte proportions also show significant correlations (adjp<0.05) in metabolically healthy directions with abdominal obesity (ρ=-0.33), BMI (ρ=-0.41), fat mass (ρ=-0.47), and MI (ρ=0.38). We observed 331 differentially expressed genes (DEGs; logFC>0, adjp<0.05) in IS adipocytes (e.g. *ADIPOQ*, *IRS1*) compared to IR adipocytes, enriched for pathways in fatty acid oxidation, lipid metabolism, glucose transport, and insulin response (FDR<0.05), while 584 IR adipocyte DEGs (e.g. *APOE*, *CD74*) are enriched for cell motility and adhesion. Building cell-type level co-expression networks using hdWGCNA identified 5 adipocyte networks. The eigengene of one of these networks is upregulated in the healthy direction across all four IS phenotypes using Wilcoxon tests, and genes in this IS network are highly expressed in IS adipocytes and

functionally enriched for healthy SAT pathways (FDR<0.05). To assess the genetic risk of the 166 adipocyte DEGs with a high module membership in the IS network, we used their *cis* regional SNPs to build partitioned network polygenic risk scores (PRSs) for T2D in METSIM and for BMI in the UK Biobank using PLINK. These network PRSs were significant predictors but not enriched for variance explained in T2D or BMI ($p_{perm1000}$>0.05 for both). Similarly, we saw no enrichment of T2D or BMI GWAS SNPs among the 166 genes using MAGMA (p>0.05 for both). In independent METSIM SAT snRNA-seq data (n=84), we replicated (adjp<0.05) the IS/IR subtypes, scDRS enrichment, IS trait correlations, and the IS adipocyte network, confirming that our results are robust. Overall, we discover a replicated adipocyte subtype and network of 166 IS genes related to SAT and cardiometabolic health.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6049F: Leveraging genome-wide effects on gene expression to identify new disease-critical genes with *trans*-genetic components

Authors:

K. Brunton[1], T. Amariuta[2,3]; [1]Dept. of Bioengineering, Univ. of California, San Diego, San Diego, CA, [2]Div. of BioMed. Informatics, Dept. of Med., UC San Diego, San Diego, CA, [3]Halıcıoğlu Data Sci. Inst., UC San Diego, San Diego, CA

Abstract:

A transcriptome-wide association study (TWAS) identifies genes that may mediate genetic associations found in GWAS. Since ~70% of genetic regulation of gene expression occurs distally, conventional TWAS models considering a gene's *cis* genetic variants are likely underpowered to detect disease-critical genes. The critical role of distal gene regulation in disease is highlighted by the fact that > 10% of genome-wide significant variants do not reside in the *cis* region *of any* gene. We hypothesized that modeling the *trans*-genetic component of gene expression in TWAS would implicate new disease genes.
Here, we augment *cis*-genetic gene expression prediction models with 3 methods for detecting *trans*-eQTLs (expression quantitative trait loci). First, using GBAT (gene-based association test), we identify cis-genetic variants of upstream genes that may regulate downstream target genes. Second, using transPCO, we identify variants associated with co-expression patterns of gene sets. Third, we use BAtch Screening Iterative Lasso (BASIL) to detect variants that are predictive of gene expression across the genome. We estimate priors on each of these approaches using xtune and subsequently learn posterior predictive effects on gene expression. Our simulations demonstrate a 40% increase in the power to detect disease-

associated genes in TWAS at realistic parameter settings.
We show the utility of our approach by constructing genome-wide gene expression prediction models for blood (N=574) and muscle-skeletal (N=602) tissue gene expression data. For 1,338 genes, our approach doubled the accuracy of *cis*-genetic gene models (average model R^2 increase from 0.046 to 0.109), 56% of which had an accuracy that was not significantly greater than 0 with standard *cis*-genetic models. Although *trans*-eQTLs are thought to be highly cell-type-specific, amongst genes with substantial increases in model accuracy (increase in R^2 > 0.05), we observed 15% are shared between tissues and about half of the predictive *trans* loci for these genes are shared between tissues. Intriguingly, our models implicated gene regulatory roles for >2,000 GWAS loci that were >500 kb from the nearest gene. We implemented a version of TWAS that includes *trans*-components of gene expression and identified 3,122 new genes associated with 8 immune-mediated GWAS traits. Notably, we associated inflammatory bowel disease (IBD) with IL-10*,* an established IBD therapeutic, whose *cis*-genetic model association was non-significant (p = 0.3). Overall, modeling the distal genetic component of gene expression greatly boosts the detection of disease-critical genes while improving characterization of GWAS variants.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6050F: Leveraging long-read Oxford Nanopore Technologies and a national genome program to better understand structural variation inheritance and medical impact

Authors:

D. Matias Sanchez[1], A. Al Awadhi[1], A. El-Khani[1], A. Al Mannaei[2], F. Al Marzooqi[1], F. Aldhuhoori[1], F. Sedlazeck[3], H. Wu[1], H. Sajad[1], I. Eltantawy[4], J. Quilez[1], J. Mafofo[1], L. Paulin[3], M. Ibrahim[4], M. Al Ameri[2], S. Elavalli[1], S. Zhang[1], T. Magalhaes[1], W. Abdulrahman[2]; [1]M42, Abu Dhabi, United Arab Emirates, [2]Dept. of Hlth.of Abu Dhabi, Abu Dhabi, United Arab Emirates, [3]Baylor Coll. Med., Houston, TX, [4]M42; IROS, Abu Dhabi, United Arab Emirates

Abstract:

The Emirati Genome Program (EGP) aims to perform ~30X whole-genome sequencing (WGS) of the approximately 1 million Emirati nationals in the United Arab Emirates. Its scale makes the EGP one of the largest population genome sequencing programs in the world. Unique to the EGP are (i) the random sampling of virtually the entire national population; (ii) the distinct genetic structure and extensive families resulting from the demographic history and mating practices in the Gulf; and (iii) the use of long-read Oxford Nanopore Technologies (ONT) as one

of the three high-throughput sequencing (HTS) technologies (more than 100,000 ONT WGS to date) - ONT is arguably superior to short-read sequencing (SRS) technologies to characterize structural variants (SV). These features combined allows us to conduct an unprecedented study of structural variation in humans.
Here we analyzed an initial dataset of 8,547 EGP samples with WGS from ONT. First, we demonstrated that WGS data derived from ONT is as high quality as Illumina and explored differences in SV calling between the two - for that we sequenced in-house 24 Coriell reference samples with WGS on both platforms. We found that ONT detected ~2X more SV per sample compared to Illumina, with that increase largely contributed by deletions and insertions.
Second, we produced the most comprehensive catalogue of SV in an Arab population. We identified a total of ~4 million SV in this cohort averaging 25,517(SD=1,808) SV per individual.
Finally, we investigated how SV are inherited in families. For that we initially estimated pairwise relationships and reconstructed families from the genomics data. We inferred 12,345 relationships resulting in 41 unique trios and 656 families. In those we studied the inheritance of SV spanning 5,027 medically relevant genes since understanding their inheritance can explain disease phenotypes and support diagnosis. We honed in on 386 cases that are proven to be challenging to accurately genotype by SRS to assess the added value of ONT in SV calling.
This study provides the most comprehensive and detailed characterization of SVs through long-read sequencing to date. It offers unprecedented insights that promise to revolutionize health outcomes and expand our understanding of genetic complexities, setting the stage for advanced genetic analysis and the development of personalized medicine in the United Arab Emirates and beyond.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6051F: Leveraging population cohorts to identity new pathogenic variants

Authors:

C. Moreau[1], É. Michel[1], L. Gagnon[2], J. Mathieu[3], L. Bouchard[4], M. Desmeules[4], S. Girard[1]; [1]Université du Québec à Chicoutimi, Chicoutimi, QC, Canada, [2]Univ. of Quebec in Chicoutimi, Chicoutimi, QC, Canada, [3]CIUSSS du Saguenay-Lac-St-Jean, Chicoutimi, QC, Canada, [4]Université de Sherbrooke, Saguenay, QC, Canada

Abstract:

Rare diseases affect 10% of the population and diagnosis remains a significant challenge for patients living with a rare disease who often experience prolonged diagnostic odyssey. In certain

populations, the prevalence of some rare diseases may increase due to demographic events such as a founder effect. It is the case in the Quebec province of Canada, predominantly settled by people of French origin by the early 1600s. The initial European founder effect was followed by regional founder effects, notably the one observed in the Saguenay-Lac-Saint-Jean (SLSJ) region. Consequently, many rare diseases are more frequent in Quebec and the SLSJ than elsewhere in the world. This study aims to identify new potentially pathogenic variants that have increased in frequency following the founder effects. Instead of looking at the clinical signs among patients, we aim at finding ClinVar established pathogenic variants in a population cohort that are at a higher frequency in Quebec and the SLSJ individuals compared to gnomAD non-Finnish Europeans. To confirm that a variant was introduced by one or few ancestors and has increased in frequency due to the founder effect, we ensured that there was a shared IBD genomic segment surrounding this variant in at least half of the carriers. Applying these criteria to 25,041 individuals with imputed data (including 1,879 who also had WGS) enabled us not only to validate 101 previously reported founder variants, but crucially to identify 87 founder variants that had never been reported with carrier rates ranging from 1/25 to 1/1000 in the SLSJ. Importantly, we identified individuals who were homozygous for 2 of these variants and had been diagnosed with the corresponding diseases in our data. We also compared the number of carriers in the SLSJ to those in other urban regions of Quebec (UQc) for the previously reported and newly identified founder variants. We found that 55% of individuals in the SLSJ and 30% in UQc carry at least one such variant. Despite this, 69% of the variants found in the Quebec province were not considered as founders in the SLSJ, indicating that while there are fewer variants present, a higher proportion of individuals carry at least 1 variant in the SLSJ compared to UQc. These results show that leveraging population cohorts to identify new pathogenic variants more frequent within specific populations could be a promising approach toward enhancing and fastening the molecular diagnosis. This could shorten the patients' diagnostic odyssey and reduce the economic burden associated with the undiagnosed rare diseases.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6052F: Linking regulatory variants to target genes by integrating single-cell multiome methods and genomic distance

Authors:

E. Dorans[1,2], K. A. Jagadeesh[1,3], K. K. Dey[4], A. L. Price[1,3,5]; [1]Dept. of Epidemiology, Harvard Sch. of Publ. Hlth., Boston, MA, [2]PhD Program in Biological and BioMed. Sci., Harvard Med. Sch., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard,

Cambridge, MA, [4]Sloan Kettering Inst., Mem. Sloan Kettering Cancer Ctr., New York, NY, [5]Dept. of Biostatistics, Harvard Sch. of Publ. Hlth., Boston, MA

Abstract:

Methods that analyze single-cell RNA-seq and ATAC-seq multiome data have shown great promise in linking disease GWAS variants to target genes by correlating chromatin accessibility with gene expression across cells (e.g. SCENT; Sakaue et al. 2024 Nat Genet). However, existing methods have low concordance in their linking scores and do not capture the effects of genomic distance.

We propose pgBoost, a gradient boosting approach that trains a non-linear combination of peak-gene linking scores (SCENT, Signac, ArchR, Cicero) computed separately across cell types and data sets, plus genomic distance, on fine-mapped eQTL data to assign a probabilistic score to each candidate SNP-gene link. As positive and negative training sets, we use fine-mapped eSNP-eGene pairs with maximum PIP > 0.5 and maximum PIP < 0.01 across GTEx tissues, restricted to off-chromosome data.

We applied pgBoost to single-cell multiome data from 85k cells across 6 major immune/blood cell types. pgBoost outperformed other single-cell methods and genomic distance in enrichment across 4 independent SNP-gene link evaluation sets. For fine-mapped eSNP-eGene pairs, pgBoost attained enrichment of 21x at distance >10kb (vs. 1.2-3.1x for other single-cell methods; each p for difference < 4e-35). For CRISPR-validated links, pgBoost attained enrichment of 4.8x at distance >10kb (vs. 1.6-2.1x for other single-cell methods; each p for difference < 2e-5). We observed even larger improvements at larger distances (e.g. >100kb). Similar improvements in enrichment were observed for evaluation links derived from GWAS data and Activity-By-Contact (ABC) scores. Restricting pgBoost to features from a focal cell type further improved the identification of SNP-gene links in relevant cell types, e.g. pgBoost trained using erythrocyte features attained enrichment of 13x for ABC links in erythrocytes (vs. 3.6-4.7x using features in other unique cell types or all cell types; each p for difference < 0.007).

We highlight two examples. pgBoost linked rs1427407 (fine-mapped for RBC count, PIP=0.99) to *BCL11A* (experimentally validated link; top 5% of links for pgBoost and SCENT but not for Signac, ArchR, Cicero, genomic distance). pgBoost also linked rs2295348 (fine-mapped for mean corpuscular volume, PIP=0.58) to *RNF24* (210kb to TSS, 11th closest TSS; top 5% of links for pgBoost but not for SCENT, Signac, ArchR, Cicero, genomic distance). Notably, erythroid phenotypes are impacted by knockout mouse models for *RNF24* (but no other genes within 500kb). We conclude that a non-linear combination of linking strategies, including genomic distance, improves power to identify target genes underlying GWAS associations.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6053F: Monocyte Inflammation and Resilience to Alzheimer's disease: Novel Genetic Risk Genes Identified by Transcriptome-wide Association Study

Authors:

Y. Mustafa[1], V. A. Janve[2], T. Hohman[3], J. Haines[4], L. Main[5], Y. Song[6], M. Pericak-Vance[7], W. Scott[8], A. Naj[9], J. Below[3], L. Dumitrescu[10], W. Bush[4]; [1]Case Western Reserve Univ., Westlake, OH, [2]Vanderbilt Univ. Med. Ctr., Nashvillle,, TN, [3]Vanderbilt Univ Med Ctr., Nashville, TN, [4]Case Western Reserve Univ, Cleveland, OH, [5]Case Western Reserve Univ., Lakewood, OH, [6]Case Western Reserve Univ., Cleveland, OH, [7]Univ. of Miami Miller Sch. of Med, Miami, FL, [8]Univ. of Miami, Miami, FL, [9]Univ. of Pennsylvania, Philadelphia, PA, [10]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Alzheimer's disease (AD) is a highly heritable, progressive, and increasingly common neurodegenerative disease of aging. Recent GWAS have attempted to discover genetic elements that are associated with resilience to AD, which can be defined as better-than-predicted cognitive performance relative to measured neuropathology or genetic risk. We performed a series of transcriptome-wide association studies (TWAS) combining monocyte genetic and expression data with GWAS summary statistics of cognitive and global AD-resilience (residuals of CSF biomarkers Aβ42 and total tau regressed on cognitive/memory performance and/or hippocampal volume) and Amish cognitive preservation (absence of cognitive impairment despite advanced age or APOE4 status) in order to discover genetic risk genes of AD in monocytes, an understudied axis of innate immune processes and more accessible target for intervention in AD. Training data of monocyte expression in several states of immune stimulation (naive, LPS 2h, LPS 24h, IFN 24h), as well as corresponding genotype data from 432 individuals (Fairfax et al. 2014) were accessed and subject to QC. We developed models of cis-genetic expression using two complementary strategies: elastic-net and LD-pruning/MASH; the former is adept at capturing polygenic structures, and the latter independent loci significant across all states of stimulation. Elastic-net resulted in an average of 4,497/18,184 models that explained more than 1% of gene expression (avg. 3%, range 1-40%), and 177/422 using LD-pruning/MASH (avg. 5%, range 1-50%). Genes significantly modeled were tested for association with resilience phenotypes using S-PrediXcan. 33 genes met a suggestive threshold of $P < 0.0008$. Whole-blood RNA-seq data from a separate AD cohort was used for validation, with 6 of our putative targets (*COPS8, CRISPLD2, MOB3A, NUDT13, SQOR, UPF3A*) showing differential expression between AD cases and controls. These genes are involved in Hippo signaling, cytokine secretion, and RNA metabolism pathways. Expression and eQTL profiles across cell

types support that these associations are likely specific to monocytes. These results confirm a distinct axis of genetic risk in myeloid inflammatory processes contributing to AD and can represent novel therapeutic targets.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6054F: Natural variants in SEL1L modify lethality, ERAD, and proteasome function in a model of NGLY1 deficiency

Authors:

T. Tu'ifua[1], C. Chow[2]; [1]Univ. of Utah, Salt Lake city, UT, [2]Univ. of Utah, Salt Lake City, UT

Abstract:

N-glycanase 1 (NGLY1) deficiency is a rare disease caused by autosomal recessive loss of function mutations in the *NGLY1* gene. Patients suffer from movement disorder, developmental delay, liver dysfunction, and alacrima. NGLY1 removes N-linked glycans from glycoproteins in the cytoplasm and is thought to help clear misfolded glycoproteins from the endoplasmic reticulum (ER) through the ER associated degradation (ERAD) pathway. Despite this, the physiological significance of NGLY1 in ERAD is not understood. The best characterized substrate of NGLY1 is NRF1, a transcription factor that upregulates proteasome expression and the proteasome bounce back response.

We performed a genetic modifier screen using a *Drosophila* model of NGLY1 deficiency and identified a number of modifiers that reduce the lethality of the model. We identified two protein coding variants in *SEL1L*: S780P and Δ806-809. Both variants are localized in the SEL1L cytoplasmic tail, an uncharacterized domain of the protein. SEL1L is a component of the ERAD complex that retrotranslocates misfolded proteins from the ER to the cytoplasm for degradation.

To test the interaction between these *SEL1L* variants and *NGLY1*, we created CRISPR mutant fly lines that carry these *SEL1L* variants in a common genetic background and tested them with our model of NGLY1 deficiency. Validating our screen, the $SEL1L^{P780}$ and $SEL1L^{\Delta 806\text{-}809}$ variants increase the survival of the NGLY1 deficiency model, compared to the $SEL1L^{S780}$ variant. To determine how these *SEL1L* variants were modifying lethality in NGLY1 deficiency, we interrogated the ERAD and NRF1 signaling pathways.

We found that the $SEL1L^{P780}$ and $SEL1L^{\Delta 806\text{-}809}$ variants improve general ERAD function. However, this improved ERAD is lost when we reduce NGLY1. This result further implicates NGLY1 in general ERAD function, however it does not explain how the *SEL1L* variants decrease NGLY1 deficiency lethality. Interestingly, we found that these variants are protective against

proteasome inhibition, leading to changes in larval size and increased survival, in heterozygous *NGLY1* null flies. We hypothesize that these *SEL1L* variants modify NGLY1 deficiency through NRF1 signaling. These results will provide new insights into the role of NGLY1 in ERAD and the etiology of NGLY1 deficiency. *SEL1L* is a strong candidate modifier gene in patients, where variability in presentation is common.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6055F: New multi-omics xQTL colocalization method improves the discovery of causal variants for Alzheimer's disease

Authors:

K. Dey[1], X. Cao[2], H. Sun[3], R. Feng[2], R. Mazumder[4], The Alzheimer's Disease Functional Genomics Consortium, G. Wang[3]; [1]Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Columbia Univ., Fort Lee, NJ, [3]Columbia Univ., New York, NY, [4]Massachusetts Instiitute of Technology, Boston, MA

Abstract:

The expansive nature of molecular quantitative trait loci (xQTL) studies, spanning multiple phenotypes (expression, splicing, protein abundance) and multiple cell types, necessitates a scalable colocalization framework that can harness the power of these multiple modalities to identify functionally causal variants underlying diseases. Current colocalization methods are inadequate in scaling to many phenotypes (COLOC), or in accounting for multiple causal variants at a locus (HyPrCOLOC). To this end, we propose a multi-phenotype colocalization method, *ColocBoost*, that employs a novel proximity-adapted gradient boosting algorithm to simultaneously identify multiple colocalization events and the phenotypes they are shared in from individual-level and summary statistics data. We performed systematic benchmarking of *ColocBoost* against COLOC (V5), HyProColoc and MOLOC for simulation settings with varying number of causal variants (1 to 3) and phenotypes (2 to 50). *ColocBoost* exhibited 2.34x-4.23x significant higher power than HyPrColoc and MOLOC at >1 causal variants (p-value<0.001) while also showing 21.1%-75.2% reduced FDR. We applied *ColocBoost* to 10 different xQTL data from the brain cortex of ROSMAP subjects (N=424); this included 6 pseudobulk cortical cell types and 4 bulk eQTLs from different cortical regions and monocytes. We identified colocalization events for 10,793 genes, in which 3,882 genes shared >1 causal variants. 63.5% of colocalized sets of variants were shared across more than 2 phenotypes; with Excitatory and Inhibitory neurons sharing 1.9x-12x more colocalization events compared to other brain cell-type pairs. For genes with multiple colocalization events, weaker secondary colocalization events were more gene-distal and 27.5% more cell-type specific than the primary strongest colocalization events.

ColocBoost identified 2x more AD colocalizations in ROSMAP compared to GTEx brain. Variants colocalized across multiple brain cell types by *ColocBoost* showed 39% and 48% higher disease heritability information for immune and brain traits meta-analysis compared to the fine-mapped variants. Next, we performed a targeted colocalization of the xQTLs with Alzheimers' (AD) disease GWAS. *ColocBoost* identified 31% of AD colocalization events to be specific to microglia, including for some well-known genes like *BIN1* and *INPP5D*. For the *PICALM* gene, we observed 2 distinct AD colocalization events: one specific to microglia and the other shared between microglia and astrocytes. Overall, *ColocBoost* provides a novel framework to identify colocalized disease-critical functional signals for varying numbers of phenotypes.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6056F: Non-coding variants in *SCN1A* reveal novel poison exons and therapeutic targets

Authors:

S. Tang[1], G. Carvill[2]; [1]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [2]Northwestern Univ., Chicago, IL

Abstract:

Dravet syndrome is a developmental and epileptic encephalopathy caused by pathogenic variants in *SCN1A,* a gene encoding for a subunit of the voltage-gated sodium channel. Previous work demonstrated the existence of a poison exon in intron 20 of *SCN1A*. Aberrant inclusion of this poison exon results in nonsense-mediated decay of the *SCN1A* transcript, reducing protein abundance and leading to disease phenotypes by haploinsufficiency. This poison exon has been leveraged as a therapeutic target for antisense oligonucleotides that modulate *SCN1A* splicing and rescue protein abundance. Here, we use targeted long-read sequencing of human iPSC-derived neurons to validate two novel poison exons in *SCN1A*. We show that disease variants within these poison exons lead to their aberrant inclusion in an *in vitro* splice reporter assay, as well as in patient iPSC-derived neurons. Additionally, we use antisense oligonucleotides (ASOs) to target aberrant *SCN1A* splicing. Finally, we demonstrate that in human fetal and adult brain, *SCN1A* poison exon inclusion is developmentally regulated. Taken together, our results reveal novel pathogenic mechanisms, identify new therapeutic targets, and provide important preclinical data for genetic therapies in *SCN1A*-related disorders.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6057F: *NSD2* New variant leading to Global Developmental Delay and Autism

Authors:

S. Carlo[1,2], N. Arciniegas[1], A. Cornier[3], G. L. Lacourt Sosa[1], D. Verges Santiago[1], D. P. Torres Rodriguez[1]; [1]Ponce Hlth.Sci. Univ., Ponce, PR, [2]Mayaguez Med. Ctr., Mayaguez, PR, [3]San Jorge Children s & Women s Hosp., San Juan, PR

Abstract:

Rauch-Steindl syndrome (RAUST) is a rare genetic disorder that affects multiple body systems and is characterized by a wide range of symptoms that include dysmorphic facial features, developmental delay, small head circumference, hypotonia, and speech delay, among others. It is known to manifest as a less severe form of Wolf-Hirschhorn syndrome that is present in less than one percent of the world's population. The rarity of RAUST imposes a clinical diagnostic and treatment challenge.
A 2-year-old female was brought for a consult due to speech delay. Upon evaluation, a diagnosis of Global Developmental Delay and Autism Spectrum Disorder was made. Because her family history was non-contributory, a chromosomal microarray was performed in hopes of pinpointing the specific cause of the diagnosis. Results for this test were unremarkable, and as a follow up, a whole exome sequencing was ordered. This test revealed a missense mutation, c.3040C>G (p.Pro1014Ala) on *4p16.3*, specifically on gene *NSD2*. With an autosomal dominant heterozygous inheritance pattern, this genetic mutation is known as a variant of uncertain significance and is present in the general population with a frequency 0.000006573. The gene affected is responsible for providing instructions for the assembly of proteins MMSET2 and RE-IIBP, which function as histone methyltransferases. Tasked with epigenetic regulation of genetic expression, defects in these proteins are thought to affect normal development, even though the exact cascade of genes affected is still under investigation. The *NSD2* mutation, together with the associated phenotype, are suggestive of Rauch-Stiendl syndrome.The clinical presentation of this case, as well the limited amount of registered disease incidence, demonstrates the necessity of more ongoing research related to this rare genetic disorder.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6058F: pcQTLs: Identifying variants impacting the expression of multiple genes

Authors:

K. Lawrence, T. Gjorgjieva, S. Montgomery; Stanford Univ., Stanford, CA

Abstract:

Genome wide association studies (GWAS) have linked thousands of variants across the human genome to complex traits and diseases. However, because GWAS are agnostic to the mechanism by which a variant impacts a phenotype and because a majority of GWAS variants (~90%) fall into noncoding regions of the genome, interpreting the functional impact of these variants remains challenging. About one third of GWAS variants can be linked to some intermediate molecular phenotype, such as gene expression or splicing usage via quantitative trait loci (QTL) mapping. Underpinning QTL approaches is the assumption that each variant exerts its effect via one, causal gene. Here, we describe a framework to map QTLs impacting the expression of multiple nearby genes simultaneously. We apply this framework using RNA-seq data from 13 tissues collected as part of the Genotype-Tissue Expression (GTEx) project. We identify correlated gene clusters and use local principal components (PCs) of gene expression as the dependent variable in QTL analysis to map pcQTLs. Using this approach, we are able to capture signals of shared regulatory architecture and reduce multiple testing burden to detect dispersed, multi-gene effects. In summary, pcQTLs extend the traditional framework of single causal genes acting independently to one that identifies causal loci where multiple genes impact complex traits.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6059F: Pervasive gene overlap across pathways leads to false positives in GWAS pathway enrichment testing

Authors:

A. Cote, L. Liou, J. Garcia Gonzalez, P. O'Reilly; Icahn Sch. of Med., New York, NY

Abstract:

In a genome-wide association study (GWAS), genetic variants are tested for association with risk for a disease. One of the many downstream analyses to interpret GWAS is gene set enrichment testing, where researchers test whether genes harboring risk variants are more overrepresented in a pathway than expected by chance. To perform enrichment analyses researchers typically use curated pathway databases based on experimental evidence. These databases include Gene Ontology, Reactome, KEGG, and DrugBank, among others. While informative, these resources demonstrate an unbalanced representation of genes, where some genes are annotated to many pathways. For example, in the Gene Ontology and Reactome databases, most genes belong to

at least 3 pathways, and 1% of genes are found in a striking 28 or more pathways. This overlap of genes across pathways can be due to redundant or highly similar pathways, gene multifunctionality, or the repeated study of key disease-relevant genes. As a result, gene set enrichment results can often be biased towards well-studied genes. While this issue of gene overlap has been noted previously, here we systematically examine it in the context of GWAS pathway enrichment. GWAS enrichment methods often account for factors such as gene size, gene set size, or gene density, but these approaches do not explicitly account for the frequency of genes across pathways. In this study we assess the impact of gene overlap on the robustness and specificity of GWAS gene set enrichment results across complex disorders. We first quantify the overlap between gene sets across major pathway databases, co-expression modules, and other functional pathway resources. Next, we measure GWAS gene set enrichment with and without the inclusion of influential, highly annotated genes. We assess the impact of gene overlap on both accuracy and reproducibility of gene set enrichment results across different GWAS and across enrichment methods. Lastly, we test whether accounting for gene overlap improves disease prediction in pathway-based polygenic risk score analysis. We hope this work may serve as a resource and provide important context for the interpretation of gene set enrichment results in future GWAS follow-up.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6060F: Polygenic disease association of individual cells in single-cell ATAC-seq data

Authors:

B. Truong[1], M. Zhang[2], A. Price[3]; [1]Harvard Univ. / Broad Inst., Cambridge, MA, [2]Ray and Stephanie Lane Computational Biology Dept., Sch. of Computer Sci., Carnegie Mellon Univ., Pittsburgh, PA, [3]Harvard Sch. of Publ. Hlth., Boston, MA

Abstract:

Associating individual cells to disease by integrating single-cell ATAC-seq with GWAS can deepen our understanding of disease etiology, complementing insights from single-cell RNA-seq. Previous individual cell-level methods restrict to fine-mapped GWAS loci, limiting power. Here, we introduce PolyADS, a powerful method that leverages genome-wide polygenic signals to associate individual cells in scATAC-seq data to disease.
PolyADS evaluates whether an individual cell exhibits excess accessibility across a set of disease-associated peaks, accounting for dependency between nearby peaks to ensure accurate calibration. First, we apply MAGMA to GWAS summary statistics to identify the top 100,000

disease-associated peaks (disease peaks). Second, we define 1,000 sets of control peaks by matching GC content and average accessibility in clusters of disease peaks. Third, we score each cell for higher accessibility in disease peaks, comparing the resulting disease score to all control scores for all other cells to assess statistical significance. We also score cell types by comparing disease scores vs. control scores of constituent cells. We applied PolyADS to scATAC-seq profiles of 1.2 million cells (Zhang et al. 2021 Cell) in conjunction with GWAS summary statistics for 66 diseases/traits (average *N*=339K). First, in empirical null data, PolyADS produced far fewer unexpected associations at FDR<10% compared to a baseline of applying chromVAR to disease peaks (84 vs. 814 for 178K brain cells-8 blood traits and 2 vs. 189 for 527K blood cells-10 brain traits), implying superior calibration of PolyADS. Second, at the cell type level, PolyADS identified well-established associations that were not identified by chromVAR (e.g., erythroblast-red blood cell count, astrocyte-educational attainment; AUC = 0.83 for PolyADS vs. 0.79 for chromVAR with respect to a curated set of cell type-trait associations). PolyADS also identified biologically plausible associations supported by previous S-LDSC analyses that were not identified in previous analyses of scRNA-seq data (scDRS method; Zhang et al. 2022 Nat Genet) (e.g., fetal macrophage-HbA1c, adult natural killer T cells-total bilirubin). Third, at the individual cell level, PolyADS identified significant heterogeneity within many cell types (e.g., glutamatergic neuron-smoking status, T lymphocyte-inflammatory bowel disease; 45% of cell type-trait associations had significant cell-level heterogeneity via Geary's C statistic). In conclusion, PolyADS is a well-calibrated and powerful method for distinguishing disease-associated individual cells and cell types in scATAC-seq data.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6061F: Population level-diversity of human IGHC haplotypes is linked to inter-individual variation in antibody isotype levels in the expressed repertoire

Authors:

U. Jana[1], E. Ford[1], O. Rodriguez[1], W. Lees[1], E. Engelbrecht[1], Z. Vanwinkle[1], K. Shields[1], S. Schultze[1], D. Smith[1], G. Yaari[2], M. Smith[1], C. T. Watson[1]; [1]Univ. of Louisville, Louisville, KY, [2]Bar-Ilan Univ., Ramat Gan, Israel

Abstract:

The Immunoglobulin heavy chain constant (IGHC) domain in antibodies (Abs) and B cell receptors (BCRs) is crucial for immune function, facilitating antigen (Ag) neutralization and elimination via various effector mechanisms. Despite known effects of IGHC genetic variants on

Ab/BCR function, the extent of population-level diversity in the IGHC locus remains unexplored, especially in understudied genetic ancestries. As a result, we lack a comprehensive understanding of the influence of IGHC genetic variants on the transcriptional regulation of these genes, and the downstream impact on circulating isotype levels and differential Ab effector functions in disease.
To address these limitations, we employed targeted long-read sequencing (Pacific Biosciences) of the IGHC locus to generate locus-wide locally phased assemblies in 110 donors from 5 diverse super populations. These represent the richest catalog to date of IGHC SNVs and SVs, including a small inversion, insertions and deletions ranging in size from 20-120 Kb, leading to the discovery of novel IGHC genes and alleles. In total, we discovered 5 novel genes, and 211 currently undocumented alleles, representing the majority (88%) of observed alleles. Alleles were enriched for non-synonymous variants, including positions known to influence Ab function. In addition, we observed differences in SV and gene allele frequencies between populations. To link IGHC variants to functional signatures in the expressed Ab repertoire, we integrated long-read genomic sequencing with full-length repertoire sequencing (FLAIRR-seq) of IgM, IgG, and IgA transcripts from peripheral blood mononuclear cells (PBMCs) in a cohort of healthy adult donors (n=50). This allowed us to comprehensively profile IG transcripts, including expressed IG variable (V), diversity (D) and joining (J) genes, linked to nearly-full-length IGHC alleles of the same rearranged chromosome. This facilitated complete descriptions of expressed Ab/BCR transcripts at the isotype, subisotype, and IGHC allele levels. We used these data here for the first time, to identify IGHC SNVs and SVs associated with IGHC gene and allele usage patterns, Ab/BCR Ag-binding domain diversity, and clonal lineage structure in the circulating repertoire. Together, these data provide a novel framework for integrating long-read genomic and Ab/BCR transcriptome sequencing to dissect intricate variation in the Ab expression. Future applications of this approach will lead to a more precise understanding of genotype-phenotype relationships and provide valuable insights into Ab-mediated mechanisms underlying disease and potential therapeutic interventions.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6062F: Population-scale Single-nucleus Cell-type-specific Chromatin Dynamics and Nucleosome Positioning in Skeletal Muscle ⭐

Authors:

X. Wang[1], C. Robertson[1,2], A. Varshney[3], N. Manickam[4], P. Orchard[1], M. Laakso[5,6], J. Tuomilehto[7], T. Lakka[5], K. Mohlke[8], M. Boehnke[1], L. Scott[1], H. Koistinen[7], F. Collins[2], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]NHGRI, NIH, Bethesda, MD, [3]Univ. of Michigan, Canonsburg,

PA, [45]Univ of Eastern Finland, Kuopio, Finland, [6]Kuopio Univ Hosp., Kuopio, Finland, [7]Finnish Inst for Hlth.and Welfare, Helsinki, Finland, [8]Univ North Carolina, Chapel Hill, NC

Abstract:

Nucleosomes are the basic repeating units of chromatin, and nucleosome positioning relative to gene regulatory processes, such as transcription factor (TF) binding, plays a key role in chromatin dynamics and cell-type-specific gene regulation. Existing chromatin accessibility (ca) quantitative trait locus (QTL) studies do not partition the ca-regions into nucleosome-free regions (NFR) and positioned nucleosomes. As a corollary, no QTL studies have investigated the effects of genetic variation on nucleosome positioning in type 2 diabetes (T2D)-related tissues. To address these gaps, we mapped NFR and nucleosome occupancy in skeletal muscle using single nucleus chromatin accessibility (snATAC) profiles from 281 individuals in the FUSION tissue study. We focused on the 5 most abundant cell types: Type 1 fibers, Type 2a, Type 2x, Endothelial, and Fibro–adipogenic Progenitors. Using Multi-Otsu statistical thresholding, we partitioned snATAC fragments into NFR and nucleosomal components. We called nucleosome-free peaks (nfrPeaks) and nucleosome occupancy peaks (nucPeaks) from the resulting profiles and quantified NFR accessibility at nfrPeaks and nucleosomal occupancy at nucPeaks. Finally, we mapped genetic effects on NFR accessibility (nfrQTLs) and nucleosomal occupancy (nucQTLs). We discovered 76,027 nfrQTLs and 10,583 nucQTLs across the 5 cell types (FDR < 5%). Analyzing GWAS summary statistics for 384 traits through LD score regression, we found nfrQTL peaks enriched in traits including T2D, fasting insulin, and childhood BMI ($p < .05$). Formal colocalization analysis identified 115 nfrQTLs that colocalized with T2D GWAS signals (posterior probability, PPH4 > 0.5). Notably, nfrQTLs near genes including *C2CD4A, AP3S2,* and *IRS3* were observed among all three muscle fiber cell types. Colocalization analysis also revealed 6,396 pairs of colocalized nfrQTL and nucQTL (PPH4 > 0.8). Causal inference test analyses indicated that nfrQTLs causally influence nucQTLs for > 90% of colocalizing QTL pairs in Type 1, Type 2a, and Type 2x cells (FDR < 5%), supporting a genetic model whereby NFRs causally position adjacent nucleosomes.
These analyses provide candidate causal high-resolution chromatin mechanisms at metabolic disease loci where disease variants exert their regulatory effects by first modifying NFR accessibility, likely through TF binding, which then alters adjacent nucleosome positioning.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6063F: pQTL discovery using the SomaScan®high-plex proteomics platform in the COPDGene study

Authors:

B. Epstein, E. Troth, J. Chadwick, M. A. Hinterberg; SomaLogic, Inc, Boulder, CO

Abstract:

Protein Quantitative Trait Loci (pQTLs) link genetic variation to proteomic variation, facilitating the discovery of mechanisms underlying genetic disease risks. Here we present a pQTL study based on the SomaScan® Assay, a high-plex affinity-based proteomics assay that uses single-stranded DNA aptamers (SOMAmer® Reagents) to bind target proteins. We identified pQTLs in 3,243 non-Hispanic White (NHW) and 1,477 African American (AA) subjects from visit two of the COPDGene study assayed with the SomaScan Assay, V4.0, which enables approximately 5,000 protein measurements. COPDGene is a multi-center observational study designed to identify genetic factors that affect susceptibility to COPD among smokers. We tested for associations between protein measurements and ~ 6.5 million (NHW) or 13 million (AA) SNPs by using linear regression with 10 genetic principal components (PCs), 3 proteomic PCs, sex, age, and pack-years of tobacco used (included because tobacco use is an important factor in COPDGene) as covariates. Genomic inflation factors were calculated using LD (linkage disequilibrium) score regression. We identified significant ($p < 0.05$ after Bonferroni correction) index variants, clumped variants in linkage disequilibrium ($R^2 > 0.7$) with the index variants, and labeled the resulting clumps of variants as cis-pQTLs if the index variant was within 500 kb of the transcriptional start site. Protein-altering variants (PAVs) can affect affinity-based assays by changing the binding of reagents to the target protein, thus we searched for PAVs among clumped variants. There were 736 and 1,018 protein analytes with cis-pQTLs identified in the AA and NHW populations, respectively. Of those, 617 had cis-pQTLs in both populations and 553 had either the same index SNP, or the index SNPs in the two populations were in LD ($R^2 > 0.7$) with each other. Cis-pQTLs explained a median of 2.7% and 4.8% of the variance in measured protein abundance in NHW and AA, respectively. We found that 457 protein analytes had a cis-pQTL that included a PAV in the NHW population (376 in AA). When considering multiple independent ($R^2 < 0.7$) significant hits to the same analyte separately, 993 analytes in NHW (97%) and 704 in AA (96%) had a cis-pQTL hit that did not include a PAV. These results indicate that while genetics influences many protein measurements, the protein measurements are also sensitive to environmental factors, and different pQTLs may be discovered from populations with different ancestries. Genomics and proteomics data can provide a complementary and more comprehensive biological representation than either type of data alone.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6064F: Proteome-wide association study of blood mitochondrial DNA copy number

Authors:

G. Caltagirone[1,2,3], J. Feiner[1,2,3], N. Perrot[1,2,3], T. Pocuca[1,2,3], G. Pare[1,2,3], M. Chong[1,2,3]; [1]McMaster Univ., Hamilton, ON, Canada, [2]Population Hlth.Res. Inst., Hamilton, ON, Canada, [3]Thrombosis & Atherosclerosis Res. Inst., Hamilton, ON, Canada

Abstract:

Mitochondrial DNA copy number (mtDNA-CN) is a measure of the average intracellular concentration of mitochondrial genomes. Observational studies suggest associations between low mtDNA-CN and greater risk of various age-related diseases, including neurodegenerative disorders and cardiovascular disease. While previous studies posited that mtDNA-CN reflects the number of mitochondria in cells, the biological relevance remains unclear. Protein correlations may provide clarity into the biological pathways associated with variability in mtDNA-CN. Given that mitochondria are essential for many cellular functions, we hypothesized that mtDNA-CN levels are systematically associated with variation in the circulating proteome. We derived buffy coat mtDNA-CN estimates from whole-genome sequencing data in 40,643 UK Biobank participants. Associations between mtDNA-CN and 2923 circulating protein biomarkers were tested using linear regression models adjusted for age, sex, collection conditions, and blood cell composition. 1699 (58%) proteins were significantly associated (p<1.71x10-5). For 97% of significant proteins, higher intracellular mtDNA-CN levels were associated with lower circulating protein levels. This extreme proportion of proteins exhibiting an inverse relationship was specific to mtDNA-CN as similar analyses for body mass index resulted in only 23% of proteins being inversely associated. Pathway analysis revealed enrichment in various biological pathways including catabolic processes, stress response, organelle organization, apoptotic signalling, and programmed cell death (p<0.05). In parallel, enrichment was observed in several intracellular components, most notably the cytoplasm (p=3.31x10-19) and the mitochondrion (p=2.53x10-2). Despite proteins being measured extracellularly, enrichment was observed for intracellular proteins, suggesting that these proteins are released from the cells. In addition, no significant enrichment was observed in secreted proteins, suggesting that proteins are leaking out due to cell death or lysis instead of normal active secretion. Supporting this, a 3.67-fold enrichment for proteins related to the apoptotic pathway was observed among the top tertile of the most significant proteins (p=1.23x10-5). Altogether, the systemic downregulation of circulating proteins observed suggests that low mtDNA-CN may reflect cell death and turnover, which may explain previous observations of lower mtDNA-CN marking higher risk of several chronic diseases.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6065F: Rare and Common Variants Risk of Thoracic Aortic Aneurysm and Dissection

Authors:

D. Guo[1], J. DePaolo[2], D. R. Murdock[1], A. C. Cecchi[1], S. A. LeMaire[3], A. L. Estrera[1], J. S. Coselli[4], J. Smith[5], M. Bamshad[5], G. Biagetti[2], S. Abramowitz[2], R. Judy[2], M. G. Levin[2], S. M. Damrauer[2], D. M. Milewicz[1]; [1]UT McGovern Med. Sch. at Houston, Houston, TX, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Res. Inst. and Heart & Vascular Inst., Geisinger, Danville, PA, [4]Baylor Coll. of Med., Houston, TX, [5]Univ. of Washington, Seattle, WA

Abstract:

Normal 0 false false false EN-US ZH-CN X-NONE /* Style Definitions */ table.MsoNormalTable {mso-style-name:"Table Normal"; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:""; mso-padding-alt:0in 5.4pt 0in 5.4pt; mso-para-margin:0in; mso-pagination:widow-orphan; font-size:12.0pt; font-family:"Calibri",sans-serif; mso-ascii-font-family:Calibri; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Calibri; mso-hansi-theme-font:minor-latin; mso-bidi-font-family:"Times New Roman"; mso-bidi-theme-font:minor-bidi; mso-font-kerning:1.0pt; mso-ligatures:standardcontextual; mso-fareast-language:EN-US;} Up to 20% of aortic dissection (AD) cases have evidence of a pathogenic or likely pathogenic (P/LP) variant in 11 genes validated to cause heritable thoracic aortic disease (HTAD). To determine if these genes also contribute to early onset (≤60 years old) sporadic dissections, exome sequencing was performed on 551 patients with ESTAD cases. We determined that 9.3% of ESTAD patients have a P/LP variant in an HTAD genes, with *FBN1* having the highest burden (4.2%). Interestingly, cases presenting Stanford type B aortic dissections have a higher burden of P/LP variants than type A dissections. The burden of P/LP variants in HTAD genes was also assessed in patients with ADs and thoracic aneurysms (TAAs) in the Penn Medicine Biobank (PMBB, AD = 262, TAA =1377) and the United Kingdom Biobank (UKB, AD = 490, TAA =1134). The burden of P/LP variants in HTAD genes was higher in patients with AD (PMBB, 9.2% and UKB 4.9%) compared to those with TAA (PMBB, 2.0% and UKB 2.8%). Further analyses identified that individuals with a P/LP variant in an HTAD gene had a 20-fold increased risk of a diagnosis of AD or TAA (95% confidence interval [CI] 13.5 to 29.5, P<0.001). Using GWAS data for ascending thoracic aortic diameter (AscAoD), a polygenic risk score (PRS) was constructed that strongly associated with prevalent TAD in the PMBB data (OR=1.43 per 1 standard deviation increase in PRS, 95% CI 1.36 to 1.50, P<0.001). We found that TAD prevalence was increased among individuals carrying a P/LP HTAD gene

variant in the highest PRS quintile (58%, 95% CI 37% to 78%) compared to carriers in the lowest PRS quintile (24%, 9.4% to 45%). Thus, our data indicate: (1) the burden of P/LP in HTAD genes is greater in cases presenting with ADs than TAAs; (2) the presence of a P/LP variant confers a 20-fold increased risk of thoracic aortic disease in the population; (3) common variants reflected in a PRS modify the penetrance of the P/LP variants in HTAD genes.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6066F: Reducing Variant Classification Disparities Using Multiplexed Functional Data

Authors:

M. Dawood[1], S. Fayer[2], S. Pendyala[2], M. Post[2], D. Kalra[3], K. Patterson[2], E. Venner[4], L. Muffley[2], D. Fowler[5], A. Rubin[6], J. Posey[1], S. Plon[3], J. Lupski[1], R. Gibbs[3], L. Starita[5], C. Robles-Espinoza[7], W. Coyote-Maestas[8], I. Gallego Romero[9]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Washington, Seattle, WA, [3]Baylor Coll. Med., Houston, TX, [4]HGSC - Baylor Coll. of Med., Houston, TX, [5]Univ of Washington, Seattle, WA, [6]Walter and Eliza Hall Inst. of Med. Res., Melbourne, Australia, [7]Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, [8]Univ. of California San Francisco, San Francisco, CA, [9]St Vincent's Inst. of Med. Res., Fitzroy, Australia

Abstract:

Multiplexed Assays of Variant Effects (MAVEs) can test all possible single variants in a gene of interest. The resulting saturation-style data may help resolve variant classification disparities between populations, especially for variants of uncertain significance (VUS). We analyzed genome sequencing data from *All of Us* and replicated our analyses in two independent versions of the Genome Aggregation Database (overall 213,663 individuals of European-like versus 206,975 individuals of non-European-like genetic ancestry). Using two orthogonal statistical approaches, we show a higher prevalence ($p \leq 5.95e-06$) of VUS in individuals of non-European-like genetic ancestry across all medical specialties assessed in all three databases. Further, in the non-European-like genetic ancestry group, higher rates of Benign or Likely Benign and variants with no clinical designation ($p \leq 2.5e-05$) were found across many medical specialties, whereas Pathogenic or Likely Pathogenic assignments were higher in individuals of European-like genetic ancestry ($p \leq 2.5e-05$). We show over 86% of VUS in each medical specialty are missense variants, and missense VUS are significantly higher in individuals of non-European-like genetic ancestry ($p \leq 6.25e-07$). When missense VUS were excluded, a significant difference in VUS rates persisted between the two superpopulations ($p \leq 2.5e-05$). In-frame indels, splice region, and synonymous variants also drive the VUS disparity ($p \leq 6.25e-07$). Then, we

automated the Clinical Genome Resource's Variant Curation Expert Panel rules for *BRCA1*, *TP53*, and *PTEN* and incorporated clinically calibrated MAVE data. We achieved a cumulative 73.7% VUS reclassification rate. We reclassified VUS in individuals of non-European-like genetic ancestry at a significantly higher rate in comparison to reclassified VUS from European-like genetic ancestry (p=9.1e-03) effectively compensating for the VUS disparity. Further, essential code analysis showed equitable impact of MAVE but inequitable impact of allele frequency (p=7.47e-06) and computational predictor (p=6.92e-05) evidence codes towards VUS reclassification. We show likely hundreds of thousands of individuals of non-European-like genetic ancestry have had inequitable variant interpretations due to statistically significant bias in commonly used computational predictors such as REVEL and BayesDel (p≤2.5e-05). In part, this bias in computational predictors explains the VUS disparity for which MAVE evidence compensates. These findings highlight the need for proactive, saturation-style MAVE data production to enhance inclusive understanding of clinical variation.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6067F: Regulatory variation in MTOR dampens Neutrophil-T cell crosstalk during sepsis

Authors:

P. Zhang[1], P. MacLean[1], A. Jia[1], K. Cano-Gamez[1], H. Qin[1], C. Wainwright[1], A. Allcock[1], G. Scozzafava[1], K. Burnham[2], E. Davenport[2], J. Knight[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Wellcome Sanger Inst., Hinxton, United Kingdom

Abstract:

Sepsis is a heterogeneous syndrome and personalised strategies are essential to successful targeted therapeutics. A sepsis response signature, SRS1, identifies a subset of severe patients that were associated with immunosuppressive phenotypes. We aim to identify and characterise genetic markers that interact with SRS endotypes, and mediators involving key cellular phenotypes that could inform novel therapeutics. We mapped the epigenomic determinants of variation for *MTOR* expression in sepsis to identify causal cell-cell interplay. We assayed the genetic and cellular interactions using an *ex vivo* co-culture model and omics sequencing. We identify genetic variation has an inverted effect on *MTOR* expression between activated T cells and neutrophils, decreasing and increasing expression respectively. We find significant interaction between the variation and prognostic markers SRS endotype and neutrophil to lymphocyte ratio. In co-culture, upon TCR stimulation T cells promote sepsis neutrophils, which in turn suppress T cell activation. This crosstalk is dampened by hypoxia and rapamycin involving

IL3/CD123 axis. The G allele with reduced mTOR signalling and cytokine release in T cells increases survival during sepsis. This study offers evidence of how common variation can interact with disease endotypes and present a novel patient stratification strategy in sepsis. More widely, these findings may be informative for understanding severe infection associated with CAR T therapy.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6068F: Responsible genes for triplet repeat diseases are involved in the ubiquitin proteasome system

Authors:

H. Hohjoh[1,2], H. Shimizu[2], T. Okura[2], N. Hattori[1]; [1]Juntendo Univ., Tokyo, Japan, [2]Natl. Inst. of NeuroSci., NCNP, Tokyo, Japan

Abstract:

Fragile X syndrome (FXS) and Huntington's disease (HD) belong to the family of triplet repeat diseases caused by an abnormal expansion of trinucleotide repeats in the responsible *Fragile X mental retardation 1* (*FMR1*) and *Huntingtin (HTT)* genes, respectively. The disease-causing *FMR1* gene carries abnormally expanded CGG trinucleotide repeats that inhibit the expression of *FMR1* and cause FXS. The disease-causing *HTT* gene has an abnormally expanded CAG repeat in exon 1, which is translated into polyglutamine causing abnormal aggregation of the HTT protein, and the aggregated HTT protein is harmful in the brain. Unwanted and defective proteins, including aggregates, are systematically degraded in the cell. In other words, cells have a proteolytic system to eliminate unwanted and defective proteins from the cell. Therefore, dysfunction (or abnormality) of the intracellular proteolytic system may be related to the accumulation and formation of disease-causing gene products. The ubiquitin-proteasome system (UPS) is the primary ubiquitin-dependent proteolytic pathway in the cell. Our previous study examined the UPS in cells, where disease-causing and related genes for FXS were knocked down by RNA interference (RNAi). Mouse neuro2a cells, a mouse neuroblastoma cell line, were subjected to gene silencing for *Fmr1*, *Fmr1 autosomal homolog 1 (Fxr1)* and *discs large MAGUK scaffold protein 4* (*Dlg4*) genes by RNAi, and UPS activity was examined. The results indicated that knockdown of *Fmr1*, *Fxr1* and *Dlg4* genes resulted in increased ubiquitination and that *Fmr1*- and *Fxr1*-knockdown further increased proteasome activity. Knockdown of these genes also affected process formation in N2a cells. Therefore, the findings suggested that these genes are involved in the UPS, and that UPS may contribute to neurite outgrowth. Based on the previous study, we now performed similar experiments with the *Htt* gene. Knockdown of HTT in

N2a cells resulted in increased UPS activity. UPS activity was also increased in R6/2 mice, a mouse model of HD. These results suggest that HTT may be a factor associated with the UPS. Furthermore, the results of *Htt*-knockdown were similar to those of *Fmr1*-knockdown in either UPS activity or process formation in N2a cells, suggesting that both genes may be involved in similar cellular functions.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6069F: RNA-based analysis of potential splicing variants facilitates genome diagnostics and reveals limitations of in silico prediction tools

Authors:

M. Drost[1], J. Dekker[1], F. Ferraro[1], E. Kasteleijn[1], H. Douben[1], L. van Unen[1], M. Hoogeveen-Westerveld[1], E. Kroon[1], P. Elfferich[1], R. Schot[1], I. Vogt[1], E. Korpershoek[1], F. Sleutels[1], H. Bruggenwirth[1], I. Hollink[1], L. Meerstein-Kessel[1], L. Hoefsloot[1], M. van Slegtenhorst[1], M. Wilke[1], M. Weerts[1], R. van Minkelen[1], A. Wagner[1], A. Bouman[1], B. van Paassen[1], G. Verheijen-Mancini[1], I. de Graaf-van de Laar[1], A. Kievit[1], J. Verhagen[1], K. Stuurman[1], L. Donker Kaat[1], M. van Dooren[1], M. Wessels[1], R. Oldenburg[1], S. Zeidler[1], T. van Dijk[1], S. Barakat[1], V. Verhoeven[1], Y. van Bever[1], Y. van Ierland[1], N. Bannink[2], S. Van Koningsbruggen[3], P. Lakeman[3], L. Leeuwen[4], N. Verbeek[5], M. Sinnema[6], M. Heijligers[6], C. van Asperen[7], J. Saris[1], M. Nellist[1], T. van Ham[1]; [1]Dept. of Clinical Genetics, ErasmusMC, Rotterdam, Netherlands, [2]Dept. of Pediatrics, Franciscus Gasthuis & Vlietland, Rotterdam, Netherlands, [3]Dept. of Human Genetics, Amsterdam Univ. Med. Ctr., Amsterdam, Netherlands, [4]Dept. of Genetics, Univ. Med. Ctr. Groningen, Groningen, Netherlands, [5]Dept. of Med. Genetics, Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [6]Dept. of Clinical Genetics, Maastricht Univ. Med. Ctr., Maastricht, Netherlands, [7]Dept. of Clinical Genetics, Leiden Univ. Med. Ctr., Leiden, Netherlands

Abstract:

Background/Objectives: DNA variants that affect pre-mRNA splicing are an important cause of genetic disorders. Because the effects are diverse and can be difficult to predict, international variant classification guidelines recommend experimental characterization to confirm defective pre-mRNA splicing prior to DNA variant classification. However, in many diagnostic laboratories, in vitro investigations of mRNA splicing are not routinely performed. Therefore, insight into the added value of routine experimental investigation is scarce. We assessed the utility of in vitro analysis of pre-mRNA splicing in a diagnostic setting and compared the outcomes with in silico mRNA splicing prediction algorithms. **Methods:** We investigated DNA variants, identified in individuals for whom diagnostic testing had been requested. We analyzed pre-mRNA splicing in

blood or skin fibroblasts by RT-PCR, followed by agarose gel electrophoresis and Sanger sequencing, and/or by performing exon trap assays. We compared the outcomes to the in silico splicing prediction tools SpliceAI, Squirls, SPiP, Pangolin, VulExMap as well as the prediction tools integrated in Alamut. **Results:** An effect on pre-mRNA splicing was demonstrated in 128/202 (63%) of the tested variants and facilitated reclassification of 96/177 (54%) variants of uncertain significance (VUS). In total, 85/177 variants (42%) were reclassified as (likely) pathogenic. Our comparison of splice prediction algorithms showed that their performance differed, is dependent on variant type and location and that no single tool classified all variants equally well. We describe several examples of hard-to-predict effects and unexpected results highlighting the limitations of prediction tools. **Conclusion:** Experimental analysis of pre-mRNA splicing in a diagnostic setting facilitated variant classification, enabling reclassification of 85/177 (48%) VUS as (likely) pathogenic. In silico predictions of pre-mRNA splice sites often correlate with in vitro analyses, but should be used with caution, consistent with existing guidelines. We provide a framework for RNA-based analysis in a molecular diagnostic setting, showing the added value of routine testing of RNA from individuals with a variant that could affect pre-mRNA splicing.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6070F: Role of Trisomy 21 in Hirschsprung Disease

Authors:

K. Singh, T. Davoli, A. Chakravarti; NYU Grossman Sch. of Med., NEW YORK, NY

Abstract:

Hirschsprung disease is a rare neurodevelopmental disorder marked by the partial or total absence of enteric neurons in varying lengths of the gastrointestinal tract. The disorder is one of a functional intestinal blockage owing to the absence of motility from the enteric nervous system (ENS). Notably, individuals with trisomy 21 (Down syndrome) have an ~100-fold increased risk of Hirschsprung disease. This study aims to explore the underlying molecular causes for this heightened susceptibility in Down syndrome patients. To understand the underlying molecular processes, we differentiated induced pluripotent stem cells (iPSCs), both with and without trisomy 21, into enteric neuronal crest cells (ENCCs). Our initial findings indicate that during differentiation, iPSCs from both groups lose the stem cell marker *OCT4*. As differentiation proceeds into neural crest cells, and subsequently into ENCCs, there is significant increase in *HOXB3* gene expression levels. We also looked at the chromosome 21 gene *SOD1*, as a candidate mediator of ENCC changes. First, *SOD1* gene expression levels do not change during

the differentiation of iPSCs into ENCCs. Second, however, the ENCC marker gene *RET*, one of two major genes for Hirschsprung disease, also shows increased expression during differentiation. Third, we did not observe changes in other Hirschsprung disease genes, like the second major gene *EDRNB*. Fourth, interestingly, our results reveal that iPSCs with an extra copy of chromosome 21 exhibit ~1.5 fold *lower* levels of *RET* expression as compared to disomic cells with chromosome 21. This suggests that trisomy 21 can itself modulate some early markers of enteric neural crest cells, particularly *RET*, indicating a cell intrinsic effect of the extra chromosome. These findings provide the basis for further experiments to dissect how trisomy 21 increases the risk of Hirschsprung disease in Down syndrome patients.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6071F: Scalable single-cell models for robust cell-state-dependent eQTL mapping

Authors:

J. Alquicira Hernandez[1,2,3,4], A. Nathan[1,2,3], M. S. Jain[1,2,3], S. Raychaudhuri[1,2,3]; [1]Brigham and Women's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Garvan Inst. of Med. Res., Sydney, Australia

Abstract:

Disease risk variants identified through genome-wide association studies (GWAS) are enriched in non-coding genomic regions, suggesting a potential regulatory role. However, regulatory variants identified through expression Quantitative Trait Locus (eQTL) mapping exhibit only a modest overlap with GWAS variants. This may be because traditional eQTL models aggregate RNA profiles from many cells and fail to capture the distinct regulatory programs of causal cell states. We recently proposed a single-cell-resolution eQTL modeling approach (Nathan A, *et al*, 2022) that offers a high-resolution approach to linking genetic variation to gene expression in cell types (such as T cells) or states (e.g. cytotoxicity and proliferation). However, accurately modeling single-cell expression represents a challenge for single-cell eQTL mapping. In some instances, individual genes may exhibit distinct count distribution patterns in single-cell data due to data sparsity and heterogeneous contributions of various cell states. This violates parametric assumptions and may lead to p-value inflation and a higher false-positive rate. Furthermore, current eQTL methods fail to scale to large datasets of millions of cells. Therefore, they are limited to testing only a few variants across a whole locus already defined from pseudobulk analysis. To solve these problems, we propose a non-parametric linear mixed-effects modeling approach to map cell-state-dependent single-cell eQTLs. Key features of this

model include: 1) adaptive bootstrapping to derive statistically calibrated p-values without any assumption about the gene expression distribution, 2) pseudocell aggregation and covariate regression prior to eQTL mapping to reduce computation time while retaining statistical power. We applied this method to 500,089 memory T cells from 259 Peruvian individuals and identified cell-state-dependent eQTLs for autoimmune disease-risk genes such as IL2RA and CTLA4 which had no evidence of eQTL using pseudobulk analysis. Some eQTLs from single-cell resolution analysis were not detected with bulk analysis and had regulatory effects specific to rare states. These results suggest that testing eQTL interactions with cell states can identify cell-state-dependent regulatory variants that would be obscured in traditional eQTL mapping.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6072F: Single cell RNA sequencing-based modeling of the mismatch repair pathway to support clinical variant classification for Lynch Syndrome ⭐

Authors:

J. Reuter, S. Padigepati, D. A. Stafford, K. Nykamp; Invitae, San Francisco, CA

Abstract:

Intro: Experimental data is recognized as an important feature in variant classification frameworks, especially when interpreting rare missense variants with limited clinical data. Single cell RNA-seq (scRNA-seq) is emerging as a new approach for characterizing the effects of mutations at scale. Here, we describe scRNA-seq based experiments for four genes in the mismatch repair (MMR) pathway (*MSH2*, *MSH6*, *MLH1* and *PMS2*), including a saturation-level experiment on *PMS2*. We further demonstrate that scRNA-seq profiling can be used to generate a highly performant multi-gene model for the MMR pathway.
Methods: HEK293 cells harboring variants for *MLH1* (n=252), *MSH2* (n=257), *MSH6* (n=269) and *PMS2* (n=221) were exposed to a mutagen (6-thioguanine, 6-TG) and profiled using scRNA-seq. Supervised machine learning approaches were applied to discover gene expression-based patterns that accurately distinguished known pathogenic and benign mutations and could be used to make predictions on variants of uncertain significance. We scaled this experimental approach to provide predictions for nearly all possible single nucleotide accessible missense mutations for the *PMS2* (n=5702 variants). Finally, we developed a training and testing framework to build and validate a multi-gene model for MMR genes.
Results: Using a scRNA-seq based multiplex assay of variant effect (MAVE) strategy, we generate cellular models for each of the MMR genes. We show that model performance is dependent on

6-TG exposure and that the *MSH2* predictions are correlated (Pearson=0.89) with a previously published *MSH2* MAVE. We generated a saturation-level functional map for *PMS2* and determined that approximately 10% of the possible single nucleotide accessible missense mutations produce gene expression changes consistent with pathogenicity. Conceptually, scRNA-seq based MAVEs have the potential to share information across genes in a pathway. To test this hypothesis, we developed and validated a multi-gene model, which exhibited an AUROC of 0.96 and was MMR-gene specific.
Conclusions: We report highly performant MAVE models to support variant classification across the MMR pathway genes responsible for Lynch Syndrome. For *PMS2*, a gene where variant classification is complicated by both lower penetrance and high sequence homology to a pseudogene, we proactively generated high quality evidence for nearly all clinically relevant missense mutations. Finally, we construct a novel MMR-specific, multi-gene MAVE model, an advancement that has implications for generating experimental evidence for genes with limited clinical knowledge as well as learning pathway level mechanisms of pathogenicity.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6073F: Single-cell long-read sequencing of normal lung tissues from Asian women characterizes transcript heterogeneities and facilitates splicing-QTLs

Authors:

B. Li[1], T. Luong[1], E. Sisay[1], J. Yin[1], J. Shin[2], J. Byun[3], Y. Chang[2], N. Rothman[1], Q. Lan[1], C. Amor[3], E. Long[4], T. Zhang[1], H. Zhang[1], J. Shi[1], J. Lee[2], E. Kim[2], J. Choi[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [3]Baylor Coll. of Med., Houston, TX, [4]Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China

Abstract:

Background: Splicing quantitative trait loci (sQTL) analysis can link genetically controlled alternative splicing to GWAS signals unexplained by expression QTLs. However, current studies using bulk short-read sequencing often rely on computational estimations of isoform levels or splice junction reads, failing to capture full-length and novel isoforms or cell type specific splicing events. Lung cancer is a prevalent and life-threatening disease, influenced by both environmental and genetic risk factors. Genome-wide association studies (GWAS) have identified > 50 genomic loci associated with lung cancer across diverse populations, but most loci are non-protein coding and not yet functionally characterized.**Methods**: To uncover cellular transcriptome variations and potential genetic regulatory mechanisms relevant to lung cancer,

we performed PacBio single cell MAS-seq on tumor-distant normal lung tissues from 129 never-smoking Korean women. Since lung cancer originates from epithelial cells, we enriched epithelial cells using fluorescence-activated cell sorting and multiplexed ~six individuals per batch. DNA genotyping and imputation of matched blood were used for demultiplexing individuals and facilitating QTL analyses. For doublet removal, genotype-based Vireo and RNA expression-based Scrublet were adopted.**Results:** Following quality control, we retained 377,643 cells from 129 individuals. Cells were annotated by mapping to the Human Lung Cell Atlas (HLCA) using scArches and Azimuth and further refined by iterative clustering, which resulted in 37 cell types, including 9 epithelial populations accounting for 45% of cells. At the isoform level, transcriptomes were classified using SQANTI3 and integrated across batches with TALON, leading to 831,469 isoforms with 93.3% being novel. TransDecoder assessed the functional significance of novel transcripts and removed potential artifacts. Collectively, 326,530 isoforms from 21,616 genes were retained and utilized to create the long-read augmented genome annotation. Notably, 39,574 isoforms including 16,411 novel ones from 13,518 genes, were consistently detected in more than 50% of individuals.**Conclusion**: We plan to further characterize isoform-level cellular heterogeneities and decipher genetic effect on the transcript structure regulation by performing isoform-level sQTL and allele-specific analysis within each cell type. Integrating lung cancer GWAS data will help identify cell-type specific susceptibility genes. We aim to establish a large-scale long-read cell atlas of lung tissues to facilitate the understanding of lung cancer etiology.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6074F: Single-neuron somatic mutation in the context of Alzheimer's disease tau pathology

Authors:

B. Jin[1], S. M. Naik[1], S. L. Kirkham[2], D. H. Oakley[3], M. P. Frosch[3], B. T. Hyman[3], Y. Huang[4], M. Miller[5]; [1]Brigham and Women`s Hosp., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Boston Children s Hosp., Boston, MA, [5]Brigham and Womens Hosp, Boston, MA

Abstract:

Despite being the most clinically correlated pathology, the precise role of hyperphosphorylated tau protein (P-tau) in Alzheimer's disease (AD) remains largely unknown. Tau pathology parallels brain regions experiencing atrophy and neuronal loss. However, even in the advanced stages of AD, less than 20% of neurons in affected brains exhibit tau pathology. Are these tau-bearing

neurons physiologically distinct from nearby tau-free neurons? Meanwhile, AD neurons harbor a higher somatic mutation burden than healthy neurons, with oxidative and other disease-specific mutational signatures. This observation prompts the question of whether somatic mutations are driven by disease pathology. We hypothesized that neurons with pathological P-tau will carry more somatic mutations than neurons without P-tau. We developed a method to isolate single neuronal nuclei based on cytoplasmic P-tau aggregation with fluorescence-activated nuclear sorting and then performed single-cell whole genome sequencing (scWGS) on these nuclei. This sorting method effectively enriches P-tau+ nuclei (>380 fold) with high accuracy (98%). We then performed scWGS on P-tau sorted neuronal nuclei from 5 AD prefrontal cortex samples, targeting 4 P-tau+ and 4 P-tau- nuclei per individual. We compared the genomes of P-tau-sorted neurons with 29 tau-undetermined neurons from the same AD samples and 52 neurons from neurotypical control individuals. We used SCAN2 to identify single-cell somatic mutations and performed mutational signature analysis. We found that total somatic single-nucleotide variant (sSNV) burden and oxidation-related mutations are both elevated in the P-tau+ and P-tau- neurons (P-tau- vs control: +183 sSNV/neuron, p = 2.74e-3; P-tau+ vs control: +163 sSNV/neuron, p = 2.9e-3). Notably, P-tau+ and P-tau- neurons do not differ significantly in sSNV burden (P-tau+ vs P-tau-: -20 sSNV/neuron, p =0.75) and oxidative signature mutations. Similarly, P-tau+ and P-tau- neurons carry more indels (insertions and deletions) than control neurons, but we observed no significant difference between P-tau+ and P-tau- neurons. We found that a specific indel mutational signature, ID4, is elevated in both P-tau+ and P-tau- neurons compared to controls. ID4 has previously been reported as related to defective ribonucleotide excision repair. In summary, our results suggest that P-tau accumulation, the most prominent hall marker for AD, occurs independently of somatic mutation. This study is the first to investigate the somatic genomic profiles of tau-specified single neurons and demonstrates an independent relationship between DNA damage accumulation and P-tau aggregation.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6075F: Somatic Variants in Focal Cortical Dysplasia Create Disease-Associated Transcriptional Programs without Altering Cell Identities

Authors:

S. Khoshkhoo[1], S. Bizzotto[2], M. Talukdar[3], E. J. Stronge[3], Q. Hu[4], B. Chhouk[3], A. M. Tillett[1], A. D'Gama[3], P. V. Kharchenko[4], A. Poduri[3], C. Walsh[3]; [1]Brigham and Women s Hosp., Boston,

MA, [2]Sorbonne Univ., Paris, France, [3]Boston Children's Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA

Abstract:

Somatic variants are the most common cause of intractable pediatric focal epilepsies and frequently associated with malformations of cortical development such as focal cortical dysplasia (FCD). Histopathological hallmarks of FCDs are cortical mislamination and the presence of morphologically abnormal cells such as dysmorphic neurons (DNs) and balloon cells (BCs) which are indicative of pathogenic variants activating the PI3K-mTOR signaling pathway. Electrophysiologic studies have shown that DNs are inherently hyperexcitable, but the identities of DNs and BCs is still debated and the mechanisms through which a small mutant cell population causes disease is unknown. To elucidate the transcriptional mechanisms of FCD we performed single nucleus RNA sequencing (snRNA-seq) on 20 surgically resected FCD samples, representing a range of pathogenic variants in *PIK3CA, MTOR, DEPDC5,* etc., and 8 neurotypical postmortem control samples. Differential gene expression analysis in cases vs controls showed significant and likely compensatory downregulation of gene pathways downstream of mTORC1 signaling in all the major cell types, but most significantly in neurons. Moreover, the FCD-associated microglia exhibited a transcriptional program consistent with a pro-inflammatory state with corresponding activation of immune-related gene pathways in astrocytes and oligodendrocytes. Surprisingly, no new cell clusters representing DNs or BCs were identified and all major cell type proportions were unaffected in the FCD tissue. To determine the cell autonomous mechanisms of somatic variants in FCD, we developed GoTEN which combines 10X snRNA-seq with Oxford Nanopore (ONT) long-read amplicon sequencing to simultaneously obtain cell identities and transcriptional states associated with mutant cells. We applied GoTEN to 3 samples harboring *PIK3CA* and *MTOR* variants. GoTEN analysis detected the pathogenic variants mostly in excitatory and in some inhibitory neurons, although a small percentage of mutant glial cells were also seen, which is consistent with shared neuroglial progenitors. Strikingly, the mutant cells had indistinguishable cellular identities relative to their non-mutant counterparts although they did have disease-associated transcriptional changes mostly restricted to PI3K-mTOR pathway genes. In summary, using a combination of snRNA-seq and a new method, GoTEN, our study highlights genotype-specific transcriptional changes and a pro-inflammatory state in FCD which could be potential targets for precision therapies in FCD. Additionally, despite drastic morphologic changes *in situ,* we find little to no change in mutant cell identities.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6076F: Systematic cell-specific assessment of childhood obesity GWAS loci using massively parallel reporter assay

Authors:

W. Bruner[1], K. Trang[1], S. Grant[2]; [1]Children s Hosp. of Philadelphia, Philadelphia, PA, [2]Children s Hosp. of Philadelphia/UPenn, Philadelphia, PA

Abstract:

The global prevalence of childhood obesity continues to rise, driving an increase in disease outcomes such as diabetes and cardiovascular traits. While previous GWAS meta-analyses for pediatric obesity conducted by the Early Growth Genetics (EGG) consortium discovered risk loci for childhood obesity, such efforts alone cannot provide mechanistic insights in to how they confer susceptibility. In order to determine which variants at these loci drive differential transcription, we utilized a massively parallel reporter assay (MPRA), which enables assessment of many putative regulatory sequences simultaneously. Our MPRA experimental design was based on summary statistics derived from individual populations in the latest EGG trans-ethnic meta-analysis of the trait. We elected to pursue all genome wide significant signals, along with any sub-threshold signals that either achieved population-specific $P<1x10^{-6}$ or overall trans-ethnic results of log_{10}Bayes factor >4.4 (equating to ~ $P<1x10^{-6}$). Each variant which met the required threshold was then subjected to LD expansion at an r^2 cutoff of 0.8 based on genome build GRCH37 (hg19) and variant sets within 1000 Genomes, Phase 3.v5 via SNiPA v3.4, and subsequently validated using the NIH LDLink online resource. Final biallelic SNP set consisted of 2,453 putative target elements for a total of 4,906 complete sequences included in the final MPRA library. Our selected variants were then assessed by MPRA for their differential transcriptional regulatory activity in neuronal-based cell types. Recognizing the central nervous system's pivotal role in the pathogenesis of obesity, our initial efforts were focused on HMC3 neuronal microglial and the SH-SY5Y neuroblast-like cell lines. Using the linear model framework *mpralm* R package analysis tool, we observed >300 variants with allele-specific signatures exclusive to either HMC3 or SH-SY5Y cells, with three variants (*rs75973443-ADCY3*, *rs34033929-ADCY9,* and *rs2749796-LINC01441*) yielding consistent differential allelic activity across both cell types. These findings suggest a robust cell-specific activity for these tested variants. This method therefore offers an opportunity to gain mechanistic insights into the pathogenesis of childhood obesity.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6077F: TET2 clonal haematopoiesis is associated with improved antibody responses to SARS-CoV-2 vaccination and potentiates a vaccine-response endotype

Authors:

P. Maclean, A. Jia, **C. O'Neill**, I. Dyne, A. Chong, A. Mentzer; Ctr. for Human Genetics, Oxford, Oxford, United Kingdom

Abstract:

Background: Older age is strongly associated with impaired responses to vaccination, including for SARS-CoV-2. The prevalence of clonal haematopoiesis of indeterminate potential (CHIP) is also strongly age-associated, but some CHIP mutations appear to augment adaptive responses. We aimed to test for an association between CHIP and SARS-CoV-2 vaccine responsiveness in a large cohort. Methods: UK Biobank (UKB) participants who received two doses of SARS-CoV-2 vaccine and returned a binary finger prick lateral flow test detecting anti-Spike IgG were profiled for myeloid and lymphoid CHIP status using whole exome sequencing data. We tested for associations between seroconversion and CHIP status, as well clonally-expanded mosaic chromosomal alterations (mCAs). We adjusted for factors known to affect post-vaccination humoral responses including age, sex, smoking, BMI, vaccine type, prime-boost interval, class II HLA risk alleles, incident cancer diagnosis and a composite measure of immunocompromise, and excluded participants with prior SARS-CoV-2 infection using UKB and Public Health England data. We also used UKB plasma proteomics data to test the effect of CHIP on an innate immune endotype predictive of vaccine responsiveness.Results: 44,372 UKBB participants with no known prior infection were tested at least 14 days after their second dose. 30,823 were seropositive, and 13,549 seronegative. After adjustment for known determinants of vaccine responsiveness, CHIP in myeloid driver genes was associated with higher odds of seroconversion after two doses of SARS-CoV-2 vaccine (OR 1.09, 95% CI 1.00 - 1.20). Among the most common CHIP driver genes, mutations in TET2 were associated with a higher rate of seroconversion (OR 1.32, 95% CI 1.06-1.63) while DNMT3A mutations were not (OR 0.95, 95% CI 0.84 - 1.08). Rarer CHIP mutations in lymphoid driver genes were not associated with a difference in the antibody response (OR 1.05, 95% CI 0.89 - 1.24). Clonal somatic mCAs, including loss of the X or Y chromosome and autosomal aneuploidy, did not affect seroconversion. In participants with targeted proteomics data available, TET2 CHIP potentiated an innate immune signature previously associated with superior antibody responses after vaccination.Conclusions: TET2 CHIP mutations were linked to a higher likelihood of seroconversion after two SARS-CoV-2 vaccine doses, independent of other factors affecting vaccine responsiveness. The association

with TET2, and not other clonally-expanded haematopoietic mutations, suggests a specific effect. TET2 CHIP may amplify an inflammatory endotype linked to superior antibody responses.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6078F: The allelic spectrum of *SCN4A*-associated disease in over 7000 individuals: lessons learned from genetic testing.

Authors:

C. Lopes[1], M. McCarthy[1], F. Lebron[1], G. Maston[1], Z. Powis[1], I. Karbassi[2]; [1]Quest Diagnostics, Seacaucus, NJ, [2]Quest Diagnostics, Marlborough, MA

Abstract:

Background

SCN4A belongs to a gene family that encodes Na-channels, which are critical for cell electrical signaling. Pathogenic variants in the *SCN4A* gene cause a group of related autosomal dominant disorders that include periodic paralysis and myotonia and are prevalent in approximately 1/100,000 of the population. To identify regions associated with pathogenicity, we investigated the allelic spectrum of variants in a large cohort of individuals undergoing *SCN4A* testing.

Methods and Results

We retrospectively reviewed variants from 7,080 individuals undergoing single- and multi-gene panel *SCN4A* testing at a CLIA-certified diagnostic laboratory. Variants were classified using published guidelines in line with ACMG/AMP recommendations.

Variant spectrum: Of 444 unique variants detected, 66% were missense, 26% were synonymous, 8% were intronic, and less than 1% were frameshift. From 7,080 tests, reportable genetic findings, including variants classified as pathogenic (P), likely pathogenic (LP), and of uncertain significance (VUS), were identified in approximately 19% of individuals, with variants classified as P/LP detected in 7%. Autosomal dominant *SCN4A*-associated disorders are caused by a gain-of-function mechanism. Accordingly, our data showed that P/LP variants were almost exclusively missense (96%) and were limited to 50 unique variants in 38 amino acids. In contrast, intronic and synonymous variants accounted for 65% of the benign (B) and likely benign (LB) variants. Missense VUS accounted for 66% of the total missense variants observed, highlighting the need to develop *SCN4A* specific variant interpretation guidelines.

Hotspot regions: We investigated the location of B/LB and P/LP missense variants observed and identified regional hotspots (PM1), where P/LP variants were enriched and B/LB variants were absent. Consistently, these hotspots were in important protein functional domains. Approximately 50% of all patient reports with VUS had variants located in these regions. We re-

examined all variants currently classified as VUS in the hotspot regions and approximately one third of these variants could be re-classified based on new and/or newly evaluated information, including hot spot classification.

Conclusion

Our data suggest that defining hotspot regions in *SCN4A* associated with higher risk of pathogenicity can help classify missense VUS in this gene and lead to more informed prediction algorithms, thus affecting variant classification and accurate diagnosis by clinicians. Because these Na-channel regions are highly homologous, these data may have a broader impact in channelopathies.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6079F: The effect of rare gonadotropin pathway variants on PCOS pathogenesis.

Authors:

G. Stokes[1], C. Parker[1], R. Bauer[1], C. Welt[2], R. Legro[3], M. Urbanek[1]; [1]Northwestern Univ., Chicago, IL, [2]Univ. of Utah, Salt Lake City, UT, [3]MS Hershey Med. Ctr, Hershey, PA

Abstract:

Background: Polycystic ovary syndrome (PCOS) is a common, complex endocrinopathy affecting 6-20% of reproductive age women. It is defined by elevated androgens, irregular or absent periods, and polycystic ovarian morphology. Physiological conditions contributing to PCOS pathogenesis include insulin resistance, and ovulatory and gonadotropin signaling dysfunction. Twin and family studies have shown PCOS to be highly heritable. Furthermore, genome-wide association studies (GWAS) implicated genes in insulin resistance, diabetes, ovulation, and gonadotropin signaling. The gonadotropin hormones, follicle stimulating hormone (FSH) and luteinizing hormone (LH), are neuroendocrine hormones that direct follicle development and overall ovarian health. Reproductive neuroendocrine dysfunction is a major pathophysiological component of PCOS. Women with PCOS have elevated LH levels contributing to hyperandrogenism and oligo/anovulation—two cardinal characteristics of PCOS. **Purpose:** Given the physiological relevance of gonadotropin signaling dysfunction and genetic implications of gonadotropin genes from PCOS GWAS, we hypothesize that protein coding variation in gonadotropin genes contributes to PCOS pathogenesis in a subset of women with PCOS. Therefore, our *objective* is to identify the extent that potentially pathogenic variants (PPVs) in gonadotropin signaling genes contribute to PCOS and elucidate the molecular mechanisms of gonadotropin signaling dysfunction in PCOS. **Design/Methods:** Targeted (discovery cohort) and

whole exome sequencing (WES; replication cohort) identified protein coding variation in *FSHR, FSHB, LHCGR,* and *LHB* genes in women with PCOS. We applied the American College of Medical Genetics (ACMG) criteria to identify PPVs and test for association with PCOS using gene-based tests. **Results:** After applying ACMG criteria, we identified four variants as benign, one as likely benign, and 39 variants as PPVs (26 variants of unknown significance (VUS), nine likely pathogenic, and four pathogenic) in *FSHR, FSHB, LHCGR, and LHB*. Furthermore, PPVs in *FSHR* and *LHB* show significant and reproducible evidence for association with PCOS in the two cohorts. **Conclusion:** PCOS GWAS loci have mapped near gonadotropin genes *(FSHB, FSHR, LHB)*, but no studies have identified pathogenic variants in these genes or performed functional evaluations of gonadotropin gene variants in PCOS. We will identify how gonadotropin dysfunction is an underlying cause of PCOS, which members of the gonadotropin signaling pathway are impaired in PCOS, and how the pathogenic variants in this pathway impair gonadotropin signaling at the cellular level.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6080F: The human host genetic determinants of immune function across ancestry and disease

Authors:

J. Botey-Bataller[1,2,3], N. van Unen[1,3], ImHoGen consortium, M. Netea[2,4], Y. Li[1,2,3]; [1]Ctr. for Individualised Infection Med., Hannover, Germany, [2]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [3]TWINCORE, Hannover, Germany, [4]Univ. of Med. and Pharmacy, Craiova, Romania

Abstract:

Immune responses are highly diverse among individuals. The genetic causes of this diversity and their effect on immune-mediated diseases are yet to be fully characterized. To study the effect of host genetics on immune function, we established ImHoGen, a multi-ancestry, multi-disease consortium partnered with the Human Functional Genomics Project (HFGP), encompassing 12 independent cohorts recruited in six different countries in Africa, Asia and Europe. In total, ImHoGen includes 5,629 genotyped individuals with comprehensive immune function profiling, assessed by measuring the production capacity of 12 cytokines upon 22 stimulants. QTL mapping of cytokine responses (cQTL) revealed 56 genome-wide significant cQTLs. Among these, cQTLs exert either cis effects on cytokine genes or trans effects mediated by immune receptors and effector molecules. To decipher the role of genetic ancestry on immune function, we associated admixture proportions with cytokine responses in each cohort, finding a link between African genetic ancestry and the strength of the pro-inflammatory responses, and

European genetic ancestry and anti-inflammatory cytokine responses. Lastly, we developed polygenic scores (PGS) of immune responses, which we tested in an independent cohort, finding a positive correlation with measured immune responses. Furthermore, the PGS were associated with infection severity in 235 patients with COVID-19, finding a higher genetically predicted IL-6 responses to LPS in more severe patients. Collectively, we present the first results from a multi-ancestry, multi-disease consortium studying the role of host genetics in immune function, allowing us to pinpoint genetic loci regulating immune response, the collective effect of genetic ancestry, and a polygenic score for immune function that is related to infection severity.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6081F: The influence of trinucleotide repeats in the androgen receptor gene on androgen-related traits and diseases

Authors:

T. Sasako[1], Y. Ilboudo[2], K. Y. H. Liang[3], Y. Chen[4], S. Yoshiji[5], B. Richards[6]; [1]McGill Univ., Montréal, QC, Canada, [2]Jewish Gen. Hosp., Montreal, QC, Canada, [3]McGill Univ., Montreal, QC, Canada, [4]Lady Davis Inst., Montreal, QC, Canada, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada

Abstract:

Background: Trinucleotide repeats in the androgen receptor have been proposed to influence testosterone signaling in men, but the clinical relevance of these trinucleotide repeats remains controversial.
Aim: To examine how androgen receptor trinucleotide repeat lengths affect androgen-related traits and disease risks and whether they influence the clinical importance of circulating testosterone levels.
Methods: We quantified CAG and GGC repeat lengths in the androgen receptor (*AR*) gene of European-ancestry male participants in the UK Biobank from whole-genome and whole-exome sequence data using ExpansionHunter, tested associations with androgen-related traits and diseases, and examined whether the associations between testosterone levels and these outcomes were affected by adjustment for the repeat lengths. We also explored circulating proteins associated with the repeat lengths among the 2,941 proteins whose circulating level was examined.
Results: We successfully quantified the repeat lengths from whole-genome and/or whole-exome sequence data in 181,217 males. Both repeat lengths were shown to be positively associated with circulating total testosterone level and bone mineral density, whereas CAG

repeat length was negatively associated with male-pattern baldness, but their effects were relatively small and were not associated with most of the other outcomes. Circulating total testosterone level was associated with various outcomes, but this relationship was not affected by adjustment for the repeat lengths. Further, circulating total testosterone level was shown to be associated with various circulating protein levels, but the repeat lengths were associated with almost none of them after adjustment for circulating total testosterone level.
Conclusion: In this large-scale study, we found that longer CAG and GGC repeats in the *AR* gene influence androgen resistance, elevate circulating testosterone level via a feedback loop, and play a role in some androgen-targeted tissues. Generally, however, circulating testosterone level is a more important determinant of androgen action in males than repeat lengths.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6082F: The Trisomy 21 Model Atlas

Authors:

A. Timkovich[1], B. F. Niemeyer[1], L. Dunn[2], K. Schade[3], N. Paul Eduthan[1], K. Waugh[1], P. Araya[1], J. Espinosa[1], M. Galbraith[4], K. Sullivan[3]; [1]Linda Crnic Inst. for Down Syndrome, Univ. of Colorado Anschutz, Aurora, CO, [2]Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, [3]Linda Crnic Inst. for Down Syndrome, Aurora, CO, [4]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Trisomy 21 (T21), also known as Down syndrome (DS), is the most common chromosomal abnormality and a leading cause of intellectual and developmental disability. T21 affects the development or function of nearly every organ system, though little is known how T21 affects gene expression across different tissues and organs. Therefore, we are developing a T21 Model Atlas of tissue-specific murine and human transcriptomes matched to detailed metrics of organ development, architecture, and pathology. Tissues from Dp16 (segmental duplication of MMU16) and TcMAC21 (HSA21q mouse artificial chromosome) mice and littermate controls across three developmental timepoints were examined using bulk transcriptome analysis, histopathological scoring, and multiplex immunofluorescence imaging. In parallel, a human atlas for key organs will be created from tissue-specific cells and organoid derived from age/sex matched euploid ad T21 human iPSCs. Bulk transcriptome profiles of the brain, heart, kidney, and lymph node showed tissue-specific dysregulation of gene expression in the Dp16 mouse model, for both triplicated and non-triplicated genes. Additionally, histopathologic assessment shows the presence of congenital heart defects including atrial septal defects and ventricular septal defects in Dp16 embryos. Dp16 mice also demonstrate elevated liver pathology scores

which when paired with bulk transcriptome analysis identified genes associated with increased pathology scores in damaged Dp16 livers. Moreover, the brain of DP16 mice displays significantly reduced density of mature neurons, but no difference in microglial density within the dentate gyrus of the hippocampus. Additionally, bulk transcriptome analysis demonstrated dysregulated gene expression across differentiated cells derived from human iPSCs. The Trisomy 21 Atlas will improve our understanding of the effects of T21-mediated gene dysregulation on organ development and pathology, while enabling cross-validation of findings in mouse models of DS with human cell types. This atlas will act as a valuable and lasting resource to the DS research community.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6083F: Trans-pQTL network highlights the role of the NLRC4 inflammasome in regulating circulating IL-18 level

Authors:

D. Cousminer[1], C. Robins[1], S. Jariwala[2], J. Betts[3], L. McCarthy[3]; [1]GlaxoSmithKline, Collegeville, PA, [2]GlaxoSmithKline, Boston, MA, [3]GlaxoSmithKline, Stevenage, United Kingdom

Abstract:

IL-18 is a proinflammatory cytokine with complex roles in mucosal and systemic inflammation. Here, we identified genes associated with IL-18 plasma protein levels in UK Biobank proteomics data: gene-level rare-variant burden analyses and trans-pQTL analysis of common genome-wide genetic variation. Subsequently, we performed variant-to-gene mapping based on fine-mapped coding variants, distance and pathway-based priority scores, and MetaBase direct interactors to identify likely effector genes at the trans-pQTL loci. We then performed biological pathway enrichment analyses to provide functional annotation of these genes. Eight genes harbored a significant burden of deleterious rare genetic variants associated with IL-18 protein levels, including *NLRC4* and *IL18*. Genome-wide, eighteen trans-pQTLs associated with plasma IL-18 levels. Variant-to-gene mapping at the trans-pQTL loci highlighted several genes encoding components of the NLRC4 inflammasome, including *NLRC4*, the locus encoding caspase-1 and 4, *GSDMD*, and *VDR*. *NLRC4*, *SH2B3*, and *IL18* (as a *cis* interaction) were identified via both approaches. Common themes emerging from the pathway enrichment analyses included inflammasomes/pyroptosis, IL-1 family signaling, and TNF-mediated and JAK-STAT signaling pathways. Inflammasomes are key mediators of endogenous or exogenous stimuli like pathogens or tissue damage. They consist of a sensor (such as NLRC4) that activates caspase, which cleaves pro-IL-18 into mature IL-18 and activates gasdermin (encoded by *GSDMD*). IL-18

is secreted via gasdermin pores to carry out its effector functions in a variety of immune responses. This finely tuned system of inflammasome activation and inhibition includes the vitamin D receptor (VDR), which enhances NLRC4 inflammasome activation, and negative feedback from gasdermin, reducing caspase cleavage and IL-18 release. Aberrant inflammasome activation contributes to various diseases, including autoinflammatory, cardiovascular, and metabolic diseases, highlighting the importance of maintaining a balance between inflammasome activation and inhibition. Further work will assess the colocalization of trans-pQTL loci with disease associations to investigate how these genes interact with and impact IL-18 in disease-relevant contexts.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6084F: Triplication of HSA21-encoded genes modulates Alzheimer's Disease-related phenotypes in vivo via interferon signaling

Authors:

K. Schade[1], P. Araya[1], K. A. Waugh[1,2], L. N. Dunn[1], N. Busquet[3,4], H. R. Lyford[1], E. C. Britton[1], J. M. Espinosa[1,2], K. D. Sullivan[5,6]; [1]Linda Crnic Inst. for Down Syndrome, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Dept. of Pharmacology, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Animal Behavior Core, NeuroTechnology Ctr., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [4]Dept. of Neurology, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Linda Crnic Inst. for Down Syndrome, Aurora, CO, [6]Dept. of Pediatrics, Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

Introduction: Alzheimer's disease (AD) is the most commonly occurring neurodegenerative disorder, affecting up to 6 million people in the United States. People with Down syndrome (DS), the genetic condition caused by Trisomy 21 (T21), exhibit AD-related neuropathology by the age of 40, likely due to the triplication of the Amyloid Precursor Protein (APP) gene. However, the impact of many chromosome 21 (chr21) genes on AD phenotypes in the context of DS remains unknown. **Methods:** We generated a DS-AD mouse model by crossing Dp16 with 5xFAD (models for DS and AD, respectively), generating four genotypes: WT, Dp16, 5xFAD, and Dp16/5xFAD. In this study, we evaluated the cognitive/behavioral profiles of these mice at 3, 5, 7, and 9 months of age, and investigated the cell density and activation of microglia, astrocytes, and neurons using a multispectral immunohistochemistry (IHC) analysis. **Results:** We found that Dp16/5xFAD mice exhibit increased cognitive impairment at 7 and 9 months of age in the Radial Arm Water Maze test compared to all other genotypes. Furthermore, Dp16/5xFAD mice showed

hyperactivity but no significant alterations in anxiety-related behavior in the Open Field test. Conversely, no differences were observed in motor coordination and associative learning and memory in these mice. IHC staining revealed that Dp16/5xFAD mice exhibit a significantly decreased number of neurons in the somatosensory cortex, dentate gyrus, and cerebellum, while there is an increased number of activated microglia cells. **Conclusions:** These results indicate that triplication of chr21 genes impacts neurological dysfunction in AD, exacerbating impairments in spatial learning and memory. However, deeper characterization is needed to understand the mechanisms by which T21 contributes to AD-pathogenesis and may also provide insights into possible **therapeutic strategies** to address the needs of this population. **Funding Support:** This work was supported by NIH R01AI145988 and R01AI150305

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6085F: *TSC2* loss in neural progenitor cells suppresses translation of ASD/NDD-associated transcripts in an mTORC1- and MNK1/2-reversible fashion

Authors:

R. Beauchamp[1], P. Martin[1], K. J. Szkop[2], F. Robert[3], S. Bhattacharyya[1], J. Brenner[1], N. E. Redmond[1], S. Huang[3], S. Erdin[1], O. Larsson[2], V. Ramesh[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Karolinska Inst., Stockholm, Sweden, [3]McGill Univ., Montreal, QC, Canada

Abstract:

Tuberous sclerosis complex (TSC) is an inherited neurocutaneous disorder where, in addition to hamartomatous tumors in many organs, patients often present with neurodevelopmental (ND) manifestations such as epilepsy and TSC-associated neuropsychiatric disorders (TAND) that includes autism spectrum disorder (ASD). TSC is caused by mutations in the *TSC1* or *TSC2* tumor suppressor gene encoding hamartin (*TSC1*) and tuberin (*TSC2*), which form a functional complex inhibiting mechanistic target of rapamycin complex 1 (mTORC1) signaling. This has led to FDA approval of first-generation, allosteric mTORC1-selective rapamycin analogs ("rapalogs") as a lifelong therapy for TSC tumors. Yet, while clinical trials report reduced epilepsy in 40% of TSC patients, rapalogs have been ineffective in treating TAND. mTORC1 signaling plays an essential role in protein synthesis, and aberrant mTORC1 activation due to *TSC1/TSC2* loss leads to phosphorylation of key substrates including translational repressor eIF4E-binding proteins (4E-BPs) that are largely insensitive to rapalogs. When unphosphorylated, 4E-BPs bind and sequester the eIF4E subunit, impairing formation of heterotrimeric eIF4F translational complex. Upon 4E-BP1 phosphorylation by activated mTORC1, eIF4E is released and, together with

MNK1/2-dependent eIF4E phosphorylation, initiates 5' cap-dependent translation via the eIF4F complex. Our recent studies of TSC1 patient-derived, CRISPR-edited isogenic neural progenitor cells (NPCs) revealed *TSC1*-dependent alterations in early ND phenotypes and mRNA translation. Notably, many genes related to synaptic regulation or ASD were rapamycin-insensitive, yet showed response to third-generation, bi-steric mTORC1 inhibitor RMC-6272. Here, we extend our studies to isogenic CRISPR-modified TSC2 patient-derived NPCs where we find ND phenotypic defects in *TSC2*-Null cells that are rescued upon treatment with RMC-6272, but not rapamycin. Polysome-profiling reveals significant overlap in differentially translated genes between *TSC1*- and *TSC2*-Null NPCs. Intriguingly, gene enrichment analyses show decreases in ASD/NDD- and epilepsy-associated genes that are reversed by treatment with either RMC-6272 or MNK1/2 inhibitor eFT508, indicating translational downregulation in a subset of neurodevelopmentally important genes that is dependent on mTORC1-eIF4F pathway activation. These results serve to establish the importance of mTORC1-4E-BP1- and MNK-eIF4E-mediated translation in TSC, ASD and other ND disorders, laying the groundwork to clinically evaluate drugs targeting this pathway as a potential treatment strategy for TAND and ASD/NDD.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6086F: Uncovering the Genetic Basis of Systemic Lupus Erythematosus in Diverse Populations

Authors:

C. Atighetchi[1], F. De Angelis[1], C. Guerra Amorim[2]; [1]California State Univ. Northridge, Woodland Hills, CA, [2]California State Univ. Northridge, West Hollywood, CA

Abstract:

Systemic lupus erythematosus (SLE) is a female-biased autoimmune disorder characterized by the immune system attacking the body's own tissues and organs resulting in inflammation and damage. SLE presents heterogeneously, from skin and joint damage to severe kidney, brain, or heart complications. Higher incidences are observed among Hispanics and African ancestry populations. However, most studies focus on individuals of European ancestry, highlighting the need for inclusive research on SLE across different ethnic groups.To address this, we are using two approaches. First, we identify SLE risk factors and assess disease severity in 3,808 SLE-affected individuals from the U.S. using NIH's All of Us Research Program data. Severity was assessed based on perceived health, SLE remission status, and comorbidities, analyzed across self-reported race/ethnicity to determine if severity varies. We observed a higher prevalence of females among SLE patients, consistent with prior research, showing a 9:1 gender ratio. SLE

cases included about 15% fewer individuals self-defined as White/European compared to controls, aligning with lower SLE prevalence in this group. There was a higher proportion of African-Americans among SLE patients. We uncovered notable differences in SLE severity among African American, Middle Eastern, Asian, and Pacific Islander individuals compared to those of European ancestry.Secondly, we conducted a Genome-Wide Association Study (GWAS) (n=245,394) to identify SLE-associated variants. Among the significant associations are variants in NRAS, a gene previously associated with autoimmune lymphoproliferative syndrome, and DUSP22, a gene previously associated with T cell activation in the JKAP pathway of lupus nephritis, linked to increased SLE risk in diverse populations. Given the lack of SLE studies in non-European backgrounds, these findings will help ensure more comprehensive and inclusive healthcare solutions, advancing the understanding and treatment of SLE across diverse ancestries.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6087F: Uncovering the Role of an African-specific *ABCA7* Frameshift Deletion on Lipid Metabolism and Alzheimer's Disease

Authors:

Y. Nam[1], B. DeRosa[2], C. Golightly[3], S. Simon[4], J. Arvizu[5], A. Ramirez[1], P. Whitehead[6], L. Adams[3], K. McInerney[6], T. Starks[7], H. Cukier[3], M. Cuccaro[8], J. Haines[9], G. Byrd[10], G. Beecham[3], D. Dykxhoorn[3], J. Young[3], J. Vance[3], M. Pericak-Vance[11]; [1]Univ. of Miami, MIAMI, FL, [2]Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Univ. of Miami, Miami, FL, [45]Univ. Of Miami, Miami, FL, [6]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [7]Wake Forest Sch. of Med., Winston-Salem, NC, [8]John P. Hussman Inst. for Human Genomics, Miami, FL, [9]Case Western Reserve Univ, Cleveland, OH, [10]NC A&T State Univ, Greensboro, NC, [11]Univ. of Miami Miller Sch. of Med, Miami, FL

Abstract:

Background: we previously identified a 44-base pair deletion in the ATP-binding cassette sub-family A member 7 (*ABCA7*) that is significantly associated with Alzheimer's disease (AD) in African Americans (AA), with a frequency 15% in AD. It produces a frameshift mutation resulting in a truncated protein (p.Arg578Alafs). *ABCA7* is a lipid transporter across cellular membranes. We have previously shown that the mutant mRNA is expressed in neurons, but whether it is translated into a stable protein is not known. We studied the effect of the deletion on the metabolism of lipids using isogenic induced pluripotent stem cells (iPSC) and differentiated

neuronal models. The aim of this study is to discover that *ABCA7* 44bp deletion to see if it has an effect on lipid metabolism in neurons. Methods: we analyzed scRNAseq data to determine the cell-type expression pattern of ABCA7. We generated a recombinant version of the *ABCA7* gene with and without the deletion and bearing an N-terminal flag tag which was transfected into HEK cells to analyze truncated ABCA7 stability and localization. We utilized CRISPR-based genome editing in iPSC lines from three independent AA-cognitively unimpaired individuals (WT) to introduce the isogenic AA-specific *ABCA7* deletion (-/+ and -/-). These were differentiated into induced neuron-like cells. A multi-omics study of the *ABCA7* deletion lines were performed using bulk-RNAseq, proteomics, and untargeted lipidomics. Results: the truncated *ABCA7*-tagged protein appeared stable and localized in the plasma membrane in transfected HEK cells. scRNAseq data confirmed the highest expression of *ABCA7* in neurons. ABCA7 mutant lines exhibited a normal neuronal phenotype. Transcriptome comparison of unedited control vs homozygous *ABCA7* deletion iPSC-derived neurons showed that the *ABCA7* transcripts were expressed at similar low levels in both lines. Proteomics analysis failed to detect ABCA7 protein expression. However, the lipidomics data showed that the homozygous *ABCA7* deleted neurons show a 2~5-fold increase in diaglycerol, phosphoinositol, phosphoethanolamine, cholesterol ester, and triacylglycerol compared to the isogenic control neurons. Conclusion: the ABCA7 deletion significantly alters lipid content within the carrier neurons. Importantly, altered lipid content is of increasing interest in studies of the mechanism of AD pathogenesis. Our data suggest that the African ancestry-specific deletion in ABCA7 generates a transcript that escapes NMD in neuronal cells. Further studies into why some individuals appear to escape this mutation's effect may highlight compensational mechanisms in lipid metabolism.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6088F: Unravelling the effect of multiple sclerosis risk loci on splicing using targeted single-cell long-read sequencing

Authors:

M. David[1], C. Wang[2], J. S. Jabbari[2], M. E. Ritchie[2], I. Cleynen[1], B. Dubois[1]; [1]KU Leuven, Leuven, Belgium, [2]The Walter and Eliza Hall Inst. of Med. Res., Parkville, Australia

Abstract:

Multiple sclerosis (MS) is a complex, autoimmune disease of the central nervous system and is the leading cause of non-traumatic disability in young adults. Thus far, no cure is available and current therapies can come with severe adverse effects as they are based on broad

immunodepletion. Large genome-wide association studies (GWAS) enabled the identification of 236 MS risk loci implicating both adaptive and innate immune cells in MS pathogenesis. Understanding the mechanism of action of these risk loci is a key challenge, likely to advance our understanding of the MS immunopathogenesis and the development of novel, more targeted therapies.
MS-associated variants are >90% non-coding but they display significant enrichment for variants affecting gene expression and/or splicing. Increasing evidence points to the cell specificity of these processes. However, approaches to study these have been limited to bulk RNA from broad immune cell subsets in blood. In an attempt to fill this important gap, we use a state-of-the-art multimodal single-cell method to investigate the effect of a set of 130 prioritized MS risk variants, on the splicing of 54 genes in *cis*. The investigated variants were selected based on their likelihood to impact splicing. Our unique approach to delve into these effects combines two layers of information: we overlay (1) the unbiased short-read transcriptomics landscape of thousands of cells with (2) a targeted long-read sequencing layer (scRaCH-seq, Thijssen *et al* (2022)), enabling concomitant cell clustering and annotation as well as capture of full-length transcripts for optimal detection of alternative splicing of our target genes, in a cost-effective way. This multi-omics technique is applied to paired blood-CSF samples from MS patients and non-inflammatory controls (n=8).
Using 3,016 biotinylated oligonucleotide probes, we were able to enrich and detect all 54 target genes as well as generate an average of 17,007,732.5 sequencing reads per sample. Preliminary analyses show a strong positive correlation (min tau = 0.66; max p = 3.86e-12) between short- and long-read counts, indicating that the targeted long-read method likely did not cause a technical bias in the observed read counts. Moreover, the long-read sequencing data enabled identification of new, unannotated gene isoforms. It also revealed cluster specificity of particular isoforms, which is concealed when working exclusively with short-read data that (mostly) provides gene-level information.
Additional results from differential transcript expression/usage analyses and evaluation of the effects of genotype on splicing will be presented during the 2024 ASHG annual meeting.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6089F: Using AllofUs to investigate genetic factors influencing the measles vaccine response

Authors:

A. Majd[1], S. Pfeifer[2], M. Martinez-Vargas[3], L. Arroyo-Cruz[4], A. Saravanan[5], U. Bhattarai[6], F. Abebe-Akele[7], O. Rodriguez[8]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Max F.

Perutz Lab., Vienna, Vienna, Austria, [3]Baylor Coll. of Med., Houston, TX, [4]Univ. Central Del Caribe, San Juan, Puerto Rico, [5]Northern Illinois Univ., DeKalb, IL, [6]Adams State Univ., Alamosa, CO, [7]Elizabeth City State Univ., Elizabeth City, NC, [8]Univ. of Louisville Sch. of Med., Woodside, NY

Abstract:

Understanding the effectiveness of vaccine responses is crucial for public health. The heritability of antibody responses to vaccines varies across different types, with the measles vaccine response having an estimated heritability rate of 89%. This vaccine has proven highly effective, significantly reducing the incidence of measles infections and related deaths. However, it is important to acknowledge that the vaccine may fail to induce sufficient protective antibodies in 2% to 10% of individuals. To better understand the genetic factors influencing the response to the measles vaccine, we performed a genome-wide association study (GWAS) across three immunoglobulin loci—heavy, kappa, and lambda—that harbor 136 to 171 antibody genes. We utilized the NIH All of Us cohort, which contains extensive genetic, clinical, and survey data from more than 250,000 individuals. We evaluated whether single-nucleotide polymorphisms (SNPs) were associated with measles antibody levels in 437 ancestry diverse individuals with measles antibody levels after vaccination. We identified a region within the lambda locus on chromosome 22 with significantly associated SNPs (p-value = 1.57e-06, adjusted for 18077 number of tested SNPs). Using the Ensembl Variant Effect Predictor, we found a missense variant in VPREB1 and SNPs upstream of the gene IGLV1-51 and the pseudogene IGLV1-50. VPREB1 is an intriguing association due to its role in B cell development and antibody regulation. These results demonstrate that genetic factors may influence an individual's response to vaccination and could provide an additional needed avenue for therapeutic development in low vaccine responders.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6090F: Using DOLPHIN protein domain information to classify human variations in rare diseases

Authors:

M. Corcuff, j-p. desvignes, C. Guien, G. Collod-Beroud, E. Ménoret, D. Salgado, C. Beroud; Aix Marseille Université, Marseille, France

Abstract:

Using the ACMG-AMP guidelines for interpreting sequence variants, it remains difficult to meet the protein domain criterion, PM1, which is only assigned in about 10% of cases, whereas the

variant frequency criterion, PM2/BA1/BS1, is reported in 50% of cases. To improve the classification of human missense variants, we developed the DOLPHIN system (https://dolphin.mmg-gbit.eu). Using Pfam eukaryote protein alignments, we generated a DOLPHIN score to identify protein domain residues and variants that have a significant effect. In addition, we derived a DOLPHIN global domain residue variation frequency from GnomAD by aligning all similar domains from all human proteins. Applying this tool to all potential human transcript variants included in dbNSFP, 30.0% were assigned a PM1 label and 33.2% qualified for a new benign support criterion, BP8. This criterion indicates that the protein domain variation was observed with high frequency in at least one identical domain of any human protein. The DOLPHIN extrapolated frequency allowed the annotation of 31.8% of the variants, compared to the original frequency available in GnomAD for only 7.6% of them. Overall, DOLPHIN allows a simplified application of the PM1 criterion, an extended application of the PM2/BS1 criteria and the creation of a new BP8 criterion. DOLPHIN facilitates the classification of amino acid substitutions in protein domains covering nearly 40% of proteins and representing the sites of most pathogenic variants.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6091F: Utilizing cortical organoids to understand cellular mechanisms in patients with the developmental delay-associated 16p12.1 deletion

Authors:

S. Noss, J. Sun, S. Girirajan; The Pennsylvania State Univ., University Park, PA

Abstract:

The 16p12.1 deletion is a rare copy number variant (CNV) of approximately 520 kb covering 7 genes on chromosome 16. Deletion carriers present with a wide range of neurodevelopmental disorders (NDDs), including autism spectrum disorder, schizophrenia, epilepsy, and developmental delay, as well as craniofacial abnormalities such as macrocephaly and microcephaly. With high phenotypic variability, the 16p12.1 deletion serves as an effective paradigm to understand how complex genetic architecture can result in varying cellular phenotypes, and ultimately different clinical outcomes for the patient. We used induced pluripotent stem cells (iPSCs) generated from 16p12.1 deletion carriers and their families, as well as a control line from an unrelated healthy donor with an isogenic CRISPR/Cas9-mediated deletion line. Our previous work using 2D cell culture revealed abnormal cell proliferation and apoptosis, as well as alterations in cell fate determination by performing cellular assays and RNA

sequencing on iPSC-derived neural cells. Here, to further understand the mechanisms underlying these cellular phenotypes, we utilized 3D suspension culture and dual SMAD inhibition to generate region-specific (cortical) organoids from the control and CRISPR/Cas9-mediated deletion lines. Cortical organoids provide an ideal platform to recapitulate early brain development, allowing us to investigate more details in cellular heterogeneity and mechanisms during development. Immunostaining performed on organoids to label genetic markers demonstrated enhanced maturation and over-expression of a ventral marker, NKX2.1, as well as an overexpression of GABAergic neuron marker, VGAT, in the CRISPR/Cas9-mediated deletion line, suggesting that the deletion can cause dysregulation in neuronal development-associated pathways. To characterize alterations in neuronal function caused by the deletion, multielectrode array (MEA) assays will be performed on both 2D cultured neurons and cortical organoids. Additionally, cortical organoids will be generated from probands with opposing head-size phenotypes (macrocephaly and microcephaly) to better understand the mechanisms and genetic etiology that contributes to this heterogeneity. Overall, our study provides insights into the cellular mechanisms that contribute to phenotypic heterogeneity in neurodevelopmental disorders.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6092F: Variability in Self-reported Race Among Individuals with Middle Eastern Inferred Genetic Ancestry in the UK Biobank, All of US and MGB Biobank

Authors:

Y. Sui[1,2], S. Koyama[2], M. Wang[3], P. T. Ellinor[1,2], P. Natarajan[1,2], F. S. Alkuraya[4], A. C. Fahed[1,2]; [1]Cardiovascular Res. Ctr., Dept. of Med., Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [2]Cardiovascular Disease Initiative, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]CAS Key Lab. of Genome Sci. and Information, Beijing Inst. of Genomics, Chinese Academy of Sci. and China Natl. Ctr. for Bioinformation, Beijing, China, [4]Dept. of Translational Genomics, Ctr. for Genomic Med., King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia

Abstract:

While many people living in the US and UK identify as Arab, “Middle Eastern” or “Arab” is not routinely included as an option in demographic data surveys collected in the context of clinical care and biobanks. This might miss an opportunity at improving health for this group, especially that it is already massively under-represented in genomic studies. Here, we used principal component analysis (PCA) combined with K-nearest-neighbor classification on Middle Eastern

individuals from the 1000 Genomes Project and the Human Genome Diversity Project (HGDP) to identify participants of the UK Biobank, All of Us, and Mass General Brigham (MGB) Biobank who have a Middle Eastern inferred genetic ancestry. We compared their PCA to an indigenous Arab cohort from Saudi Arabia and studied how self-reported race and ethnicity varied in each. In total, we identified 2,725 individuals who have Middle Eastern inferred genetic ancestry - including 1,161 of 488,120 (0.24%) UK Biobank, 942 of 245,394 (0.38%) All of Us, and 632 of 65,119 (0.97%) MGB Biobank participants. The PCA distribution of this group overlapped with that of a cohort of 6,416 indigenous Arabs from Saudi Arabia. There was a large discordance in self-reported race/ancestry and inferred genetic ancestry that varied by the self-report options provided. Out of the 1,161 UK Biobank participants with Middle Eastern inferred genetic ancestry, 584 (50.3%) selected “Other ethnic group”, 307 (26.4%) selected “Any other white background”, and 81 (7.0%) selected “British”. In the All of Us, “Middle Eastern or North African” was provided as an option and was selected by 698 of 942 (74.1%) participants with inferred Middle Eastern genetic ancestry. The remainder identified as “Black or African American” (7.4%), “White” (6.8%), and “More than one population” (5.5%). In the MGB Biobank, 496 of 632 (78.5%) participants selected “White” and 75 (11.9%) selected “Other”. In conclusion, we identify 2,725 participants in three US and UK biobanks with inferred genetic ancestry similarity to indigenous Arabs from Saudi Arabia, and report a large variability in self-reported race and ethnicity.

Session Title: Molecular Effects of Genetic Variation Poster Friday Session

Board 6093F: Versatility of variant libraries and deep mutational scanning in diagnostics and therapeutics for transmembrane proteins

Authors:

M. Gilbert[1], E. Keefer-Jacques[2], T. Jadhav[3], D. Antfolk[4], Q. Ming[4], N. Valente[1], G-W. Shaw[1], C. Sottolano[1], G. Matwijec[1], V. Luca[5], K. Loomes[1], R. Rajagopalan[2], T. Hayeck[3], N. Spinner[6]; [1]The Childrens Hosp. of Philadelphia, Philadelphia, PA, [2]Children s Hosp. of Philadelphia, Philadelphia, PA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]H. Lee Moffitt Cancer Ctr. & Res. Inst., Tampa, FL, [5]H. Lee Moffitt Cancer & Res. Inst., Tampa, FL, [6]The Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Technical advances in sequencing of the human genome have been astounding over the past decade, however our comprehension of the functional consequence of DNA variants has not keep pace. This disconnect is detrimental to the entire continuum of patient care, from

diagnostics to therapeutics. Deep mutational scanning (DMS) is revolutionizing our ability to study variant effects in disease genes. These high-throughput assays characterize libraries of variants using functional tests, creating prospective repositories of data. Once created, libraries can be used in diverse applications. We developed variant libraries for the disease gene Jagged1 (*JAG1*), an autosomal dominant cause of the multi-system disorder, Alagille syndrome (ALGS), and show how DMS applications tailored for membrane proteins can be designed to support both diagnostics and therapeutics. We developed a high-throughput membrane expression assay to measure the cell localization of the transmembrane protein, JAG1. To advance diagnostics, we used this assay and characterized 486 out of 2,832 JAG1 variants as functionally abnormal and translated our results into evidence weight for use during variant classification. Review of ClinVar and an internal ALGS cohort identified 27 VOUS with abnormal membrane expression data of which 26 could be reclassified as likely pathogenic. We then looked at how this assay could be used to develop therapeutics for ALGS. Protein misfolding and intracellular sequestration is a common pathomechanism for mutant transmembrane proteins and restored membrane expression has proven efficient for functional rescue of other misfolded proteins. We tested the ability of low temperature incubation to support membrane expression of mutant JAG1 and showed correction of three out of eight tested mutants (37.5%). These results reveal differences in variant properties and highlight the need for a personalized approach. Using our established JAG1 cell libraries and high-throughput membrane expression pipeline, we are now identifying missense variants that are amenable to cell trafficking correction by performing both a small molecule screen and optimizing an analytic strategy to identify responsive missense variants. Application of DMS assays in different contexts for the same disease gene can address different clinical concerns. Here, we highlight a single assay with utility in both reducing the VOUS burden and in identifying variants amenable to a therapeutic intervention for transmembrane proteins. It is estimated that 20-30% of all proteins are transmembrane, supporting expansion of these types of diverse DMS applications to ~5,000 other membrane proteins.

Session Title: Omics Technologies Poster Friday Session

Board 1100F: "CellmarkerAI", an artificial intelligence tool for Single-cell transcriptomics gene marker identification and annotation

Authors:

A. Ali[1,2], N. Nassir[2], A. Ahmed[2], A. Nazir[3], M. Uddin[2,4]; [1]Sheikh Khalifa Med. City, Abu Dhabi, United Arab Emirates, [2]Ctr. for Applied and Translational Genomics (CATG), Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, [3]Zayed Univ., Dept. of

Information Systems and Technology Management, Coll. of Technological Innovation, Abu Dhbai, United Arab Emirates, [4]GenomeArc Inc., Cellular Intelligence Lab, Mississauga, ON, Canada

Abstract:

Background/Objectives:Using single-cell transcriptomics, we can obtain crucial understanding of the development and differentiation of multicellular organisms at the cytological level. Nonetheless, the challenge arises in annotating cell types due to the absence of reliable marker genes. This study aims to create a method to identify markers within single-cell clusters and subclusters using Deep Learning and Machine Learning artificial intelligence tools.Methods:We developed CellMarkerAI, an Artificial intelligence tool that leverages deep learning neural networks and the machine learning tool Support Vector Machines (SVM) to annotate cell types from single cell transcriptome data. CellMarkerAI was utilized on experimental single nuclei mRNA sequencing data of Progressive Supranuclear Palsy (PSP) inflected brain tissue, along with publicly accessible datasets like single cell transcriptome data from the Developmental Single Cell Atlas of Gene Regulation and Expression. This enabled the identification of strong gene markers associated with various cell types.Results:We effectively classified and annotated cell types using a pre-trained CellMarkerAI, achieving classification accuracy of over 90% with minimal execution times. Utilizing CellMarkerAI on metanephric data from the fetal single cell transcriptome atlas, we detected known canonical gene markers of sub clusters such as *NPHS2* and *SLC12A1/NKCC2* in podocytes and the thick ascending limb of the loop of Henle, respectively. Likewise, CellMakerAI identified the canonical *GFAP* in astrocytes subclusters when used on the experimental PSP data. Furthermore, we uncovered less-known gene markers such as *OBI1-AS1* in astrocytes and *KIRREL3* in podocytes, which exhibited strong clustering within their respective cell types (P < 1.00 X 10-3). To unveil additional novel markers, we repeated this analysis across various cell types in different tissues.Conclusion:CellMarkerAI, employing a combination of Deep Learning and Machine learning, efficiently identifies gene markers with high performance on actual single-cell genomics data, excelling in sub-cluster annotation and the discovery of novel markers.

Session Title: Omics Technologies Poster Friday Session

Board 1101F: 2022 Trans-Antarctic Expedition: Microbiome and multiomic health tracking during a 2,150km unsupported kite-ski expedition across Antarctica, with robust control and twin pair participant comparison ⭐

Authors:

A. Brooks[1], M. Snyder[2], A. Celli[1], D. Thota[1], M. babu[1], S. Lancaster[1], L. McGuire[1], P. Maguire[1], D. Lipson[3], S. Pollock[4]; [1]Stanford Univ., Palo Alto, CA, [2]Stanford Dept. of Genetics, Palo Alto, CA, [3]Ultima Genomics, Newark, CA, [4]Ultima Genomics, Fremont, CA

Abstract:

Understanding the range of human variation across multiomic measures requires not only controlled subjects living regular lives, but also comparison to subjects under the most extreme pressures such as previous work in the NASA Twin-Study. In late 2021 and into 2022 two adventurers performed a 2,150km unsupported expedition across the continent of Antarctica from the Novolazarevskaya research base to the South Pole using only skis, sleds and kites for locomotion and equipment hauling without resupply. Blood, urine, stool, saliva, clinical, wearable and environmental measures for multiomic analysis were collected for five months prior to the expedition, roughly weekly for nine timepoints during the expedition, and for five months after the expedition for the two expedition subjects and two matched control subjects which includes a twin-pair and 110 total timepoints. Samples were processed using the same methodology as subjects from the Human Microbiome Project Phase II Diabetes cohort, on which we have performed novel unpublished metagenomics of 105 subjects for up to a decade providing a robust control cohort for comparison. This submission focuses on wearable, clinical, and primarily metagenomic profiling from these two studies, where we examined patterns in gut and oral microbiomes. Environmental measures reveal unique stresses of -37C average temperatures, average altitude above 3km, and wind speeds of >100km per hour during the expedition, which correlate with wearable measures of heart rate, calories burned, sleep and stress measures. Comparison with metagenomic data reveals significant shifts in microbiome composition and drops in diversity at the beginning of the expedition, but microbiome stability returning toward the end which parallels personal logs of negative/positive mood and increases in sleep stability and reductions in stress scores. Control subjects do not show such concerted shifts in the microbiome over such short periods, although longitudinal variation at certain timepoints is commonly observed in most subjects. Metabolomic samples from blood and urine are being processed and should be incorporated shortly. Finally, this dataset provides a unique if small comparison to robust longitudinal multiomic sampling, and an example of microbiome perturbation under some of the most extreme environmental conditions.

Session Title: Omics Technologies Poster Friday Session

Board 1102F: 39 unique proteins have differential responses associated with ancestry during severe COVID-19 infection

Authors:

T. Zheng[1], Y. Ilboudo[2], T. Lu[3], G. Butler-Laporte[4], T. Nakanishi[1], D. Morrison[2], D. Adra[2], B. Richards[5]; [1]McGill Univ., Montreal, QC, Canada, [2]Jewish Gen. Hosp., Montreal, QC, Canada, [3]Univ. of Toronto, Toronto, ON, Canada, [4]Univ. of Oxford, Montreal, QC, Canada, [5]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada

Abstract:

Background: COVID-19 led to a disruption in nearly all aspects of society, yet these impacts were not the same across populations. During the pandemic, it became apparent that ancestry was associated with COVID-19 severity and morbidity, such that individuals with African ancestry tended to have worse outcomes than other ancestries. One factor that may influence COVID-19 outcomes is the circulating proteomic response to infection. This study aims to examine how different ancestries had differential circulating protein levels in response to severe COVID-19 infection. **Methods:** 4,979 circulating proteins from 1,272 samples were measured using the SomaScan platform. We used a linear mixed model to assess the ancestry-specific association between the level of each protein and severe COVID-19 illness, accounting for sex, age, and days since symptom onset. We then compared each ancestry-specific effect size of severe COVID-19 illness on protein level to one another in a pairwise manner to generate Z-scores. These Z-scores were then converted into p-values and corrected for multiple comparisons using a Benjamini-Hochberg false discovery rate of 5%. **Results:** Comparing ancestries, we found that 62% of the tested proteins are associated with severe COVID-19 in European-ancestry individuals, compared to controls. We found that 45% and 22% of the tested proteins were different between COVID-19 infected and control individuals in people of African and East Asian ancestry, respectively. There was a strong correlation in effect size between ancestries. We found that individuals of European and African ancestry had the most similar response with a Pearson correlation of 0.868, 95% CI [0.861, 0.875] while European and East Asian ancestries had a Pearson correlation of 0.645, [0.628, 0.661] and, East Asian and African ancestries had a Pearson cohort of 0.71, [0.695, 0.722]. However, we found 39 unique proteins that responded differently (FDR < 0.05) between the three ancestries. **Conclusions:** Examining 4,979 protein levels in 1,272 samples, we identified that the vast majority of proteins had similar responses to infection across individuals of European, African and East Asian ancestry. However, there were 39 proteins that may have a differential response to infection, when

stratified by ancestry. These proteins could be investigated to assess whether they explain the differences in observed severity of COVID between ancestral populations.

Session Title: Omics Technologies Poster Friday Session

Board 1103F: A comprehensive comparison on clustering methods for multi-slices spatially resolved transcriptomics data

Authors:

M. Zhou, C. Xiong, J. Chen, S. Huang, Y. Zhang, W. Wu, Y. Li; UNC Chapel Hill, Chapel Hill, NC

Abstract:

Spatial transcriptomics (ST) enhances gene expression analysis within tissues by providing spatial information. Spatial domain detection, also known as clustering, is a crucial task in ST data analysis. Clustering facilitates the identification of similar expression patterns across various tissue regions, revealing spatial distribution of different cell types and states valuable for understanding biological mechanisms (e.g., complex tumor microenvironment). Earlier ST clustering methods were designed for single tissue sections. The emergence of multiple contiguous tissue sections derived from the same or similar tissue specimens within or across individuals has led to the development of multi-slice clustering methods. In this study, we evaluated the performance of ten state-of-art ST clustering methods, including seven single-slice methods and three multi-slice methods, on two simulated and three real datasets. Specifically, we investigated the effectiveness of pre-processing techniques, including spatial coordinate alignment (for example, PASTE) and gene expression batch effect removal (for example, Harmony). Most methods achieved great performance in single-slice clustering. MAPLE performed well on simulated datasets. BASS, graphST, and STAGATE showed better performance in multi-slice clustering on real datasets. Batch effect correction and coordinate alignment show mixed results, benefiting certain methods and datasets while impairing some others. In summary, our study provides a comprehensive comparison of clustering methods for multi-slices ST data, serving as a reference for future method development and providing practical guidelines for methods selection in various scenarios.

Session Title: Omics Technologies Poster Friday Session

Board 1104F: A computational approach to capturing infant gut microbiome and metabolic changes with diet

Authors:

A. Price, S. Rasolofomanana-Rajery, D. Frank, N. F. Krebs, M. Tang, A. Krishnan, A. Hendricks; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

While studies have explored differences in gut microbiome development in response to infant liquid diets (breastmilk, formula), little is known about the impact of complementary foods on the infant gut microbiome. As infants transition from the liquid diet to complementary foods, gut microbial communities change in composition and functional capacity. Additionally, this transition is a key time for establishing anthropometric trajectories and influencing future cardiometabolic disease risk potentially mediated through the microbiome. Thus, elucidating the impact of complementary foods on the gut microbiome during this transition is critical.
We investigated how two common protein-rich foods - meat and dairy - affect the gut microbiome and metabolome during early complementary feeding in U.S. formula-fed infants using untargeted metagenomic sequencing and metabolomics profiling on longitudinal stool collections during a randomized controlled feeding trial from 5 to 12 months. Understanding the landscape of microbes and their metabolites here is challenging as it changes across both diet type and time and is often not well represented by differential abundance analysis. To better capture high-dimensional shifts, we developed a novel computational approach to model the time-dependent, complex interactions as a network with node embedding within a common latent space containing 11,549 microbe features, 7,896 metabolite features, and samples from 109 infants that consumed either a meat-enriched or dairy-enriched diet.
We used regularized regression on the embeddings to classify diet (meat or dairy) and compared their performance to classifiers built on the original microbe and metabolite feature data. We find classifiers built with embeddings improve diet prediction performance by an average of 15% compared to classifiers built with the original data (mean F1 score=81% and 65%, respectively). Next, we used the embedding-based classifiers to prioritize microbe and metabolite features into diet groups. We found meat groups had a higher abundance of *Lachnospiraceae*, bacteria previously found to be associated with increased protein intake in infants- while diary groups had a higher abundance of *Lactobacillaceae*, bacteria previously associated with milk-based feeding in infants. Lastly, we used the latent space to identify both known and novel microbe-metabolite pairs based on their closeness to sample clusters. By modeling the microbe-metabolite-sample environment with node embeddings, we identify previously undetectable infant microbe-diet associations and generate new hypotheses around microbe-metabolite interactions.

Session Title: Omics Technologies Poster Friday Session

Board 1105F: A heterogeneous differentiating cell culture system for studying gene regulatory dynamics in endoderm-derived cell types

Authors:

A. Cormack, K. Barr, Y. Gilad; Univ. of Chicago, Chicago, IL

Abstract:

Characterizing the effects of gene regulatory variants is useful for interpreting genetic associations with complex diseases, particularly when gene regulation is measured in relevant cell types and contexts. While human iPSCs can be differentiated into a variety of disease-relevant cell types, their utility for population-level studies of gene regulation is hindered by laborious, time-intensive methodologies that are often difficult to reproduce across multiple individuals.

We recently developed an iPSC-derived heterogeneous differentiating cell culture (HDC) system that, when combined with single-cell RNA-sequencing, allows us to efficiently study gene regulatory variation in dozens of cell types from multiple individuals and across different exposure or treatment conditions. Moreover, cells within HDCs spontaneously and asynchronously differentiate into cell types from all three germ layers, allowing us to capture transient, developmentally dynamic effects on gene regulation. HDCs have greatly expanded the breadth of biological contexts that can be queried in population-level studies of gene regulation. Still, for studies that focus on a particular tissue, lineage, or disease, a model enriched for the cells of interest may be preferable to one that necessitates sequencing dozens of irrelevant cell types. We therefore sought to develop HDCs that are enriched for cells from each of the three germ layers.

To generate HDCs that are enriched for endodermal cell types, we adapted a conventional hepatocyte differentiation protocol such that iPSCs are guided toward the liver trajectory while still permitting differentiation into other endodermal cell types. In this pilot study, we aimed to optimize our guided differentiation approach and characterize the cell types originating from endodermal HDCs. and evaluate their fidelity in replicating in vivo gene expression profiles, and discern biological and technical contributions to gene expression variability. Using 3 iPSC lines (3 replicates each), we generated endodermal HDCs and collected single-cell RNA-sequencing data on days 14 and 18 of cell culture. We targeted 5K cells per individual and 20K reads per cell, for a total of 30K single cells. These data will reveal the cell type composition of endodermal HDCs at each time point and allow us to evaluate their fidelity in replicating in vivo gene expression profiles. In future work we will leverage this system to characterize gene regulatory variation in

cell types representing the liver and other endoderm-derived tissues, which may ultimately be useful for exploring tissue- and individual-specific responses to drugs and other exposures.

Session Title: Omics Technologies Poster Friday Session

Board 1106F: A Novel Bioinformatic Tool for Large-Scale Primer Design and Specificity Analysis

Authors:

J. Pitsch[1], S. Wirth[1], X. Yang[2], M. Breuss[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Utah, Salt Lake City, UT

Abstract:

Since its inception in 1983, polymerase chain reaction (PCR) has become ubiquitous in biological research labs. It is a fast, flexible, cost-effective technique that is easy to perform. Despite the many benefits of PCR, there remains a prevalent challenge when performing large-scale and parallelized PCR experiments: manual primer design can be an error-prone and time-consuming process depending on the number and composition of target sites. As a solution, Primer3 has emerged as an accessible tool that offers a graphical user interface, as well as command line capabilities. While the latter enables parallelized primer design, additional pipelines are required for appropriate scaling. Moreover, this does not replace manual confirmation of primer specificity (i.e., the assessment of off-targets). To overcome the challenge associated with large-scale primer design, we fused the functionality of Primer3 and a tool to assess off-targets, In-Silico PCR (ISPCR), into a single integrated pipeline that performs primer design and specificity analysis for any given number of target sites at scale. We provide this utility through an integrated Python script that reads a human-readable position list and returns appropriate primer pairs and off-target information. For our test application, we employed our pipeline for targeted amplicon sequencing on a 150bp paired-end Illumina platform. To increase the likelihood of finding an acceptable primer pair for each target site, we tasked Primer3 with designing three alternative amplicons with varying amplicon sizes including our ideal amplicon size of 190bp. We then determined the ISPCR settings needed to identify problematic off-target matches and developed an analysis script to further refine the results from ISPCR. The final output of our tool is an appended version of the input file that contains the lead forward and reverse primer pair for each target site, a measure of their likelihood to bind to off-targets, and a visual representation of matches and mismatches for primer alignment. Initial testing of our tool has shown that the designed primer pairs are compatible with our targeted amplicon sequencing (TAS) application with minimal non-specific binding. Our future goals for this

bioinformatic tool include additional testing of output primer pairs, optimizing its runtime, and developing a web interface for open access.

Session Title: Omics Technologies Poster Friday Session

Board 1107F: A novel NGS-based sample tracking system for robust sample chain of custody and quality control in population-scale genomics studies

Authors:

K. Milligan, A. O'Hara, C. Ziemniak, C. Goswami, W. Wang, D. Corney, H. Latif; GENEWIZ by Azenta Life Sci., South Plainfield, NJ

Abstract:

When processing large cohorts of samples, sample identity and functional quality are important characteristics to ensure the derived data is accurate and attributed to the appropriate sample. This is especially critical in the era of population-scale genomics where a sample swap or poor sample quality may result in incorrect data or delays due to re-processing. While traditional methods such as qPCR can be used for functional quality control, they may not be an effective predictor of sample success in Next-Generation Sequencing (NGS) based assays due to differential susceptibility to inhibitors and interfering substances. Performing QC using the same methodology for data generation allows for true functional analysis for interfering substances, poor sample quality and/or sample degradation. We have created a scalable, cost-effective sample tracking and enhanced quality control assay that may be incorporated prior to performing Whole Exome Sequencing (WES) or Whole Genome Sequencing (WGS). A range of different sample types can be processed on the amplicon-based sequencing platform, targeting regions in the exome in addition to sex-based targets to allow tracking of sample identity. This amplicon panel was designed to target autosomal SNPs with varying minor allele frequencies in different ethnic populations, ensuring that there is no bias towards variation in European ethnic populations. The average minor allele frequency for the panel developed for the general population is 0.319, 0.317 for European background, 0.261 for African background and 0.246 for Asian background based on data from ALFA and gnomAD. Sex determination targets were selected for SNPs on the X chromosome and short unique fragments on the Y chromosome. Targets on the Y chromosome were selected to be less than 150 base pairs to reduce the effects of sample degradation and increase the likelihood of successful identification. This assay was designed and assessed to confirm consistency in identity for samples processed between WGS and WES platforms, and samples submitted for the same subject across different cohorts. In

addition to identifying poor-performing samples, as low as 10% contamination levels can be detected in contrived samples. A strong quality control assay enables NGS laboratories to identify poor sample quality or degradation prior to processing, allowing for re-extraction or new sample collection before continuing on to processing on a more complex or expensive assay, ultimately saving time and budget. This newly developed sample tracking assay ensures both sample integrity and identity, leading to an overall increase in data quality for research studies.

Session Title: Omics Technologies Poster Friday Session

Board 1108F: A web-based application for integrative and comparative analysis of multi-omic data

Authors:

S. Wy, D. Kwon, J. Ahn, J. Kim, J. Oh, J. Kim; Konkuk Univ., Seoul, Korea, Republic of

Abstract:

The accumulation of high-quality genome assemblies has facilitated a more accurate comparison of genomes among multiple species. Furthermore, the availability of various omic data has further extended the scope of such comparative studies to identify the consequences of multi-omic signatures and underlying mechanisms. When performing such comparative multi-omic analyses, the genome-wide comparison of multi-omic data and the visualization of the results are critical. In addition, the visualization needs to be efficient enough to handle a large volume of multi-omic data. However, there is still a lack of applications that fulfill such requirements. In this study, we developed a web-based application for comparative multi-omic analyses. In our application, users can easily (i) compare multiple genomes, (ii) compare the profiles of various omic data, and (iii) perform integrative analyses for finding genomic regions showing interesting multi-omic signatures. Users can perform all these analyses in an easy-to-use web interface. The application works very efficiently because of intelligent indexing and multi-level data sampling. Our application will contribute to making the use of multi-omic data easier and more effective.

Session Title: Omics Technologies Poster Friday Session

Board 1109F: Advancing ACMG/AMP variant classification through computational facial analysis for variants of unknown significance

Authors:

T-C. Hsieh[1], H. Klinkhammer[1], H. Lesmann[2], M. ten Hagen[1], H. Weiland[1], S. Kaptain[1], P. Krawitz[3]; [1]Univ. Hosp. of Bonn, Bonn, Germany, [2]Inst. of Human Genetics, Univ. Bonn, Cologne, Germany, [3]Univ. Bonn, Bonn, Germany

Abstract:

Background: Streamlining the classification for variants of unknown significance (VUS) is vital for patients with undiagnosed conditions. While next-generation phenotyping techniques like GestaltMatcher have significantly contributed to diagnosing rare disorders, the lack of computational facial analysis support in current ACMG/AMP guidelines has prompted this study to propose its inclusion. This study introduces new criteria—PP4_SUP, PP4_MOD, and PP4_STR—for supporting, moderate, and strong levels of pathogenicity. These criteria are integrated with the existing PP4 criteria using GestaltMatcher within a Bayesian framework, aiming to aid in the reclassification of VUS.
Methods: Within the Bayesian framework of the ACMG/AMP guidelines, we utilized GestaltMatcher scores to evaluate phenotypic match across various syndromes, calculating the associated likelihood ratios (LR). GestaltMatcher was initially trained on 8,547 images representing 244 disorders, encoding each image into a facial phenotype descriptor. The gestalt score for each gene was quantified by the cosine distance between the most similar image of each gene and the testing image. For each image, gestalt scores were computed across all disorders. Validation was performed on 1,499 images of 321 syndromes not included in the training set. We conducted five-fold cross-validation to establish thresholds consistent with the Bayesian framework.
Results: For each gene, we established thresholds corresponding to supporting, moderate, and strong levels of evidence. For example, for the gene *NIPBL*, the thresholds for PP4_STR, PP4_MOD, and PP4_SUP were determined to be 0.64, 0.58, and 0.54, respectively. Among the 67 test patients, 57 achieved PP4_STR, 6 achieved PP4_MOD, and 4 achieved PP4_SUP. In contrast, for the gene *TAF1*, which is associated with a less distinct facial phenotype, we obtained relatively lower thresholds. In testing 32 patients, 2 achieved PP4_STR, 9 achieved PP4_MOD, and 16 achieved PP4_SUP. Additionally, facial analysis reclassified three out of eight *TAF1* VUS and one *NIPBL* VUS as likely pathogenic. These classification criteria are available through the GestaltMatcher API and can be integrated into any variant prioritization platform.
Conclusion: We demonstrated that incorporating GestaltMatcher into PP4_SUP, PP4_MOD, and PP4_STR criteria in ACMG/AMP guidelines enhances VUS classification.

Session Title: Omics Technologies Poster Friday Session

Board 1110F: CasNovo: De-novo Cas-9 protein engineering with transformer large language models

Authors:

S. Asaigoli[1], S. Asaigoli[2]; [1]UNC Sch. of Med., Chapel Hill, NC, [2]William G. Enloe High Sch., Raleigh, NC

Abstract:

Designing Cas9 proteins that possess the stability and specificity required for CRISPR-Cas9 poses a tedious and resource-expensive challenge for genetic medicine development. Sub-par fidelity of Cas9 proteins results in off-target effects, compromising the safety of gene therapies. Conventional laboratory protein engineering has improved, however, the rise in computational design of endonuclease proteins and peptide-based binders with machine learning has presented a more efficient avenue for developing novel therapeutics using protein-based large language models. Methods shown in ProtBERT, PepMLM, and ProLLaMA show transformer-based language model architectures for the process of generating de-novo proteins and predicting protein structure and function with high novelty, stability, and low perplexity. In this work, we present CasNovo, a fine-tuned tool for the generation and validation of de-novo Cas9 proteins. CasNovo leverages ProtGPT2, a transformer large language model trained on around 50 million naturally occurring amino acid sequences. The base algorithm was then trained on a dataset of over 100,000 Cas9 proteins curated from the UniProt Database. The fine-tuned architecture learned protein sequence patterns and motifs to generate 300 proteins novel proteins in-silico. Generated sequences were pushed through a deep-learning validation pipeline employing AlphaFold and ESMFold, commonly used to determine protein structure, stability, and perplexity in transformer-based large language model tools. Validated proteins were assessed by the predicted local difference test (pLDDT), post-translational modification score (PTM), sequence coverage, and tertiary structural visualizations. The Cas9 proteins from the UniProt dataset contained pLDDT scores between 60 and 80. CasNovo's fine-tuned algorithm has presented an unprecedented mean pLDDT score of 76.6 with further validation of the generated proteins pending, a close to two-fold increase from the base generation of ProtGPT2. These results indicate the promise of the Cas9 generation pipeline. Future efforts will concentrate on extensive validation of the generated proteins studying their functional application in vitro for therapeutics and genetic medicine. With the multi-task capability of conducting protein generation and validation, CasNovo has the potential to accelerate Cas9 protein and CRISPR/Cas9 therapeutic development significantly.

Session Title: Omics Technologies Poster Friday Session

Board 1111F: Characterization of the rat osteosarcoma UMR-106 using long read methods identifies shared patterns of amplification, mutation, methylation and expression with human disease

Authors:

D. Mohr[1], A. F. Scott[2], W. Littrell[1], M. Kokosinski[3], Y. Zou[1], V. Stinnett[4], J. Madhiwala[1], J. Anderson[1], K. Gabrielson[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Sch. of Med., Baltimore, MD, [3]Johns Hopkins Univ, Baltimore, MD, [4]Johns Hopkins Hosp., Baltimore, MD

Abstract:

The rat osteosarcoma cell line UMR-106 is widely used for the study of bone cancer biology. We have characterized this cell line using a combination of optical genome mapping (OGM), nanopore long-read sequencing and short-read RNAseq. The genome sequence and methylation patterns were compared to a Sprague-Dawley control animal, the strain from which UMR-106 was derived and expression data were compared to a public rat osteoblast dataset. Nanopore sequencing identified UMR-106 specific changes in copy number, structural variants, likely oncogenic mutations in Tp53 and Myc as well as methylation differences in the tumor suppressor CDKN2A each of which have been reported in the COSMIC database for genes most affected in human osteosarcomas. OGM found numerous translocations and other genomic events characteristic of widespread chromothripsis. The combined optical mapping and nanopore sequencing identified a region of approximately 30 Mb with about 5-7 X amplification near Myc on rat chromosome 7. Many genes in this region also showed increased expression vs. that reported in osteoblasts. Several of these have been cited as biomarkers in human osteosarcoma along with Myc, including Mdm2, Angpt1, TRps1, Eif3h, Deptor, Has2, Fam91a1, Tmem65, and Rnf139. The observation that so many loci from the amplified region are seen in human osteosarcoma suggests that this may be more than coincidental. This dataset illustrates the value and complementarity of both long DNA methods, optical mapping and sequencing, for the characterization of cell lines. The study also illustrates how inter-species analyses can inform about the genetic events that underpin specific tumor types. The data should be a valuable resource for investigators studying osteosarcoma and, specifically, the UMR-106 model of this disease.

Session Title: Omics Technologies Poster Friday Session

Board 1112F: Classification of continental ancestry in the Million Veteran Program from high-throughput proteomics data

Authors:

C. Yandava[1], P. Schreiner[2], D. Panyard[3], P. Devineni[1], B. Gorman[2], P. Tsao[4], S. Pyarajan[1]; [1]VA boston, Boston, MA, [2]Booz Allen, Boston, MA, [3]Stanford Univ., Stanford, CA, [4]VA Palo Alto, Palo Alto, CA

Abstract:

Proteins have been measured using methods such as mass spectrometry, enzyme-linked immunosorbent assay (ELISA), and multiplex assay based on immunobeads. Two methods—Somascan, based on single-strand modified aptamers, and Olink, a Proximity Extension Assay (PEA) based on antibody binding —simultaneously measure thousands of proteins with high specificity and sensitivity. As a pilot study, we measured the plasma protein levels of 875 individuals in the VA Million Veteran Program (MVP) comprising both a control group and cases of abdominal aortic aneurysm (AAA), Alzheimer's Disease, or lung cancer using both Somalogic and Olink platforms. As both broad and fine-grained ancestry can be reliably inferred from genetic information, we hypothesized that ancestry could also be inferred from high-dimensional proteomics data. The pilot study included 619 European (EUR), 251 African American (AFR), and 5 Hispanic (HIS) ancestry individuals, as classified from genotype data using the Harmonized Ancestry and Race/Ethnicity (HARE) method. As there were only 5 samples from HIS group, we focused on only EUR and AFR groups in subsequent analysis. We evaluated three different classifiers including a neural net classification with input, hidden, and output layers, Random Forest (RF), and Support Vector Machine (SVM), machine learning algorithms. The populations were divided into 70% training and 30% testing datasets. Models were developed using the training dataset. Predictions of the ethnic groups were performed with models on test dataset. The performance of the classification was evaluated on the predictions with Area Under Curve (AUC) and classification Accuracy, Sensitivity, and Specificity metrics. The AUC was over 0.98 for all the classification methods for Olink and over 0.987 for Somaloigc when all the proteins were used. Similarly, Sensitivity and Specificity were also evaluated. The Sensitivity was between 0.908 and 1 for Somalogic and between 0.716 and 0.96 for Olink. The Specificity between 0.95 and 1 for Somalogic and 1 for Olink for all algorithms. The Accuracy was ranged between 0.918 and 0.988 for Olink and 0.966 to 0.992 for Somalogic. A measure of variable importance for estimating a feature is by leveraging Gini Importance by calculating Mean Decrease in Gini (MDG). The minimum number (13) of proteins required for the classification was obtained for both platforms. Three proteins of these proteins belonged to the chemokine super family. In conclusion, we used the protein measurements from 2 platforms independently to classify HARE ancestry (EUR and AFR) in MVP.

Session Title: Omics Technologies Poster Friday Session

Board 1113F: Cohort-based Integrated Multiomics from Single Sample Blood Draws Provides Comprehensive Holistic Phenotypic Insights

Authors:

A. O'Hara[1], C. Tom[2], E. Stancliffe[2], A. Richardson[2], D. Corney[3], L. Turner[1], C. Mozdzierz[1], H. Latif[1], G. Zhou[1]; [1]GENEWIZ From Azenta Life Sci., South Plainfield, NJ, [2]Panome Bio, St Louis, MO, [3]GENEWIZ From Azenta Life Sci., South Planfield, NJ

Abstract:

The omics era has greatly expanded the repertoire of approaches available for researchers and clinicians to unravel the complexity underpinning human health: Next Generation Sequencing (NGS) and Mass Spectrometry (MS) approaches can characterize genomes, epigenomes, transcriptomes, proteomes and metabolomes. Advanced DNA barcoding and automated microfluidics can take this to the next level, enabling multiomic characterization of single cells. Peripheral blood mononuclear cells (PBMCs) offer a non-invasive approach that, when combined with omics tools, can provide a near holistic view of immune processes across patient cohorts. Here we detail a proof-of-principle workflow which utilizes a single blood draw to rapidly produce a diverse set of multiomics results including genomics, epigenomics, transcriptomics, proteomics and metabolomics. Blood draws using heparin tubes were collected and processed within 24 hours of the primary blood draw to ensure high viability and yield of PBMCs, along with simultaneous plasma separation and collection. These PBMC and plasma samples are then processed such that whole exome sequencing, whole genome methylation sequencing, single cell RNA sequencing, proteomic analysis and metabolomic analysis can all be collected from patient blood draw. Integrated analysis across multiple data modalities allows for holistic views of pathways and processes that are highly impacted. While genomics, transcriptomics, and proteomics provide information about genetic and functional potential, inclusion of metabolomics grants deeper phenotypic insights within and across individuals in the patient cohort. All these datatypes can be produced within days of primary sample collection using minimal sample amounts from a single blood draw, to address the biological questions at hand. With increased accessibility of omics approaches, integrative workflows such as the one described continue to gain broader adoption and drive greater insights and innovation in human health applications.

Board 1114F: Comprehensive analysis of Cas9 off-target effects using Ultima Genomics deep whole genome sequencing

Authors:

Z. Shipony[1], G. Lithwick-Yanai[1], S. Pollock[1], D. Bogumil[1], K. Kinney[2], R. Turk[2], G. Rettig[2], A. Jaimovich[1], M. Jarosz[1], D. Lipson[1]; [1]Ultima Genomics, Fremont, CA, [2]Integrated DNA Technologies, Coralville, IA

Abstract:

CRISPR-Cas9 is a groundbreaking tool in genome editing and is becoming indispensable both for research and clinical uses. Cas9 off-target effects remain a major challenge in the use of Cas9. Current methods for the discovery and characterization of off-targets such as cut-site enrichment-based methods and PCR-based enrichment provide limited insights into the distribution and type of the off-target event. These methods often fail to detect complex genomic alterations like translocations.

Using HEK293-S.p Cas9 (Streptococcus pyogenes Cas9, ATCC) cell lines, we delivered 3 different synthetic sgRNA guides targeting AR, AAVS1 and LAG3 genomic locations. We then generated deep whole genome sequencing (250-1000X) using the UG100TM sequencer for de novo analysis of on/off-target sites as well as high depth targeted amplicon sequencing of the IDT rhAmpSeqTM panel for a list of known on/off-target sites for each of the sgRNA guides at 5,000X.

Off-target sites de novo discovery using deep WGS demonstrated a high correlation with the existing list of rhAmpSeqTM-validated off-target sites although the levels of editing prevalence varied between the targeted panel and WGS. The total number of rhAmpSeqTM validated sites was AR: 207 sites; AAVS1: 186 sites; and LAG3: 233 sites. In almost all measured sites, deep WGS demonstrated a lower prevalence of editing compared to the targeted panel though WGS detection of Cas9 off-target sites showed high correlation with rhAmpSeqTM detected sites. Deep WGS also revealed the many different flavors in which Cas9 double strand breaks edit the genome and the frequency in which the different types appear at each cut site. In most on/off-target sites the most common effect is short deletions (<20 bases). In addition, WGS revealed that a high fraction of the genomic alterations were translocation events, and that translocations tend to be between Cas9 on/off-target sites representing new genomic rearrangements.

Deep WGS provides a comprehensive and sensitive method for detecting both on-target and off-target effects of Cas9, revealing a broader spectrum of genome edits including indels and translocations. This approach addresses the limitations of current methods, offering a complementary assessment of CRISPR-Cas9 editing specificity and efficiency, and enabling de novo detection. Future increases in sequencing depth and algorithmic methods will continue to

improve our understanding and management of off-target effects, ensuring safer and more effective genome editing applications.

Session Title: Omics Technologies Poster Friday Session

Board 1115F: CRISPR-BE Screen Design Server: a free web-based service for designing accurate sgRNA libraries for CRISPR base editing screens.

Authors:

V. Chapdelaine[1,2], S. Schoenherr[3], L. Forer[3], D. Taliun[1,2], R. Cuella-Martin[1,2]; [1]McGill Univ., Montreal, QC, Canada, [2]Victor Phillip Dahdaleh Inst. of Genomic Med. at McGill Univ., Montreal, QC, Canada, [3]Inst. of Genetic Epidemiology, Med. Univ. of Innsbruck, Innsbruck, Austria

Abstract:

CRISPR-based gene editing technology offers powerful methods to elucidate the genetic mechanisms of human disease. Among the numerous CRISPR-based technologies, base editing (BE) enables the modelling and correcting of genetic mutations at a single base pair resolution, i.e. point mutations, the largest class of human genetic variations. The CRISPR-based BE functional genomics screens, where point mutations are queried *en masse*, were proven a powerful way to characterize both mutations and genes. However, the lack of user-friendly computational tools for designing and running CRISPR-based BE screens hinders broader technology adoption. In a typical deep mutagenesis screen using CRISPR-based BE, one targets a list of genomic features of interest (e.g., those known to be implicated in a disease or phenotype). First, one needs to design a library of small guide RNA (sgRNA) to mutagenize across the genomic region of interest while minimizing the off-target modifications. In this step, utilizing the latest possible versions of genome references and genomic feature coordinates is crucial. Next, one must choose a set of genes for positive and negative controls, selecting sgRNA with predicted deleterious and neutral behaviour, respectively. For instance, in a viability screen, sgRNA leading to stop-gain mutations in known essential genes and mutations in neutral regions can act as positive and negative controls, respectively. Access to the latest gene and variant functional annotations is also essential in this step. When designing CRISPR-based BE screens with dozens of genes, these steps can quickly become unmanageable and error-prone without applying significant bioinformatics and computational skills. Mistakes may lead to spurious screening results because of unintended mutations, incomplete coverage, and other artifacts. Here, we introduce a free automated CRISPR-BE Screen Design Server that streamlines the creation of sgRNA libraries for CRISPR-based BE screens. CRISPR-BE Screen Design Server is based on the Cloudgene framework and has a user-friendly interactive web interface where

researchers can upload genomic features (e.g., genes) for their experiments, choose recommended settings or adjust parameters in the expert mode. The server will run data quality checks, report any inconsistencies, and start the automated library design. Once completed, it will provide downloadable results and interactive visualizations of all relevant quality metrics (e.g., predicted efficacy scores). CRISPR-BE Screen Design Server democratizes access to BE screening technologies while adhering to best practices and guaranteeing reproducibility.

Session Title: Omics Technologies Poster Friday Session

Board 1116F: Cross platform comparison of HLA typing from whole-genome sequencing

Authors:

C. Cremin[1], A. Saad[1], A. El-Khani[1], A. Attia[1], C. Minas[1], F. Aldhuhoori[1], G. Katagi[1], H. Wu[1], H. Sidahmed[1], J. Quilez[1], J. Mafofo[1], O. Soliman[1], S. Elavalli[1], S. Behl[1], T. Cardoso[1], T. Magalhaes[1], F. Sedlazeck[2], L. Paulin[2]; [1]M42, Abu Dhabi, United Arab Emirates, [2]Baylor Coll. Med., Houston, TX

Abstract:

The HLA region is important medically, with genetic variations affecting transplant outcomes, disease susceptibility, and drug responses. HLA genes have highly diverse alleles, posing challenges for accurate typing. Current practices mostly use targeted sequencing of major HLA genes, potentially overlooking others due to technical limitations. Whole-genome sequencing (WGS) offers unbiased coverage and could improve HLA typing accuracy. However, studies on WGS accuracy for HLA typing are limited, mainly focusing on Illumina short-read sequencing (SRS). WGS could enable opportunistic HLA screening using existing genomic data from population sequencing programs. Here we assessed the accuracy of comprehensive HLA typing (13 genes called with the HLA*LA software) from WGS data and its consistency across three sequencing platforms: two SRS (Illumina and MGI) and the long-read Oxford Nanopore Technologies (ONT) R9 and R10 chemistries. In each sequencing technology, we performed ~30X WGS of 24 Coriell reference samples. First, we benchmarked HLA*LA by comparing its HLA calls for each Illumina WGS sample against those produced by the leading Illumina DRAGEN software. We observed a mean concordance of 99.3% across the 3 genes reported by both tools. We then assessed the end-to-end (WGS and analysis) HLA typing consistency across platforms. For the 13 HLA genes called by HLA*LA, we detected an overall 98.5% concordance across platforms. Moreover, we independently validated our findings using the Simons Genome Diversity Project public datasets. We next studied the effectiveness of cheaper and less invasive HLA typing from WGS. First, in 5 Coriell samples we re-called HLA alleles from decreasing WGS

coverages (5-25X down-sampling) and measured accuracy relative to the ~30X dataset. Consistently across platforms we found that HLA typing remains >90% accurate even at 15X. Second, we compared HLA calls of 7 individuals from whom DNA was extracted from blood, buccal swab and saliva and each followed WGS in Illumina, MGI and ONT R9. While saliva showed weaker performance in reproducing HLA alleles compared to other sample types (Mean accuracy, 99.5% (Blood), 97.2% (Buccal) and 90.3% (Saliva)), overall, HLA outputs exhibit strong concordance regardless of sample type or platform used (Overall calling accuracy: 95.6%) Our study demonstrates the stability of HLA calling across platforms, sample types, and coverage thresholds, overcoming variation effects from platform design and biological quality differences. In conclusion, we advocate for the incorporation of WGS data into broader population studies on HLA diversity, irrespective of platform or sample specifics.

Session Title: Omics Technologies Poster Friday Session

Board 1117F: Cross-Ancestry Comparison of Aptamer and Antibody Proteomics Measures

Authors:

J. Nicholas[1], D. H. Katz[2], U. A. Tahir[3], C. L. Debban[4], F. Aguet[5], R. P. Bowler[6], J. Chen[7], C. B. Clish[5], J. Coresh[8], E. Cornell[9], R. Deo[10], P. Durda[9], X. Guo[11], C. Johnson[12], L. A. Lange[13], J. C. Mychaleckyj[14], P. Ruan[15], R. P. Tracy[9], A. C. Wood[11], B. Yu[16], M. R. Rooney[7], A. Manichaikul[4], R. Dubin[17], K. L. Mohlke[1], S. S. Rich[4], J. Rotter[11], P. Ganz[18], R. E. Gerszten[19], K. D. Taylor[11], L. M. Raffield[1]; [1]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [2]Stanford Univ., Stanford, CA, [3]BIDMC, Lexington, MA, [4]Univ. of Virginia, Charlottesville, VA, [5]Broad Inst., Cambridge, MA, [6]Natl. Jewish Hlth., Denver, CO, [7]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [8]Johns Hopkins Univ, Baltimore, MD, [9]Larner Coll. of Med. at the Univ. of Vermont, Burlington, VT, [10]Univ. of Pennsylvania, Philadelphia, PA, [11]Lundquist Inst., Torrance, CA, [12]Univ. of Washington, Seattle, WA, [13]Univ Colorado Denver, Aurora, CO, [14]Univ Virginia, Charlottesville, VA, [15]UT Southwestern Med. Ctr., Dallas, TX, [16]UTHlth.Houston, Houston, TX, [17]UT Southwestern, Dallas, TX, [18]Univ. of California, San Francisco, San Francisco, CA, [19]Beth Israel Deaconess Med. Ctr., Boston, MA

Abstract:

Introduction Protein measures from disparate affinity-based proteomics platforms often correlate poorly. Genetic variants that alter protein structure (PAVs) or post-translational modifications (PTMs) may impact epitope binding and drive inter-platform discordance for some proteins. As allele frequencies differ between ancestries, PAVs may also contribute to

inter-ancestry heterogeneity.

Methods For 2,157 plasma proteins measured on both SomaScan 7k and Olink Explore 3072 in 1,930 Multi-ethnic Study of Atherosclerosis participants [n=939 EUR, 439 AFR, 277 EAS, 231 AMR, 44 unassigned], we characterized protein quantitative trait loci (pQTL) and inter-platform protein measure correlation to identify potential genetic drivers of assay discordance across diverse ancestries.

Results We identified *cis*-pQTL (±1MB of the protein-encoding gene TSS, P<5E-08) for 855 SomaScan and 1035 Olink protein measures, of which 28% and 26% associate with PAVs. Of the 547 proteins with platform-overlapping *cis*-pQTL credible sets, 21 exhibited platform-discordant cis-pQTL effect directions, of which 14 contained PAVs. For these proteins, adjustment for the PAVs improved cross-platform measure correlation. For example, adjustment for platform-discordant rs1859788 [*PILRA*, p.Arg78Gly] increased the correlation of SomaScan and Olink PILRA measures from -0.35 to 0.27. 602 proteins exhibited nominally different inter-platform correlation estimates by ancestry (Cochran's Q p<0.05). Of these, 8 associated with PAVs common (EAF>0.10) in one ancestry reference, but rare (EAF<0.01) in others. Notably, adjustment for ancestry-differentiated PAV rs399145 [*PLAUR*, p.Thr86Ala, AFR EAF=0.12, EAF<0.01 in others] improved cross-platform suPAR correlation in African ancestry participants from 0.24 to 0.59. Among other mechanisms, *trans*-pQTL can reveal PTMs that impact proteomic assays. Hence, we further identified distal variants associated with 441 SomaScan and 233 Olink protein measures (P<1E-11). 12 loci associated with >=5 proteins (pleiotropic) on at least one platform; many of these loci are known to alter PTMs (ex. *ABO*). Proteins associated with pleiotropic trans-pQTL exhibited poor mean cross-platform correlation compared to all proteins with a *trans*-pQTL (0.36 vs 0.56 for pQTL detected on both platforms, 0.07 vs 0.23 for pQTL detected on one platform).

Conclusions Protein measures associated with PAVs and pleiotropic *trans*-pQTL are less likely to reflect true differences in abundances. However, adjustment for PAVs may offer a novel strategy for harmonizing cross-platform, cross-ancestry differences in affinity-based protein measurements.

Session Title: Omics Technologies Poster Friday Session

Board 1118F: Cross-Phenotype Genome-wide Association Study of Brain Morphology Features Extracted from Brain MRI Images

Authors:

G. Mend Mend Arachchige[1], R. Awasthi[2], X. Zhu[3]; [1]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Case Western Reserve Univ., Cleveland, OH, [3]Case Western Reserve Univ., Shaker Heights, OH

Abstract:

Brain morphology plays a critical role in the brain mechanisms underlying neurodevelopmental disorders, cognitive functions, and neurological diseases. However, understanding the genetic basis can potentially help discover new pathways for diagnosis, treatment, and prevention strategies, ultimately advancing personalized medicine in neurology and psychiatry. In this study, our aim is to investigate the genetic association between image-derived brain morphology features in individuals with neurological disorders. We utilized a subset of the ADNI data consisting of 619 individuals (362 males and 257 females) with a mean age of 75.79 years (SD = 6.61). Among these individuals, 130 were diagnosed with Alzheimer's Disease (AD), 190 were cognitively normal (CN), and 299 had Mild Cognitive Impairment (MCI). We also used publicly available Cortical Thickness features from different regions of the brain which were extracted using Deep Learning-based Advanced Normalization Tools (ANTs) methods. First, we performed hierarchical clustering using the complete linkage method to group all morphological features and conducted Genome-Wide Association studies (GWAS) on each feature. Subsequently, using GWAS summary data of all features we had, we performed Cross-Phenotype Trait Analysis using the CPASSOC method in each cluster to cluster-level association evidence. We reported loci with association p-value < 10-6. The cluster analysis resulted in 9 clusters from all 54 features. From our cluster wise cross-phenotype association results, we observed 7 independent SNPs (rs1505842, rs366463, rs10015547, and rs13134439 in Cluster 2, rs150584 in Cluster 6, rs7963150 in Cluster 8, and rs726558 in Cluster 9) with suggestive association evidence with morphological features. One of the most interesting findings we observed is the association between NAV3 and Cluster 8 features, which have features from the brain superior.parietal responsible for cognitive spatial orientation abilities. Earlier studies have reported an association of this gene with Alzheimer's disease. Similarly, in Cluster 9 which has features from the brain region transverse.temporal responsible for processing incoming auditory information, is associated with the PPP2R5A (nearby SNP rs726558) gene, which is also linked with Alzheimer's disease in earlier studies. Our research highlights the genetic loci linked to variations in brain morphology features, explaining the potential of GWAS in recognition of the genetic foundations of neurological disorders. However, our study has limited statistical power because of the small sample size.

Session Title: Omics Technologies Poster Friday Session

Board 1119F: CUT&RUN-EM: An ultra-sensitive, multiomic method that directly links chromatin features to DNA methylation ⭐

Authors:

K. Maier[1], B. J. Venters[1], V. U Sunitha Kumary[1], A. Hickman[1], A. Vaidya[1], J. M. Burg[1], L. Williams[2], V. Ponnaluri[2], P-O. Esteve[2], I. Meek[2], S. Pradhan[2], Z-W. Sun[1], M. W. Cowles[1], M-C. Keogh[1]; [1]EpiCypher Inc, Durham, NC, [2]New England Biolabs, Ipswich, MA

Abstract:

DNA methylation (DNAme) is a class of epigenetic marks that include the methylation of cytosine resides (5mC) within CpG islands. In addition to well characterized roles regulating gene expression, imprinting, and silencing parasitic DNA elements, the misregulation of DNAme is implicated in multiple diseases. Evidence is emerging that DNAme is not an independent epigenetic mark but rather closely linked to the post translational modification (PTM) of histone proteins. However, examining the relationships between 5mC and PTMs is hampered by correlative analyses that cannot establish a direct mechanistic link. Furthermore, the traditional approach to measure 5mC relies upon harsh bisulfite chemical conversation, which introduces DNA breaks and systemic biases.

To address these limitations, we developed the CUT&RUN-EM methylation mapping system: an ultra-sensitive multiomic technology that integrates EpiCypher's CUT&RUN genomic mapping with New England Biolab's Enzymatic Methyl-seq (EM-seq). The combination of these technologies reveals the direct molecular link between PTMs or chromatin proteins and DNAme at high resolution.

To quantify the capabilities and limitations of the assay, we tested multiple chromatin targets in various cell lines. We demonstrate that CUT&RUM-EM is highly reproducible, specific, sensitive, and efficient, notably requiring as few as 5 million short-read sequences per reaction (10-50x less than whole-genome 5mC-sequencing), demonstrating the disruptive potential of the assay to enable cost-effective targeted DNAme analysis. Finally, we leveraged CUT&RUM-EM to reveal mechanistic insights in clinically relevant mutations seen in Rett syndrome, a neurodegenerative disorder resulting from mutations in the 5mC-binding domain of the *MECP2* gene.

Session Title: Omics Technologies Poster Friday Session

Board 1120F: Defining a Clinical Data Ecosystem for Genomic Health

Authors:

N. Walton; Natl. Human Genome Res. Inst., Bethesda, MD

Abstract:

As the price of sequencing decreases there is significant potential to harness whole genome sequencing (WGS) as a critical component of an individuals' healthcare record. WGS could be completed once and used longitudinally in patient care providing information that is critical for prescribing, disease prevention, disease treatment, and diagnosis. Though there are millions of individuals for which this data has been generated there is currently no technical data ecosystem to support the large scale-use of WGS in clinical care. In April of 2024, the National Human Genome Research Institute held a workshop with experts to discuss the necessary developments in technology, standards, policy, and regulation to develop such an ecosystem. Through the workshop deliberations several key areas of focus were identified to further the stated goals. It is critical that data travels with the patient across the healthcare system and that patients can control access to their genomic data allowing them to participate in clinical trials and research. Such access would likely require federated systems with standard API's that currently have no definition. Other standards also need to be developed or improved, including standards for "genetic phenotypes", and more a standard representation of various metadata associated with VCF files. Development of genomic clinical workflows is critical. Such workflows are particularly underdeveloped for newer applications such as polygenic risk scores (PRS) and in their application to primary care. Scaling the clinical use of genomic data requires engagement with and integration into primary care. Regulation of the use genomic data that is generated in a CLIA laboratory but falls outside of the clinical report is unclear. Utilization of such data is critical to drive PRS, and artificial intelligence applications in genomics. While the need for some regulation is clear, there is significant concern regarding Food and Drug Administration (FDA) regulation of clinical decision support and laboratory developed tests and the potential impact on innovation.

Session Title: Omics Technologies Poster Friday Session

Board 1121F: Development of a pipeline for RNA editing analysis in single cell RNA-Seq data

Authors:

T. Whisenant, D. Chilin-Fuentes, G. Ildefonso, B. Rosenthal; Univ. of California San Diego, La Jolla, CA

Abstract:

RNA editing is an essential mechanism in post-transcriptional gene regulation, contributing to cellular diversity and functionality. The ability to accurately identify RNA edits is crucial for

understanding their role in gene expression and their implications in human diseases, including cancer. RNA editing can induce specific protein changes through coding edits, alter isoform ratios via intronic or synonymous coding edits, and influence transcript stability and lifespan by modifying microRNA binding sites or degradation rates.
Despite the significance of RNA editing, there are currently no tools available for identifying edits per cell in single-cell RNA sequencing (scRNA-Seq) data. While single cell data is ideal for capturing cell-specific edits, the limited depth in each cell poses a challenge. To address this, we propose a method to call RNA edits at the cell type level rather than at the single cell level. This approach enables the identification of significant editing events within and between different cell types, providing insights unavailable through bulk RNA-Seq.
Our methodology begins with standard scRNA-Seq data processing using CellRanger and Seurat. Cells are clustered and annotated by cell type, followed by computation of total counts per cluster for normalization purposes. We also perform differential expression (DE) analysis between clusters and between different conditions (within clusters) to integrate downstream with the editing results. To isolate RNA edits, we subset BAM files by cluster-specific barcodes and apply them to a standard RNA editing pipeline, which includes variant calling, various filtering steps, stratification into Alu/nonAlu subsets, and annotation steps. Quality checks ensure sufficient sequencing depth and adequate cell representation within each cluster as well as verification that edited genes are expressed within respective clusters. Finally, downstream analysis computes differential editing between cell types and conditions within cell types.
The results are summarized to identify types of edits prevalent in each cell type, which are then correlated with DE data. Additionally, we perform pathway analysis on intronic edits due to their abundance, revealing potential regulatory mechanisms and pathways impacted by RNA editing. This comprehensive approach not only enhances our understanding of RNA editing at a cellular level but also bridges a significant gap in scRNA-Seq analysis. By identifying cell type specific RNA edits and their functional implications, we can better understand the molecular underpinnings of diseases, paving the way for targeted therapeutic strategies.

Session Title: Omics Technologies Poster Friday Session

Board 1122F: Digital Pathology Image Analysis Enhances 10-Year Risk Stratification for Early Stage ER+ Breast Cancer

Authors:

G. Colditz[1], H. Zhou[2], S. Jiang[1]; [1]Washington Univ. Schhool of Med., Saint Louis, MO, [2]Harvard, Boston, MA

Abstract:

Despite the long and broad focus on risk prediction models and prevention strategies, breast cancer remains the leading cancer diagnosis among women worldwide. A prevalent subtype of invasive breast cancer is estrogen receptor-positive (ER+) and lymph node-negative (LN-) which has standard treatment guidelines such as breast conserving surgery. For a subset of high-risk patients, adjuvant chemotherapy is often given, but is accompanied with significant side effects. Therefore, accurate identification of patients at high risk of breast cancer progression is crucial to guide treatment choices. Existing methods for predicting breast cancer progression include the Nottingham grading system and OncotypeDx based on 21-gene panel. However, the former exhibits between-pathologist variability and the latter has high cost. We hypothesize that making full use of the microenvironment information readily embedded within the digital pathology H&E images can more accurately identify patients at high risk of 10-year progression. Utilizing the TCGA BRCA dataset, we have identified 764 patients with diagnostic digital pathology images (104 cases with progression). The average number of years of follow-up in this cohort is 3.2 years (sd = 2.9). After subsetting to ER+ and LN-, we retained 282 patients (23 cases with progression) for analysis. In predicting the 10-year progression risk, we trained a sample-specific learning algorithm to fully capture the microenvironment information within the digital pathology images. To validate model performance, a 5-fold cross-validation was conducted within TCGA. As a benchmark, we first analyzed the OncotypeDx 21-gene panel and obtained a 10-year AUC = 0.64 (sd = 0.03). Applying our proposed sample-specifc learning algorithm we obtained a 10-year AUC = 0.78 (sd = 0.04). When we add the 21-gene panel to the proposed algorithm, we have obtained 10-year AUC of 0.79 (sd = 0.04), suggesting that our algorithm has well captured those information contained within OncotypeDx. This better separation under the proposed model can be further visualized by looking at the distribution of risk scores as well as predicted survival curves for progressors vs. progression-free patients. In summary, we have developed and applied a supervised learning method that makes full use of the H&E pathology images available at time of diagnosis to better predict 10-year progression for patients with ER+ and LN- breast cancer. The proposed algorithm exhibits better performance when compared with OncotypeDx. Importantly, using the pathology image comes at no additional cost to patients.

Session Title: Omics Technologies Poster Friday Session

Board 1123F: Direct detection and counting of trinucleotide repeat expansion using CRISPR-Cas9 and PacBio HiFi Sequencing

Authors:

B. Mehta[1], D. Corney[1], W. Wang[1], J. Chen[2], H. Bahgat[2], F. Zaidi[2], C. Mozdzierz[1], H. Latif[1]; [1]GENEWIZ from Azenta Life Sci., South Plainfield, NJ, [2]CHDI Fndn., New York, NY

Abstract:

Trinucleotide repeat expansions are a major cause of neurological and developmental disorders, including Huntington's disease, fragile X syndrome, and various spinocerebellar ataxias. When the number of repeats exceeds a certain pathogenic threshold length, they can lead to toxic gain-of-function or loss-of-function effects. Historically, PCR and Southern blot approaches have been used to identify repeat expansions. However, there are significant technical challenges associated with these methods, including successfully amplifying through hundreds or thousands of base pairs of repetitive sequences and failed PCR due to high GC content. As a result, these approaches suffer from poor accuracy and resolution of the number of repeats. Consequently, the loci containing these genes are commonly referred to “dark” regions of the genome and are poorly characterized. To circumvent these challenges, we have tested a novel PCR-free CRISPR-Cas9 approach to enrich 20 genes commonly involved in repeat expansion disorders followed by long-read PacBio sequencing to successfully sequence and align repetitive sequence. High molecular weight DNA was extracted from frozen EDTA whole blood or fresh frozen tissue, followed by PureTarget library preparation. In this workflow DNA is first digested with Cas9 and guide RNAs to enrich 20 target genes, followed by ligation of SMRTbell adapters on the cut end. After removal of non-ligated templates, libraries are pooled and sequenced on a PacBio sequencer. Here, we describe the accuracy testing results of this novel approach. Accuracy of calling the number of trinucleotide repeats was established by processing previously characterized samples from Coriell with known expansions in DMPK, ATX1, ATXN3, FMR1, and FXN. Additionally, DNA extracted from blood of healthy donors without any known expansions, and from brain tissue of Huntington disease individuals was performed and compared to whole genome PacBio data. The CRISPR-Cas9 approach combined with long-read PacBio sequencing offers a powerful method for characterizing trinucleotide repeat expansions, overcoming the limitations of traditional techniques. This novel approach paves the way for improved understanding the genotype-phenotype and molecular basis of neurological and developmental disorders caused by repeat expansions.

Session Title: Omics Technologies Poster Friday Session

Board 1124F: DNA-Diffusion: Leveraging Generative Models for Controlling Chromatin Accessibility and Gene Expression via Synthetic Regulatory Elements ⭐

Authors:

L. Pinello[1], L. Ferreira DaSilva[1], S. Senan[1], J. Kribelbauer[2], Z. Patel[1], A. Reddy[3], S. Gabbita[1], J. Rosen[4], Z. Nussbaum[5], C. Cordova[6], A. Wenteler[7], N. Weber[8], T. Tunjic[8], M. Mansoldo[9], T. Khan[10], G-H. Hwang[1], V. Gardeux[11], Z. Li[12], C. Smith[1], M. Bejan[13], L. Louis[12], P. Cornejo[14], W. Connell[10], B. Deplancke[11], M. Love[15], E. Wong[12], W. Meuleman[16]; [1]MGH/Harvard/BROAD, Charlestown, MA, [2]EPFL, Lausanne, Switzerland, [3]Berkeley, Berkeley, CA, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Nomic Ai, NY, NY, [6]Johannes Kepler Univ., Linz, Austria, [7]Queen Mary Univ. of London, London, United Kingdom, [8]TU Vienna, Vienna, Austria, [9]GSK.ai, London, United Kingdom, [10]Independent Res.er, NA, CA, [11]EPFL, Laussaune, Switzerland, [12]UNSW, Sydney, Australia, [13]Univ. of Bucharest, Bucharest, Armenia, [14]UNSW, Sydney, Austria, [15]UNC-Chapel Hill, Chapel Hill, NC, [16]Altius Inst., Univ. of Washington, Seattle, WA

Abstract:

The challenge of systematically modifying and optimizing regulatory elements for precise gene expression control is central to modern genomics and synthetic biology. Advancements in generative AI have paved the way for designing synthetic sequences with the aim of safely and accurately modulating gene expression.
We leverage diffusion models to design context-specific DNA regulatory sequences, which hold significant potential toward enabling novel therapeutic applications requiring precise modulation of gene expression. Our framework uses a cell type-specific diffusion model to generate synthetic 200 bp regulatory elements based on chromatin accessibility across different cell types. We evaluate the generated sequences based on key metrics to ensure they retain properties of endogenous sequences: transcription factor binding site composition, potential for cell type-specific chromatin accessibility, and capacity for sequences generated by DNA diffusion to activate gene expression in different cell contexts using state-of-the-art prediction models.
We employ an innovative endogenous replacement approach that enables the direct measurement of gene expression from a target enhancer-gene pair within its native context. To experimentally validate the functional relevance of sequences generated by DNA-Diffusion, we tested their ability to modulate the endogenous expression of the human gene AXIN2, a potential therapeutic target in leukemia. Our experiments confirmed that DNA-Diffusion-derived sequences can drive AXIN2 expression in a cell type-specific manner and across a range of expression levels, highlighting the importance of validating the function of synthetic sequences in a native genomic context.

In summary, our results demonstrate the ability to robustly generate DNA sequences with cell type-specific regulatory potential. DNA-Diffusion paves the way for revolutionizing a regulatory modulation approach to mammalian synthetic biology and precision gene therapy.

Session Title: Omics Technologies Poster Friday Session

Board 1125F: DNAfusionDetect: A Clinical Integrative Pipeline for Detection and Annotation of Gene Fusions from DNA Sequencing Data

Authors:

C. Kuang[1], J. Zhou[1], F. Chang[1], W. Mu[2]; [1]KingMed Diagnostics, Shanghai, China, [2]Kingmed Diagnositcs, Guangzhou, China

Abstract:

Background: Gene fusions, arising from chromosomal rearrangements, are key drivers in oncogenesis and tumor progression. The identification of gene fusions is crucial for precise diagnosis, risk stratification, and targeted therapy. Next-generation sequencing (NGS) enabled enhanced gene fusion detection compared to traditional methods, but the clinical application of gene fusion analysis using DNA-based NGS data is still hindered by lacking gold standard bioinformatics tools for informative annotation, optimal filtration, and accurate allele frequency (AF) estimation.
Methods: We developed DNAfusionDetect, a comprehensive DNA fusion detection pipeline integrating fusion calling, AF adjustment, and post-calling filtration. Manta software was employed in high-sensitivity mode for initial fusion detection. Each breakpoint in the results was annotated with gene symbols and strands. Fusion events were consolidated by matching breakpoint pairs based on variant types and proximity, and then annotated with clinical information from databases such as Mitelman and FusionGDB2. Fusion supporting reads were extracted from evidence BAM files provided by Manta, and wild-type reads surrounding breakpoints were analyzed for accurate AF adjustment. A post-calling filtration process was implemented to ensure the retention of clinically relevant fusion events and the exclusion of likely artifacts due to high degrees of polymorphism, insufficient supporting read count, or high internal population frequency, etc.
Results: Evaluating DNAfusionDetect using 12 samples with known somatic gene fusions yielded a 100% detection rate with an average of only 5 events in the final list per sample, demonstrating its precision and efficiency. Sensitivity was further assessed on 94 known immunoglobulin gene (*IG*) translocations/rearrangements in 54 samples, achieving a sensitivity of 87.23%, significantly surpassing the 73.40% sensitivity of the commonly used *IG*-specific tool

IgCaller. The pipeline's limit of detection (LOD) was rigorously assessed by simulating 10 gene fusions with AF gradients ranging from 0.2% to 5% and depth gradients ranging from 1000x to 10000x, achieving a LOD of 2% at 1000x depth, 0.4% at 2000x depth, and 0.2% at depths over 2000x. Adjusted AFs were consistent with simulated AFs, with an average relative error of 0.2 for detected events. In conclusion, we have developed DNAfusionDetect, an integrative method for detecting and annotating gene fusions from DNA NGS data. This method offers high sensitivity and low false positive rates, making it a robust tool for efficient clinical analysis of gene fusions in cancer.

Session Title: Omics Technologies Poster Friday Session

Board 1126F: Enabling workflows for determining mechanisms of action in drug discovery with flexFS

Authors:

G. Planthaber[1], M. Colosimo[1], M. Peterson[1], A. Cosgrove[1], D. Wojtasinski[2], S. Sarangi[1]; [1]Paradigm4, Waltham, MA, [2]Paradigm4, North Dartmouth, MA

Abstract:

Despite the unprecedented growth in experimental technologies, data generation, and analysis methodologies, the approval rate of clinical drugs has not seen similar success. Multiple studies show that 90% of drug targets fail during clinical trials. Moreover, the average time for a new drug to be approved for clinical use is 10-15 years with a cost of over $1-2 billion. While several challenges contribute to the low rate of effective drug development, poor understanding of disease mechanisms and rising costs in discovery pipelines play significant roles. In this work, we look at one of the technologies that improves understanding of the mechanism of action for a disease or a biological network - protein structure prediction.
Accurate prediction of protein structure from amino acid sequences, a notoriously difficult problem, has witnessed tremendous advances with the development of artificial intelligence (AI) powered models such as AlphaFold2. However, running such algorithms is typically computationally demanding and relatively expensive. Here, we present an easy to use, cost-effective, and highly performant workflow for running protein structure prediction algorithms such as AlphaFold2. This workflow leverages a POSIX-compliant networked file system called flexFS, developed by Paradigm4. flexFS is designed as a highly elastic, low latency, high throughput file system that has been built on an object store. Using publicly available datasets, we present compute time and costs for the multiple sequence alignment (MSA) and structure prediction steps for AlphaFold2. Additionally, we present results when utilizing Nvidia's Cuda

libraries for an accelerated version of the workflow.
The robust and accurate performance of protein structure prediction models such as AlphaFold2 can have a significant impact in improving understanding of mechanisms of action for diseases. Application of these models in easily deployable, compute-optimized, and cost-effective workflows can help make drug discovery and target identification more effective.

Session Title: Omics Technologies Poster Friday Session

Board 1127F: Ethidium bromide treated cell lines reveal nuclear gene expression responses to changes in mitochondrial DNA copy number

Authors:

J. Xie[1], P. Win[2], C. Newcomb[1], S. Zeng[3], C. Castellani[2], D. Arking[1]; [1]Johns Hopkins Univ Sch. of Med., Baltimore, MD, [2]Western Univ., London, ON, Canada, [3]Johns Hopkins Univ., Baltimore, MD

Abstract:

Mitochondrial DNA copy number (mtDNA-CN) is a metric of mitochondrial function that has been associated with cardiovascular disease and all-cause mortality. To investigate gene regulatory networks and processes associated with mtDNA-CN, we perturbed HEK 293T cells with ethidium bromide (EtBr), a chemical that intercalates into mtDNA, blocking transcription of mitochondrial encoded genes and leading to mtDNA depletion.
Cells were exposed to EtBr in seven different doses ranging from 0 to 150 ng/mL in increments of 25 ng/mL for each 9.6 cm2 well. Cells were harvested after 48 hours, at which time mtDNA-CN and gene expression were measured. Additionally, we tracked mtDNA-CN and gene expression longitudinally, with cells grown in media with 150 ng/mL of EtBr for the experimental group, and cells grown concurrently in media with no EtBr for the control group. Cells were harvested at 0, 24, and 48 hours. At 48 hours, the growth media is replaced, and cells in the experimental group were allowed to recover in media with no EtBr for an additional 24 hours, with another round of measurements taken at that time. In the dosage experiments, mtDNA-CN was significantly associated with 279 nuclear encoded genes. We observed that expression for these genes in the treatment and recovery experiment followed three distinct tracks: 67 genes increased and decreased linearly matching the change in mtDNA-CN; 35 genes behaved in a switch-like on-off fashion matching the pattern observed for mitochondrial encoded genes; and 89 genes showed a temporally delayed response to treatment and recovery.
Despite an observed 8-fold decrease in mtDNA-CN, we did not detect differential expression either for nuclear genes involved in mtDNA replication - *POLG*, *TWNK*, *SSBP1* - or for nuclear

gene involved in mitochondrial transcription - *TFAM*, *POLRMT*, *TFB2M*, *TEFM*. Conversely, among the 40 genes which showed a delayed response and were upregulated, those involved in biological processes such as amino acid metabolic process, tRNA aminoacylation, and amino acid import across plasma membrane were overrepresented. This suggests that nuclear response to decreased mitochondrial gene expression tends to be limited to increasing resources for mitochondrial translation, while genes for mtDNA replication and transcription are more tightly regulated.

Session Title: Omics Technologies Poster Friday Session

Board 1128F: Evaluating the dynamics of germline mutation at repetitive loci with AVITI sequencing and multi-generational Utah/CEPH pedigrees

Authors:

H. Happ, T. Sasani, D. Warner, D. Neklason, A. Quinlan; Univ. of Utah, Salt Lake City, UT

Abstract:

De novo mutations (DNMs) are a fundamental source of genetic diversity, yet they also underlie more than 40% of developmental disorders. By sequencing thousands of familial genomes, including those of 33 large, multi-generation CEPH/Utah pedigrees, we understand many of the factors that influence point and structural mutation. However, nearly all studies ignore the most mutable loci in the human genome: short tandem repeats (STRs) and homopolymers. The high mutability is largely understood to be due to high rates of polymerase slippage during DNA replication. Similar polymerase slippage used in Illumina short-read sequencing leads to an elevated error rate near repetitive sequences that overwhelms the signal of true mutations and hinders studies of STR mutation. In 2023, Element Biosciences released the AVITI sequencing platform, which achieves a substantially lower error rate within and near repetitive sequences. Using this new sequencing technology, we can more accurately study mutation dynamics at STRs. As part of the Platinum Pedigree project, we sequenced PCR-free libraries from 28 members of a 4-generation CEPH/Utah pedigree on an Element AVITI platform to a genome-wide median depth of 34X. We genotyped each individual at ~450k uninterrupted homopolymers loci to identify putative DNMs in 24 distinct trios within the pedigree. We observe a median of 51 (range: 17-70) homopolymer DNMs per individual and a median mutation rate of 9.6 x 10e-5 DNMs per loci per individual (range: 3.18 x 10e-5 - 1.35 x 10e-4). Work to identify DNMs at interrupted homopolymers and other STR types, and to assess the transmission of mutations to the subsequent generation is ongoing. Notably, we can genotype

66% more homopolymer loci using Element data compared to Illumina data. This substantial increase is largely due to a lower sequencing error rate, which yields more aligned reads for genotyping. These data allow us to interrogate several fundamental aspects of homopolymer mutations such as how mutation rate varies as a function of parental age and sex, sequence context, coding or noncoding context, and homopolymer length. Together, this work will advance our understanding of the fundamental genome biology underlying these hypermutable genomic elements and create important references of variation. Finally, this technology is a promising substrate for characterizing variation in homopolymers and other STR types at disease loci.

Session Title: Omics Technologies Poster Friday Session

Board 1129F: Exposomic, proteomic, and metabolomic analyses reveal environmental toxicant exposures in the Rio Grande Valley

Authors:

M. Arriaga[1], R. Hernandez[1], N. Garcia-Rodriguez[1], A. C. Leandro[1,2], V. Diego, P.[1,2], T. E. Howard[1,2], E. G. Manusov[1], J. G. Parsons[1], J. E. Curran[1,2], S. Williams-Blangero[1,2], J. Blangero[1,2], J. A. Galan[1,2]; [1]Univ. of Texas Rio Grande Valley, Brownsville, TX, [2]South Texas Diabetes and Obesity Inst., Brownsville, TX

Abstract:

Exposure to environmental toxicants has a significant impact on risk for human disease. Environmental toxins alter cellular pathways resulting in increased risk for chronic diseases such as CVD, T2D, Alzheimer's dementia, and cancer. Genotype-by-environment interactions may influence the individual response to toxicants, but better methods are needed to identify/measure toxicants likely to be involved in such interactions. This study aims to identify the toxicological profile of a vulnerable population in South Texas by examining and quantifying environmental toxicants and their effects on disease risk. This study was conducted using data and samples from the Mexican American Family Study (MAFS) cohort, which includes over 3,500 participants from 80 families in South Texas. Twenty samples were screened for the presence of toxicants using high-resolution mass spectrometry-based metabolomic/proteomics. Bioinformatic pathway and disease associations were assessed using a comprehensive toxicology database linking metabolomic results to known toxicological profiles. Multiomic analysis of 20 plasma samples from the MAFS identified the presence of 296 proteins and 844 metabolites that have known roles in lipid metabolism dysregulation and cell death/survival pathways of substantive relevance for neurological and cardiovascular disease pathogenesis.

The data suggest toxicants involved in the inhibition of PPARA/RXRA signaling as potential drivers of complex disease progression via lipid dysregulation and inflammation. Exposomic analysis identified compounds linked to industrial chemicals, drugs, medications, pesticides, cosmetics, carcinogens, and mycotoxins. In summary, our study employed specific exposomic, proteomic, and metabolomic profiling to identify associated complex disease pathways in the MAFS. This integrative approach enabled the cross-referencing of metabolite data with toxicology information, potentially revealing new biomarkers and molecular pathways linked to specific exposures and diseases. Integrating this data with genomics and transcriptomics will help in identifying genetic and environmental risks, assisting public health officials in developing policies to raise awareness of environmental exposure in South Texas.

Session Title: Omics Technologies Poster Friday Session

Board 1130F: Flexible, Production-Scale, Human Whole Genome Sequencing on a Bench-top Sequencer

Authors:

B. Krajacich, K. Green, K. Wiseman, P. Mains, S. Robertson, M. Sudkamp, B. Lajoie, J. Zhao, S. Levy; Element BioSci.s, San Diego, CA

Abstract:

Human whole-genome sequencing (hWGS) provides comprehensive genomic information that is valuable for research that can potentially help guide disease prevention and treatment, and ultimately improve human health. Recent advancements in sequencing technology have improved sequencing quality and further reduced sequencing costs on bench-top sized instruments, making whole-genome sequencing an accessible technology for broader use. Here, we demonstrate the feasibility of a large whole genome sequencing project on a scale previously reserved for production-scale, factory-sized machines using a benchtop sequencer in a small laboratory setting. In this project, 807 samples were prepared and sequenced across 313 flow cells, with high sequencing quality at a median %Q30 of 96.6% and a median %Q40 of 89.31%. To screen library quality and maximize sample yield, we utilized 48-plex sample pre-pool 'QC' runs to improve sample balancing and to provide >1x coverage per sample prior to full-depth sequencing. With this strategy, we consistently achieve >30x human whole genome sequencing of three-plex sample trios with standard settings or up to 4 samples per run with a high-throughput run setting. Additionally, this low-pass data provided diagnostic insights, allowing for detection of chromosome copy number variations (CNV) prior to full-depth sequencing. With three instruments running concurrently, > 4,000 30x human whole genomes

could be sequenced per year. To further demonstrate additional flexibility present in the platform, we also describe a proof of concept for rapid hWGS sequencing to maximize sample to answer turnaround time for time-critical sequencing applications. Sequencing of a 2x100 >30x human WGS can be achieved in <12 hours and subsequent generation of fastq, bam, and vcf in <1 additional hour. This study provides a cost-effective and flexible real world demonstration of achieving both high quality hWGS sequencing and instrument flexibility without the need for complex batching schemes or factory-sized sequencers.

Session Title: Omics Technologies Poster Friday Session

Board 1131F: Formaldehyde Inhalation Toxicity: A Comparative Study of Acute and Chronic Exposure Using Transcriptomic Data

Authors:

D. Shin[1], D. Kang[1], J. Lee[2], K. Lim[3], Y. Seo[1]; [1]Dongguk Univ., Goyang-Si, Korea, Republic of, [2]Dongguk Univ., Goyang-si, Korea, Republic of, [3]Ewha Womans Univ., Seoul, Korea, Republic of

Abstract:

Formaldehyde is widely used industrially, leading to frequent exposure to both workers and the general population. Acute formaldehyde exposure is known to cause irritation and allergic reactions, while chronic exposure can lead to nasal sinus cancer and nasopharyngeal cancer, leukemia. Therefore, formaldehyde exposure exhibits different toxic effects depending on the exposure duration. In this study, we explored the distinct toxicological pathways associated with different durations of formaldehyde exposure using public high-throughput transcriptomic expression profiles. The transcriptomic data were analyzed using Ingenuity Pathway Analysis software. We observed that acute inhalation of formaldehyde (6 hours) increased the Nrf2-mediated oxidative stress response. In contrast, chronic exposure (5 days, 28 days) resulted in the suppression of the Nrf2-mediated oxidative stress response and overexpression of the spindle assembly, which is highly correlated with tumor progression and metastasis. Based on these results, we suggest candidate transcriptomic markers specific to acute and chronic formaldehyde exposure. Our findings contribute to understanding formaldehyde inhalation toxicity and underline exposure duration as a critical variable in mechanism of toxicity.

Session Title: Omics Technologies Poster Friday Session

Board 1132F: Functional validation of sickle cell disease alloimmunization susceptibility locus in the spleen

Authors:

L. Goldberg, A. Haldipur, A. Motter, N. Asmus, N. Hanchard; NIH, Bethesda, MD

Abstract:

Individuals with sickle cell disease (SCD) often receive transfusion of red blood cells (RBCs) routinely as a means of symptom management and acute crisis prevention. Despite its therapeutic benefits, persistent transfusions increase risk of alloantibody development against transfused RBCs. This phenomenon, known as alloimmunization, is disproportionately exhibited by individuals with SCD compared to other regularly transfused groups. There is currently no known biological predictor of alloimmunization risk. We previously identified an African-ancestry-restricted variant associated with increased alloantibody formation in SCD patients at chromosome 5q33 annotated to an ENCODE putative enhancer. This enhancer overlaps non-coding gene *LINC01847* and adrenoceptor gene *ADRA1B*. Although its local role is not known, *ADRA1B* is highly expressed in human spleen, an organ known to be involved in the clearance of alloimmunized RBCs and production of alloantibodies. We hypothesized that *ADRA1B* expression is regulated by the 5q33 enhancer and participates in alloimmunization in the spleen. To test this hypothesis, we perioperatively acquired healthy human spleen tissue from elective surgical interventions and performed 3' single cell RNA sequencing. We annotated seven discrete cell populations in the sequenced sample, namely CD8+ and CD4+ T cells, naïve and activated B cells, natural killer cells, macrophages, and other monocytes. These cells broadly expressed genes relevant to the 5q33 locus at limited quantities, including *IL12B*, an immunoregulatory gene that physically interacts with the 5q33 enhancer and its receptor subunits *IL12RB1* and *IL12RB2*. Cells from our sample did not express *ADRA1B* at baseline; therefore, we subjected cells to stimulatory compounds that have previously been shown to induce *ADRA1B* expression in non-spleen tissues: cycloheximide, forskolin, and norepinephrine. We observed a 13-fold and 7-fold increase in bulk *ADRA1B* expression when splenic cells were exposed to cycloheximide at 10 ug/mL and 20 ug/mL, respectively, compared to minimal increases with forskolin or norepinephrine. These results suggest that *ADRA1B* is highly expressed in the spleen after appropriate stimulation, which is a prerequisite to proposed single cell studies to identify the splenic cell types that express our genes of interest and for investigating of the relationship between *ADRA1B* and the 5q33 enhancer *in-vitro*. Our research will improve understanding of the underlying mechanism of increased alloimmunization risk among individuals with SCD, thereby informing the development of precision health interventions at the time of transfusion.

Session Title: Omics Technologies Poster Friday Session

Board 1133F: Harnessing the power of 3D genomics for personalised medicine and drug discovery

Authors:

D. Turner, N. Walker, K. Chahwala, S. Schoenfelder, P. Fraser, H. Jones; Enhanc3D Genomics, Cambridge, United Kingdom

Abstract:

Background/Objectives: non-coding regions of the genome, where most disease-associated GWAS variants are found, are highly enriched in elements that control gene expression in a cell-type-specific manner. Regulatory elements can be located thousands to millions of base pairs from the genes they target, and must be brought into proximity by folding of the genome for a functional interaction to occur.Understanding these long-range, cell-specific functional interactions is essential to identifying high-confidence drug targets and drug-repurposing candidates, and stratifying patients with accuracy. However, most traditional genomics-based approaches interpret the genome as a linear structure, ignoring the intricate 3D architecture, and are consequently blind to these interactions.
Methods: The GenLink3D™ pipeline uses promoter-capture Hi-C to identify long-range interactions in a given cell type, across the entire genome in a single experiment. Combining our advanced algorithms and machine-learning capability allows us to understand the 3-dimensional organisation of the entire genome with high resolution, enabling us to decode and exploit novel targets for disease diagnosis and treatment, with high throughput and precision.
Results: Our analyses show that hundreds of thousands of regulatory elements are active in each cell type, some of which are ubiquitous and others cell-type specific. These elements interact with their target promoters, driving cell-type-specific expression. We demonstrate how systematically connecting disease-specific variants to their target genes can enable the discovery of novel disease-susceptibility genes and illuminate new drug targets. In addition, we show that variation in cell-specific SNP-risk load across ethnically distinct populations can be used to derive poly-enhancer scores, offering improved patient stratification over the use of traditional polygenic risk scores.
Conclusion: We demonstrate a hypothesis-free universal discovery tool to unlock any genetic disease by identifying the associated genes, and relevant cell types, and to de-risk the drug-discovery process.

Session Title: Omics Technologies Poster Friday Session

Board 1134F: Identification of immunogenetic parameters associated with cancer health disparities through TCGA data analysis

Authors:

A. Shanker, H. Rajakaruna, T. Kanagasabai, S. Gonzalez Ochoa, Q. Wang; Meharry Med. Coll., Nashville, TN

Abstract:

Background: Genetic diversity is a determinant factor of the differential expression of several components impacting immune response mechanisms and the variability of positive therapeutic outcomes among individuals. These differences are likely to contribute significantly to the cancer disparities observed predominantly in individuals of African descent. Nonetheless, the underlying biological mechanisms governing the correlations between immune-inflammatory elements in the tumor microenvironment and the aggressiveness of cancer in African Americans compared to Caucasian Americans remain unclear. The Cancer Genome Atlas (TCGA) database provides biologists and oncologists an opportunity to delve into the fundamental biological basis of immune-inflammatory components within the TME and their potential correlation with the aggressiveness of cancers in African Americans (AA) versus Caucasians (CA). **Methods:** We conducted differential gene expression analyses of bulk RNA-seq data from tumor samples of breast, lung, and prostate cancer of the patient cohorts of AA and CA descents given in the TCGA database. We also computed the tumor-infiltrating immune cell probabilities and gene enrichments to gain mechanistic insights into differences in protein signaling pathways between the cohorts using standard pipelines in Python. **Results:** Lung and prostate tumor samples from AA showed a significant upregulation of *PDL1* (*CD274*) ($p<0.05$), *CTLA4* (CD152) ($p<0.0001$) and *NOTCH2NL/N2N* ($p=0.0037$) gene expression compared to CA. Conversely, we observed a downregulated gene expression of markers associated with anti-tumor activity (lung: *GZMB* (Granzyme-B, $p<0.04$) and prostate: *IFNG* (IFNγ, $p<0.01$) in AA compared to CA. Moreover, analysis of infiltrating immune cell populations revealed that AA tumors exhibit significantly lower proportions of tumor-infiltrating $CD8^+$ T cells ($p<0.003$) and a higher proportion of T regulatory cells (Treg) compared to CA ($p=0.001$). Consequently, differential enrichment of gene sub-networks is observed between the CA and AA population cohorts. **Conclusions:** The data reveals a significantly increased probability of immunosuppressive and exhausted T cells in AA compared to CA, suggesting deficiencies in immune defense mechanisms based on gene expression, recruitment of T cells, and secretion of cytokines in enlisting active gene subnetworks to combat cancer. Our findings emphasize the

need to explore the immune functional differences among ethnic populations to alleviate cancer health disparities.

Session Title: Omics Technologies Poster Friday Session

Board 1135F: Identifying Key Regulators of Alzheimer's Disease Risk Using Single-Nucleus RNA-Sequencing Analysis

Authors:

M. Nwadiugwu[1], H. Shen[2], H-W. Deng[3]; [1]Tulane Univ., New Orleans, LA, [2]Tulane Univ, New Orleans, LA, [3]tulane Univ., New Orleans, LA

Abstract:

Background: Early detection of Alzheimer's Disease (AD) presents significant challenges, primarily due to the complex and poorly understood mechanisms underlying the disease. This study aims to identify key molecular and regulatory features of AD risk by employing transcriptomics in order to better understand mechanisms underlying the disease. **Method:** Utilizing single-nucleus RNA sequencing data from the Religious Orders Study and Memory and Aging Project (ROSMAP)—comprising 80,660 droplet-based snRNA-seq profiles from the prefrontal cortex of 24 individuals with Alzheimer's disease (AD; 15 early and 9 late pathology) and 24 cognitively normal individuals, matched by sex and including 2 Black or African Americans and 46 non-Hispanic Whites—we identified gene regulatory proteins significantly associated with AD. We employed the ARACNe-VIPER algorithm for gene regulatory network reconstruction and virtual inference of protein activity, to distinguish differences in regulatory networks and protein activity between the groups. Our findings were validated using the ChIP-X Enrichment Analysis Version 3 web tool (ChEA3), which performs transcription factor (TF) enrichment analysis through orthogonal omics integration, supporting the identification of candidate regulators in AD-specific cis-regulatory elements. **Result and Discussion:** Initial analysis pinpointed five (5) key TFs—ZEB2, SOX10, OLIG2, NKX62, MYRF—implicated in neurofibrillary tangle formation and cortical atrophy in AD. We also discovered two TFs (MAFB, TCF4) exhibiting variable activity across sex, cognitive states (normal, early, and late AD pathology), and amyloid burden (low and high groups). Further, we identified eight TFs (PAX6, MAFB, MYT1, ID3, GLI3, EMX2, PRDM16, NFIB) associated with abnormal retina human phenotypes. **Conclusion:** Our study uncovers genetic regulatory markers of cognitive decline, offering a transformative perspective on diagnosing and managing cognitive decline and AD, especially within aging populations. These insights emphasize the genetic foundations of the disease, setting the stage for future validations with independent

datasets. **Keywords:** Alzheimer's Disease, Transcriptomics, Cognitive Decline, Regulatory Networks, Single-Nucleus RNA Sequencing. **Funding:** This research was made possible with partial support from the National Institutes of Health (R01AR069055, U19AG055373, R01AG061917, and P20GM109036).

Session Title: Omics Technologies Poster Friday Session

Board 1136F: igh-Coverage Spatial Genomics with Barcoded Multiple Displacment Amplification

Authors:

A. Lee[1], S. Lee[2]; [1]Meteor Biotech, Seoul, Korea, Republic of, [2]Meteor Biotech, SEOUL, Korea, Republic of

Abstract:

Deciphering the mutational landscapes within spatially heterogeneous cell microniches is pivotal for advancing cancer research. Traditional genome amplification techniques, such as Multiple Displacement Amplification (MDA), offer high coverage but are not scalable for extensive spatial genomic studies. To address this, we developed barcoded MDA (bMDA), a novel technique that integrates unique cell barcodes during the amplification process, enabling high-coverage, multiplexed genomic analysis of low-input DNA samples while preserving spatial context.In our study, bMDA was applied to triple-negative breast cancer (TNBC) tissues to uncover subclonal evolutionary relationships by examining copy number alterations, single nucleotide variations, structural variations, and kataegis signatures. The application of bMDA revealed significant insights into the spatial distribution and evolutionary trajectories of cancer subclones within the tumor microenvironment, highlighting the importance of spatial context in understanding tumor biology.The introduction of bMDA provides several advantages over traditional methods. Firstly, it allows for the high-throughput analysis of numerous samples simultaneously, significantly reducing both the cost and labor involved in spatial genomic studies. Secondly, the use of cell-specific barcodes ensures that data from individual cells can be accurately traced back to their spatial origins, facilitating a detailed understanding of cellular heterogeneity within tumors.Our findings demonstrate that bMDA is a powerful tool for high-coverage sequencing in spatial genomics, offering unprecedented resolution and scalability. This technology holds immense potential for accelerating cancer research, particularly in understanding tumor evolution, drug resistance mechanisms, and identifying novel therapeutic targets.The scalability of bMDA ensures its broad application in various fields beyond cancer research, including neurodevelopmental disorders and infectious diseases. By integrating spatial

information with high-resolution genomic data, bMDA paves the way for more precise diagnostic and therapeutic strategies, ultimately contributing to the advancement of personalized medicine. The potential for bMDA to revolutionize spatial genomics makes it an invaluable tool in the quest to understand the complex biology of diseases.

Session Title: Omics Technologies Poster Friday Session

Board 1137F: Impact of IL-1 and IL-6 gene polymorphisms on the severity of COVID-19

Authors:

S. Verma[1], S. Verma[2], M. Abbas[2], Z. Siddqi[2], F. Mahdi[2]; [1]Univ. of Lucknow, Lucknow, India, [2]Era Univ., Lucknow, India

Abstract:

Background: Genetic variations among individuals significantly contribute to the heterogeneity in COVID-19 severity. Single nucleotide polymorphisms in *IL-1* and *IL-6* genes impact COVID-19 severity by increased expression and higher immuno-inflammatory responses. Specifically, certain *IL-1* polymorphisms may amplify cytokine release, while *IL-6* variants can enhance inflammatory signaling pathways. This intensified inflammation can contribute to the development of cytokine storms and lead to severe respiratory distress, pneumonia, and multi-organ failure in COVID-19 patients. **Methodology:** The current study comprised 242 COVID-19 patients. Genotyping was done using PCR-RFLP and validated by Sanger sequencing. The statistical analysis was carried out with SPSS (version 21). **Results:** The frequency of the 'T' allele of the +3953C>T polymorphism in *IL-1* was higher among mild cases compared to severe cases, indicating significant protection against severe COVID-19 (P=0.001). For the *IL-6* -174G>C gene polymorphism, individuals with the GC genotype showed approximately a 4-fold higher risk of severe disease (P <0.001). Linkage disequilibrium analysis between *IL-1* +3953C>T and -511C>T revealed that individuals with the CT* haplotype had a higher risk of severe disease (P=0.001). In the haplotype analysis of *IL-6* -174G>C and IL-6 -597G>A polymorphisms, the AC* haplotype was significantly associated with COVID-19 severity (P=0.034). **Conclusion:** These findings suggest that interindividual genetic variations significantly contribute to the heterogeneity in COVID-19 severity. Identifying these genetic factors can enhance our understanding of the disease mechanisms and guide personalized treatment strategies.

Session Title: Omics Technologies Poster Friday Session

Board 1138F: Improving Transcriptome Variant Detection Using PacBio Isoform Sequencing

Authors:

A. Carroll[1], S. Yin[2], E. Tseng[3], K. Shafin[1], P-C. Chang[1], D. E. Cook[1]; [1]Google, Mountain View, CA, [2]Verily, Mountain View, CA, [3]Pacific BioSci.s, Menlo Park, CA

Abstract:

Long-read sequencing from Pacific Biosciences (PacBio) offers unique advantages to characterize transcriptomes, due to the ability of individual reads to fully cover a transcript. Long-read transcriptomes also enable variant detection, which can determine the effect of variants on allele-specific expression or identify compound-heterozygosity in genetic disease.
DeepVariant is a deep learning variant detection method with models optimized for PacBio's DNA sequencing. Prior work (de Souza et al. 2023) demonstrated that existing DeepVariant models trained with DNA data could achieve high accuracy on transcriptome sequencing with preprocessing which manages splicing representations.
However, there are challenges specific to transcriptome sequencing, including splicing, base modifications, and variable and allele-specific expression. We hypothesized that training DeepVariant with PacBio's transcriptome sequencing data would improve variant detection accuracy.
We trained a custom DeepVariant model using transcriptome sequencing of Genome in a Bottle samples (HG002, HG004, HG005) of multiple cell lines from the Revio and Sequel II instruments. For evaluation, we set chromosome 20 aside from training to use as an independent evaluation set. We use the preprocessing identified in (de Souza et al. 2023) for both baselines and the custom model.
Comparing the custom DeepVariant model to the DeepVariant v1.6 release trained on DNA data, for exons covered at 10x or higher, we observe improvements in SNP Precision (0.844 -> 0.969), SNP Recall (0.944 -> 0.948), SNP F1 (0.891 -> 0.958), Indel Precision (0.704 -> 0.949), Indel F1 (0.757 -> 0.869), with a minor decrease in Indel Recall (0.819 -> 0.802).
We make this model available in an open source manner, to enable the community to more accurately detect variants in their PacBio transcriptome sequencing.

Session Title: Omics Technologies Poster Friday Session

Board 1140F: InterSpatial: Bi-directional Integration of Single Cell RNA-seq and Spatial Transcriptomic Data

Authors:

T. Majumder, C. Chan, J. Xie; Duke Univ., Durham, NC

Abstract:

To understand the interactions and dynamics between cells, it is essential to integrate the spatial information of gene expressions at the single-cell level. Existing spatial transcriptomic methods primarily focus on gene expression patterns across different locations within a tissue. However, our novel approach provides a spatial probability distribution for each cell within the single-cell data across the tissue space. By explicitly modeling probability distributions, we can compare our results with existing methods of cell type distribution and explore a broad scope for computing cell-cell interactions. We use the spatial coordinates of the spots to compute the Wasserstein distance between the spatial distributions of two different cells, ultimately forming a network that identifies co-localized cell types. We demonstrate our method's effectiveness on several datasets. First, we integrate spatial transcriptome and scRNA-seq data from human lungs to locate senescent and other potentially pathological cells in lungs affected by pulmonary fibrosis. We also identify the spatial distributions of different cell types and the co-localization patterns between immune cells and pathological cells. Based on the proximity of immune cells to pathological cells, we identify the receptors activated in the immune cells in response to the overexpression of corresponding ligands in the pathological cells. Secondly, we analyze the co-localization between excitatory and inhibitory neurons in the middle temporal gyrus of patients with Alzheimer's disease. Thirdly, we combine scRNA-seq and Visium data from human hearts affected by myocardial infarction to characterize the interactions between myeloid cells and their neighboring cells. In summary, our novel method provides a more comprehensive understanding of cell-cell interactions and dynamics in a pan-tissue setup, unlocking new insights into the functioning of biological tissues.

Session Title: Omics Technologies Poster Friday Session

Board 1141F: Is there anything in the swamp? A credibility assessment of biologic signals from non-secreted low-abundance proteins measured in human plasma.

Authors:

J. Dennis, J. Gogain, C. Paterson, J. Kuzma, J. Wommack, E. Troth, S. Williams; Standard BioTools, Boulder, CO

Abstract:

Background: Proteomic multiplexing has expanded and now includes measurement of thousands of non-secreted low-abundance proteins. Because the mechanism for the presence of these non-secreted proteins in plasma is unclear, the biological significance of this fraction of the proteome has been questioned. We present here methods to evaluate this question by seeking to determine: 1) if proteogenomic associations are present in the non-secreted, low abundance fraction of the proteome, and 2) if this fraction contains other biologically relevant information.
Methods: Measurement of cis-pQTLs demonstrates true signal detection and correct reagent binding, therefore we surveyed peer-reviewed publications to determine the number of cis-pQTLs identified on the SomaScan® 5K or 7K assay. Of those, we calculated the proportion of cis-pQTLs in the non-secreted, low-abundance fraction of the proteome. To demonstrate physiological relevance, we used data from ~15,000 participants in the European Prospective Investigation into Cancer and Nutrition (EPIC) study assayed on SomaScan 7K. We determined what proportion of the significant protein biomarkers associated incident lung or breast cancer were in the non-secreted low-abundance fraction. We also assessed if these proteins were shared in mouse xenograft models of the same cancers.
Results: Roughly 2,800 cis-pQTLs identified by SomaScan 5K or 7K assay have been reported in peer-reviewed publications to date. There are ~5,000 non-secreted low-abundance proteins on the SomaScan 7K menu. 30% (n=1494) of this fraction of the menu contains a cis-pQTL, vs. ~60% of secreted proteins. Of the statistically significant biomarkers for breast and lung cancer risk identified in the EPIC dataset, 68% and 60% respectively were found in the non-secreted low-abundance fraction. Furthermore, 21% and 14% of the human breast and lung cancer risk biomarkers from this fraction demonstrated statistically significant signaling in mouse xenograft models of the same cancers.
Conclusion: Non-secreted proteins comprise a greater proportion of the proteome than known secreted proteins. We show here that despite unclear mechanisms relating to their presence in plasma, they contain important biologic information, although proteogenomic associations are somewhat attenuated. Reassurance of correct protein binding in this fraction arises directly from cis-pQTLs and circumstantially from shared biomarkers of future cancer diagnosis in humans and mouse xenograft models. Proteomic technologies that can be scaled while maintaining sensitivity and precision will likely continue to yield useful discoveries.

Session Title: Omics Technologies Poster Friday Session

Board 1142F: Leveraging machine learning to predict Parkinson's disease development using proteomics data from presymptomatic patients ⭐

Authors:

S. Nieves Rodriguez, A. Doostparast Torshizi; Population Analytics & Insights, Data Sci. Analytics & Insights (DSAI), AI/ML Analytics & Insights, Janssen Res. & Dev., LLC, Spring House, PA

Abstract:

Background: Parkinson's disease (PD) is a progressive neurodegenerative disease with non-motor and motor manifestations resulting from dopaminergic neuron loss in the substantia nigra. Although non-motor symptoms can emerge up to ~12-20 years prior to motor symptoms, predicting a future PD diagnosis using non-motor symptoms, and halting the disease progression before irreversible neuronal damage occurs, remains challenging. Machine learning (ML) methods have been successful at predicting disease onset across various diseases and data modalities. Previous studies have assessed the efficacy of different models to predict PD development using phenotype and genotype data. **Methods:** The UK Biobank Pharma Proteomics Project (UKB-PPP) constitutes an opportunity to further explore the ability of ML models to predict disease using high-throughput proteomic data from plasma from up to 14 years before diagnosis, a time when early PD non-motor symptoms are estimated to have already begun. Using UKB-PPP, we analyzed 2,937 proteins from plasma of PD cases (incident=537, prevalent=101) and controls (n=15,522) to first identify a small panel of proteins as potential diagnostic biomarkers, and second assess the ability of ML models to predict disease up to 14 years before diagnosis using a small set of circulating plasma proteins. We validated our findings using the Parkinson's Progression Markers Initiative (PPMI) proteomic data consisting of 1,472 proteins from longitudinal plasma samples of PD cases (incident=142, prevalent=174) and controls (n=352). **Results:** Differential protein expression analysis identified 446 unique proteins dysregulated in at least one incident category (3-5, 6-9 and 10-14 years before diagnosis) compared to controls. After feature selection, a subset of 23 proteins predicted PD with an AUC of 0.777 in incident and 0.795 in prevalent cases. Adding age, sex and family history as additional features increased the model performance to 0.814 and 0.812 in incident and prevalent, respectively. 16 of the 23 identified proteins were also found in the PPMI dataset, and this reduced protein set predicted PD in PPMI with an AUC of 0.760 in incident and 0.608 in prevalent for first visit samples, and 0.691 in incident and 0.62 in prevalent for all visits. **Conclusions:** PD signatures can be observed in plasma up to 14 years prior to diagnosis, and the identified dysregulated proteins could provide insights into PD development.

This study is an example of how ML models can be applied to high throughput plasma proteomic data from pre-symptomatic patients to predict future disease onset.

Session Title: Omics Technologies Poster Friday Session

Board 1143F: Long-read sequencing to diagnose Mendelian diseases

Authors:

Y. Mostovoy[1,2], A. Sanchis-Juan[2], S. L. Stenton[3], M. O'Leary[2], S. Negi[4], B. Weisburd[3], S. Huang[2], W. J. Rowell[5], C. T. Saunders[5], E. Walker[2], S. Hall[2], K. Garimella[2], S. Dodge[2], S. Gabriel[2], N. Lennon[2], H. Brand[1], K. H. Miga[4], A. O'Donnell-Luria[6,2], B. Paten[4], M. Talkowski[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Broad Inst. of MIT and Harvard, Boston, MA, [4]Univ. of California Santa Cruz, Santa Cruz, CA, [5]Pacific BioSci.s, Menlo Park, CA, [6]Boston Childrens Hosp., Boston, MA

Abstract:

The introduction of short-read sequencing (SRS) has revolutionized clinical diagnosis of rare diseases, yet over 50% of individuals with a suspected monogenic condition lack a genetic diagnosis after undergoing clinical sequencing. Long-read sequencing (LRS) offers the potential for transformative advances in the discovery and characterization of genomic variation, as well as the assembly of phased haplotypes and functional signatures for interpretation of recessive variation and *cis*-regulatory effects. However, while there have been many individual examples of new clinically relevant variants discovered from LRS, the diagnostic yield of LRS over SRS remains unknown. Here, we performed a systematic assessment of the utility of LRS for rare disease diagnosis as part of the Broad Institute Center for Mendelian Genomics. We applied LRS to two cohorts of samples with suspected monogenic disorders (including neurodevelopmental, skeletal, and muscular conditions) for which SRS had been performed to capture SNVs, indels, and structural variants (SVs) with no resulting genetic diagnosis. One cohort included 61 individuals from 20 families sequenced with Oxford Nanopore (ONT), and the second comprised 200 individuals from 83 families sequenced with PacBio (PB) HiFi. Median sequencing depth was >30x for both ONT and PB. For each cohort, a comprehensive array of methods - including detection of SNVs, indels, SVs, short tandem repeat (STR) expansions, and methylation outliers - was applied to capture the largest possible range of variation. Variants were filtered using SRS and LRS population references from gnomAD, All of Us, HPRC, HGSVC, and the 1000 Genomes Project. LRS analyses optimized for sensitivity captured a median of 22k (ONT) and 29k (PB) SVs per sample, approximately a two-fold increase over SRS (~11k SVs per sample). The LRS data

included 93-106 additional rare gene-disrupting SVs per cohort in constrained genes. LRS-detected pathogenic variants included a tandem duplication spanning multiple exons of *ENG* in an individual with hereditary hemorrhagic telangiectasia, as well as a *de novo* stop-gain SNV in *CASK* in an individual with microcephaly and global developmental delay in a region where SRS had allelic dropout. Additionally, LRS provided clear phasing of a compound heterozygous case and uniquely resolved the breakpoints of multiple SVs. Analyses are ongoing to complete all samples and derive a final diagnostic assessment of LRS by technology, analytic approach, and phenotype in comparison to SRS, but LRS provides a more comprehensive picture of genomic variation and represents a potentially transformative advance for rare disease diagnosis.

Session Title: Omics Technologies Poster Friday Session

Board 1144F: Managing multi platform long read sequencing data at scale

Authors:

T. Shaffer, J. Smith, C. Hong, T. Kolar, C-L. Wei; Northwest Genomics Ctr., Dept. of Genome Sci., Univ. of Washington, Seattle, WA

Abstract:

Long-read sequencing technologies allow access to difficult regions of the genome and are increasingly being used in the study of human genetic disorders. The type and size of data, multiplicity of platforms, and frequency at which platform vendors update instruments, chemistries, and software makes analyzing a large number of samples generated under these conditions challenging. We describe how model selection affects base and variant calling and introduce a novel Nextflow pipeline we have developed to address these issues that can be run locally on a high performance cluster or cloud platform. This workflow utilizes a suite of tools for base calling and determination of modification status, read alignment, and variant calling of SNVs, indels, and structural variants. In addition, it generates numerous QC metrics that are useful to assess data quality. We have created a scalable, automated system to manage sequencing run setup through initial quality analysis, variant calling workflows, and, in the case of University of Washington Center for Rare Disease Research (UW-CRDR) samples, import into Seqr for variant analysis. This allows us to deliver a consistent data product tailored to specific project needs while making efficient use of cpu, gpu, and storage resources.

Session Title: Omics Technologies Poster Friday Session

Board 1145F: Metabolomic Analysis of a Rare Disease Cohort Identifies Disease-Specific Diagnostic Panels

Authors:

A. Kennedy[1], K. DeBalsi[1], R. Gunst[1], K. Ounap[2], M. Ilisson[2], K. Reinson[2], K. Oja[2], A. Evans[1], L. Sommerville[1]; [1]Metabolon, Morrisville, NC, [2]Univ. of Tartu, Tartu, Estonia

Abstract:

Collectively rare diseases (RDs) affect 250-450 million people worldwide. Genomic sequencing is the first line diagnostic for most RDs, yet our incomplete understanding of rare variant associations limits this approach, resulting in delayed diagnosis. Our goal is to shorten the time to diagnosis for RDs by including metabolomics in diagnostic testing. To this end, we analyzed the plasma metabolomic profiles of 133 persons suspected of having a RD but without a conclusive diagnosis or obvious genetic causality. Metabolite raw peak intensities were normalized, log-transformed, and Z-scored. Metabolite concentrations > 97.5th percentile or < 2.5th percentile of the normal range were flagged as potential diagnostic markers. 1237 compounds were detected in the cohort, covering 9 super pathways and 120 sub-pathways. Every patient demonstrated metabolic abnormalities. For some cases, the clinical significance of these changes remains an active area of research. For others, metabolic profiles suggested disease. For example, one patient who presented with a ventricular septal defect and hypertrophic cardiomyopathy showed elevated levels of ribitol, arabonate/xylonate, arabitol/xylitol, and erythritol, key metabolite markers of transketolase deficiency. Transketolase catalyzes two reactions in the pentose phosphate pathway, and its deficiency typically manifests as congenital heart abnormalities, suggesting this as a possible diagnosis. Another patient who presented with developmental delay had increased N6,N6,N6-trimethyllysine and decreased deoxycarnitine. This biomarker signature suggests deficiency of trimethyllysine hydroxylase epsilon (TMLHE), an enzyme critical to carnitine biosynthesis. TMLHE deficiency was supported by reduced levels of other carnitine compounds downstream of TMLHE. Furthermore, TMLHE deficiency is linked to autistic behaviors, which can include developmental delay, lending further credence to this diagnosis. Finally, a third patient who presented with congenital malformation of the cerebral hemisphere had reduced levels of certain metabolites in the tryptophan pathway and TCA cycle that are associated with Rett syndrome, an inherited disease that affects *FOXG1*, which is a transcription factor critical for brain development. The diagnosis of Rett Syndrome suggested by the metabolic signature and clinical symptoms could be pursued by targeted genomic profiling. While further studies are needed to confirm the suspected

diagnoses, this study shows the utility of global metabolomics as a diagnostic tool for RDs, which could complement genomics testing and shorten the time to diagnosis.

Session Title: Omics Technologies Poster Friday Session

Board 1146F: mitoScope: A New efficient de novo workflow for processing long Reads for mitochondrial analysis ⭐

Authors:

J. Smith, C. Wong, C. Frazar, C. Jacques, M-F. Huang, P. Nielsen, B. Hudson, C-L. Wei; Northwest Genomics Ctr., Dept. of Genome Sci., Univ. of Washington, Seattle, WA

Abstract:

Because mtDNA sequences are rich in repeats and segmental duplications, it can be challenging to accurately catalog the complete variation spectrum inan individual using a referenced-based alignment. To address this issue, we have developed a long read-specific de novo workflow named mitoScope to assemble and quantify the amplified circular genomes of homologous sequences along with their methylation profiles from the All of Us Research Program (AoURP) deep coverage long-read genomes. The mitochondrial genome (mtDNA) is 16.5 Kb in length and contains several protein-coding genes as well as genes encoding rRNAs and tRNAs. Structural variants and mutations in mitochondrial genomes are known to be associated with a variety of common diseases, particularly age-related dysfunctions like Parkinson's, Alzheimer's, and human cancers. mitoScope addresses the existing limitations for de novo genome assemblers, which are computationally-intensive, solely focused on linear genome assembly, and miss subpopulations present in the sample. mitoScope uses long reads as input, extracts mtDNA reads using k-mer analysis, and performs de novo metagenome assembly using meta-flye followed by polishing to generate consensus circular mtDNAs and their subpopulations. With complete mtDNA phasing, we can compute the frequencies of multiple SNP co-presences. For instance, multiple SNPs in both MT-RNR1 and MT-TS1 (encodes 12S mitochondrial ribosomal RNA and transfer RNA serine), which are 7kb apart have been associated with hearing loss that can be detected from individual DNA molecules by long-read sequencing. We will demonstrate the utility of mitoScope in detecting complex variation, methylation, circular genome subpopulations, and its application in evaluating the mutational landscape in the AoURP long-read data set.

Session Title: Omics Technologies Poster Friday Session

Board 1147F: Multi-Omics Data from Short-Duration Space Travelers

Authors:

H. Doddapaneni[1], Z. Momin[1], Q. Xiang[1], A. Santhanam[1], K. Walker[1], Q. Wang[1], z. khan[1], A. Krishnavajhala[1], C. A. Jones[2], M. Basner[2], M. Masternak[3], M-C. Gingras[1], J. Wu[4], E. Urquieta[4], D. M. Muzny[1], R. A. Gibbs[1]; [1]Human Genome Sequencing Ctr. (HGSC), Baylor Coll. Med., Houston, TX, [2]Dept. of Psychiatry at the Univ. of Pennsylvania Sch. of Med., Philadelphia, PA, [3]Sch. of BioMed. Sci., Univ. of Central Florida, Orlando, FL, [4]Ctr. for Space Med. and Translational Res. Inst. for Space Hlth., Baylor Coll. Med., Houston, TX

Abstract:

Travel to space can have significant effects on the human body. The changes in gravity, radiation exposure, and extreme g-force transitions can lead to various physiological and behavioral impacts. The extent of these changes, especially those induced during short duration missions (1 month or less) and how they can inform changes during long-term space flights (>1 month), are poorly understood. Such information is highly desirable to help guide the development of personalized countermeasures for short duration flights and to build comprehensive assessment platforms for future, longer-duration human spaceflights, as well as certain terrestrial applications. The Genomics and Space Medicine project ('Space Omics') at the BCM-HGSC is designed to collect biospecimens from participants traveling on commercial space missions (duration, under 1 month), for analysis using, an array of standardized multi-omics assays. Biospecimens (blood, urine, saliva, skin swabs and stool) are collected from multiple pre- and post-flight time-points for biobanking and for multi-omics data generation. This project, in 18 months, has enrolled ten participants from three private spaceflights and generated data for 6 participants, from two of these flights. Multi-Omics analysis of the 344 samples from the private astronaut mission showed that more transcripts, proteins and biochemical pathways were down regulated in the samples immediately upon return compared to the pre-flight samples. Multiple biochemical pathways overlapped between proteomics (Olink Explore HT, 5K high-throughput panel) and transcriptomics (bulk RNA-Seq) datasets. Single-cell transcriptome analysis of the samples collected 30 days pre-flight and R0 (day of return) showed that in the post flight samples, the proportion of CD14+ Monocytes, Plasmacytoid dendritic cells decreased, while Naive CD4+ cells and NK cells increased, in agreement with previously reported results from long-term flights underscoring the significance of this data. An added emphasis of this project is to generate ground-based data to mimic stressors experienced during space flight, as well as include appropriate age and gender matched controls to isolate 'noise' in data. In one such experiment, blood samples from two volunteers were collected before and after a centrifugation session at the NASTAR center in Philadelphia, while samples collected from a third volunteer not in this session will serve as an independent control for this stressor.

Matching bulk RNA-Seq and proteomic datasets will be generated from these samples to understand the stress experienced during the launch and re-entry of spaceflight vehicles.

Session Title: Omics Technologies Poster Friday Session

Board 1148F: Multiplexed full-length single-molecule sequencing of the human mitochondrial genome and its application for rare neuromuscular diseases and chronic lymphocytic leukaemia ⭐

Authors:

I. Gut[1], I. Keraite[1], D. Canevazzi[1], C. Frias-Lopez[1], I. Paramonov[1], R. Horvath[2], F. Nadeu[3], E. Campo[3], M. Gut[1]; [1]Centro Natl. de Análisis Genómico (CNAG), Barcelona, Spain, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]IDIBAPS, Barcelona, Spain

Abstract:

Mitochondria are organelles found in most eukaryotic cells containing several copies of a circular genome encoding for vestigial functions of what once used to be a free-living organism. Within the human population, heteroplasmy of the mitochondrial genome is common and can have substantial impact on human health depending on the variant position, frequency and endogenous or exogenous factors. Accurate detection and quantification of such variants is important for diagnosis, genetic counselling, and treatment of mtDNA diseases.
Methods to reconstruct the mitochondrial DNA (mtDNA) sequence using short-read next-generation sequencing come with an inherent bias due to amplification by PCR and mapping. They can fail to determine the phase of variants, to capture multiple deletions and to cover the mitochondrial genome evenly. Here we describe a method to target, multiplex and sequence at high coverage full-length human mitochondrial DNA as native single-molecules, utilizing the RNA-guided DNA endonuclease Cas9. Combining Cas9 induced breaks, that define the mtDNA beginning and end of the sequencing reads, as barcodes, we achieve high demultiplexing specificity and delineation of the full-length of the mtDNA, regardless of the structural variant pattern. This is the first demonstration of reading an entire human chromosome, albeit the smallest one, completely in a single read. Long-read sequencing data is analysed with our custom-developed software, baldur, which efficiently detects single nucleotide heteroplasmy to below 1%, physically determines phase and accurately disentangles complex deletions. We show the application of our method for the investigation of the mitochondrial genome of patients suffering from rare neuromuscular disorders where we observe a progressive shortening of the mitochondrial genome. We also measure the accumulation of somatic point mutations in the mitochondrial genome in B cells of patients as they progress from chronic

lymphocytic leukaemia to Richter Transformation (RT). We also identify a RT specific mutation pattern.

Session Title: Omics Technologies Poster Friday Session

Board 1149F: Net2Onto: Leveraging molecular networks and ontology structure for predicting novel genes that play a role in understudied functions, traits, and diseases

Authors:

K. Manpearl[1], L. Renming[2], A. Krishnan[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Carnegie Mellon Univ., Pittsburgh, PA

Abstract:

Even a quarter century after sequencing the human genome, only about half of the ~20k protein-coding genes have any experimental evidence linking them to at least one biological function and only a quarter to any phenotype and disease. Rapidly discovering these links for un(der-)characterized genes is critical for closing enormous gaps in our current knowledge about the molecular basis of physiological processes, disease etiology, and clinical outcomes. To address this need, we and others have developed machine learning (ML) methods that can capture interaction patterns in molecular networks to predict novel genes associated with a variety of function, trait, and disease contexts. However, the reliance of these methods on previously-annotated genes severely limits their ability to make predictions for contexts with no or very few already known annotations. Methods such as clusDCA and DeepGOZero overcome this limitation by using the relationships between functions captured in the Gene Ontology to leverage information about well-studied functions to improve gene prediction performance for under-studied functions and, further, allow for predicting genes for functions with no known gene annotations. Here, we present Net2Onto, a general-purpose neural network framework that extends these methods to make novel gene discoveries about un(der-)studied contexts by learning a mapping from representations of genes in a molecular interaction network to an embedding space with terms in a biomedical ontology. Our framework combines two desirable properties of previous methods: i) like clusDCA, it uses a representation of the ontology graph structure, ii) like DeepGOZero, it uses a neural network to learn a mapping between genes and terms, which allows for the model to capture nonlinear relationships. Net2Onto extends the application of this approach to the Mondo disease ontology and the Phenio phenotype ontology in addition to the Gene Ontology biological processes. Net2Onto achieves performance comparable to baseline network-based ML models, and through rigorous

evaluations where all annotations to certain terms are completely held-out, we show that it has the capability to make zero-shot predictions for terms that were not seen during training. Using this framework, we compared various ways to embed the network and the ontology, and explored how term properties correlated with model performance. Overall, Net2Onto is a powerful and general framework for leveraging networks and ontologies to discover novel genes related to functions, phenotypes, and diseases.

Session Title: Omics Technologies Poster Friday Session

Board 1150F: New Modular Hands-On Training Resources for Nextflow

Authors:

G. Van der Auwera, **K. Brewer**, C. Hakkaart, M. Ribeiro-Dantas; Seqera, Barcelona, Spain

Abstract:

Nextflow has emerged as a powerful tool for enabling reproducible and scalable computational workflows in life sciences research. However, despite its potential, mastering Nextflow can pose a steep learning curve for newcomers. Our traditional training resources provided comprehensive content, but their structure and pacing could feel intimidating to the point of being overwhelming.

In response to this challenge, we have developed a new set of modular hands-on training resources that offer a practical and accessible pathway for researchers to build their skills progressively as they develop proficiency in Nextflow.

Every module is centered on hands-on exercises designed to reinforce learning and encourage active engagement. Participants are guided through practical tasks, such as building and executing basic workflows, troubleshooting common errors, and customizing workflows to specific research needs. In addition, the new modules focus real-world research scenarios, ensuring relevance and applicability.

The new training materials are deposited in GitHub and can be accessed for self-service use through our training portal. A preconfigured Gitpod environment makes it possible to work through the exercises without having to manage any software dependencies. Educators are welcome to use the materials as part of trainings they deliver online or in person.

By combining structured learning modules with interactive exercises and real-world applications, we aim to empower researchers with the skills and confidence to harness the full potential of Nextflow in their scientific endeavors.

Session Title: Omics Technologies Poster Friday Session

Board 1152F: Optimal resolution of differential expression of long read RNA-seq using transcript trees and data thinning

Authors:

J. Landis[1], N. P. Singh[2], R. Patro[2], M. I. Love[1]; [1]UNC-Chapel Hill, Chapel Hill, NC, [2]Univ. of Maryland, College Park, MD

Abstract:

Differential expression analysis of long read RNA-seq data provides high confidence transcript-level inference, unavailable with previous short read assays. Due to the complexity of alternative splicing of genes, drawing meaningful biological conclusions across transcripts and across samples is challenging in current pipelines as differential signals may be spread out among many transcript isoforms with similar structure. We have recently developed efficient algorithms for building transcript trees, where similar transcripts are merged together iteratively, culminating at the gene level. We seek to optimally aggregate long read data using transcript trees to provide parsimonious summaries of multi-sample differential expression results. Here we define optimality as the highest merge point within the transcript tree while preserving significant differences of descendent nodes. We provide an R/Bioconductor pipeline that takes as input a multi-sample long read RNA-seq count dataset with known transcriptome and reports differential expression at the level of transcript groups such that key differences are preserved while redundant information is minimized. We show within simulated experiments that implementing a data thinning approach controls for false positives on the final analysis set after optimal aggregation has been performed.

Session Title: Omics Technologies Poster Friday Session

Board 1153F: Optimizing recall of RNA-seq outliers in heterogeneous rare disease cohorts

Authors:

J. Major-Mincer, A. Quinlan; Univ. of Utah, Salt Lake City, UT

Abstract:

Rare diseases collectively affect millions of individuals worldwide, yet merely half of cases are resolved. This diagnostic gap is due in part to the prevalence of variants of uncertain

significance (VUS), which lack adequate prior evidence connecting them to a disease phenotype. As a result, computational tools such as CADD and SpliceAI may inadequately predict VUS consequences due to a paucity of prior training data on relevant genomic features. In principle, RNA sequencing (RNA-seq) can address these computational shortcomings by providing a direct transcriptional readout of a VUS's impact on total transcript and isoform abundance. However, in rare disease studies an important challenge is identifying a patient's expression as an "outlier" with respect to a reference RNA-seq population that is often small yet heterogeneous in age, sex, phenotype, and technical processing. Such heterogeneity attenuates the power to detect true outliers in expression and splicing. Recognizing these challenges, computational approaches have been developed to control noise in rare disease expression data, including Z-scores and the OUTRIDER autoencoder normalization and expression outlier framework. Nonetheless, we argue that inadequate sample sizes and heterogeneous populations remain persistent challenges. Preliminary analyses with the Genotype-Tissue Expression (GTEx) dataset support this claim, as we found a weaker correlation in sample outlier statistics when comparing bootstrapped reference populations consisting of males and females against those consisting of a single, uniform sex. We then hypothesize that biological and technical confounders cause an increase in false negative expression outlier calls even after correction with existing approaches. To test this hypothesis, we use whole blood RNA-seq expression data from the GTEx dataset. We selected whole blood as the tissue of interest as it is clinically accessible and commonly used in rare disease diagnostics. We simulate outliers in expression at various magnitudes of deviance from the raw population mean using a negative binomial quantile-based approach. Using the outlier-detection methods described, we characterize the change in recall of simulated expression outliers given variant or invariant biological confounding variables. Additionally, we investigate the effect of sample size on predicted outlier genes by resampling bootstraps of populations with 50 to 150 samples. We anticipate our findings will empower confident utilization of RNA-seq in rare disease diagnostics through reasoned and considerate cohort structure.

Session Title: Omics Technologies Poster Friday Session

Board 1154F: Parent-of-origin-aware genomic analysis of hereditary cancer patients for parent-of-origin assignment to pathogenic variants without parental data

Authors:

V. Akbari[1,2], **L. Cordova**[3], T. Leung[4], K. Dixon[1], K. O'Neill[1], V. Hanlon[4], A. T. Roston[2], K. Wong[3], A. Sharma[3], E. Cheung[3], Y. Shen[1], J. Senz[5], Y. Wang[4], D. Chan[4], A. Fok[3], J. Nuk[3], Q. Hong[6], S.

Bilobram[1], R. Coope[1], E. Chuah[1], S. Chan[1], H-W. Lee[1], Y. Zhao[1], M. Bala[1], K. Mungall[1], A. Mungall[1], R. Moore[1], D. Regier[7,6], A. Virani[2], L. Lefebvre[2], F. Feldman[8], M. Marra[2,1], S. Sun[3], S. Yip[9], P. Lansdorp[2,4], S. Jones[1,2], K. Schrader[2,3,5]; [1]Canada's Michael Smith Genome Sci. Ctr., BC Cancer, Vancouver, BC, Canada, [2]Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [3]Hereditary Cancer Program, BC Cancer, Vancouver, BC, Canada, [4]Terry Fox Lab., BC Cancer, Vancouver, BC, Canada, [5]Dept. of Molecular Oncology, BC Cancer, Vancouver, BC, Canada, [6]Sch. of Population and Publ. Hlth., Univ. of British Columbia, Vancouver, BC, Canada, [7]Canadian Ctr. for Applied Res. in Cancer Control, Cancer Control Res., BC Cancer, Vancouver, BC, Canada, [8]BC Cancer, Vancouver, BC, Canada, [9]Pathology and Lab. Med., Univ. of British Columbia, Vacnouver, BC, Canada

Abstract:

Background: Most genome sequencing approaches detect genomic variants but overlook the haplotype-specific information of heterozygous variants. A few techniques enable haplotype-specific detection of heterozygous variants; however, they do not provide parent-of-origin (PofO) of genomic variants in the absence of parental data. By combining chromosome-scale haplotyping with DNA methylation at imprinted regions, we recently developed Parent-of-Origin-Aware genomic analysis (POAga) that enables PofO assignment to human autosomal variants without parental data (Akbari *et al*., 2023, Cell Genomics). Information regarding parental segregation may be useful in risk management, variant curation, and facilitating cascade genetic testing in common Tier 1 hereditary cancer conditions such as Lynch syndrome and hereditary breast and ovarian cancer (HBOC) due to germline pathogenic variants in *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM, BRCA1*, and *BRCA2* genes. POAga is also useful in rarer syndromes with PofO effects such as hereditary paraganglioma and pheochromocytoma due to paternally inherited pathogenic variants in *SDHD* and *SDHAF2* and other genes predisposing to breast cancer and gastrointestinal malignancies, such as *ATM, PALB2,* and *CDH1*. To evaluate the analytical validity of POAga, we have applied this technology to samples from patients with known parental segregation of pathogenic variants in hereditary cancer genes. **Method:** Blood samples from carriers of pathogenic variants in hereditary cancer genes of differing age, sex, ethnicity, and cancer status with known parental segregation are currently being ascertained and subjected to Strand-seq and long-read sequencing to determine PofO under an REB-approved protocol. **Results:** Thus far 138 samples have been sequenced, with 138 pathogenic variants in *BRCA2* (n=25), *MLH1* (n=19), *SDHD* (n=19), *BRCA1* (n=17), *MSH2* (n=16), *PALB2* (n=11), *MSH6* (n=9), *PMS2* (n=8), *ATM* (n=6), *CDH1* (n=5), *SDHAF2* (n=2), and *EPCAM* (n=1). PofO could be assigned to 120 variants with 99.2% (119/120) concordance with the known segregation. PofO could not be assigned to 18 variants primarily due to runs of homozygosity impeding phasing or polymorphic imprinted methylation. We are exploring avenues to solve such cases, for example

by increasing sequencing coverage and including additional imprinted regions. **Conclusion:** Our results support the ability of POAga to accurately assign parent-of-origin using blood samples from hereditary cancer patients. Ongoing ascertainment and analysis will improve PofO assignment in failed samples and give better insights into the real-world accuracy of POAga and its potential clinical applications.

Session Title: Omics Technologies Poster Friday Session

Board 1155F: Phasing Accuracy of Long-Read Whole Genome Sequencing is Coverage and Reference Genome Dependent

Authors:

J. Fu[1], G. F. Frost[1], N. Damaraju[2], D. D'Onofrio[1], W. Gahl[1], D. R. Adams[1], M. Malicdan[1], D. E. Miller[3]; [1]NIH Undiagnosed Diseases Program, NHGRI, NIH, Bethesda, MD, [2]Inst. for Publ. Hlth.Genetics, Univ. of Washington, Seattle, WA, [3]Dept. of Lab. Med. and Pathology, and Div. of Genetic Med., Univ. of Washington and Seattle Children's Hosp., Seattle, WA

Abstract:

Introduction: Long-read sequencing (LRS) technologies have greatly improved our ability to detect large structural variants, variants located in difficult-to-sequence regions, and perform haplotype phasing, improvements particularly important in clinical settings. Although short-read sequencing (SRS) can phase variants using parental samples, it falls short in phasing de novo variants or variants in the absence of parental samples. Despite the potential for LRS adoption in clinical environments, haplotype error rate construction using common LRS tools has not been systematically evaluated. **Methods:** In this study, we utilized the well-characterized reference sample HG002 to calculate phasing error rates of variant pairs within OMIM genes. We assessed error rates across multiple Oxford Nanopore Technologies (ONT) chemistries and PacBio HiFi sequencing. Phasing accuracy was evaluated using consensus VCFs HG002, HG003, and HG004 from the Genome in a Bottle (GiAB) Consortium and a "Q100" VCF generated from the fully-assembled HG002 genome. All data was aligned to both GRCh38 and the CHM13 Telomere-to-Telomere reference genome. **Results:** The phasing error rate for data aligned to GRCh38 was below 5% at nearly all genomic coverage levels, and nearly 100 times lower when aligning to CHM13, with a progressive decrease in error with additional coverage for both reference genomes. For data aligned to GRCh38, the presence of one or both variants in a pair within a difficult-to-align region caused a stable 4% increase in error. However, for CHM13, the relative contribution of variants in such regions was minimal, likely because of the resolution of low complexity and duplicated loci. Analysis of ONT data showed that higher quality data does

not reduce the minimum phasing error rate when using CHM13 but does lower the coverage required to achieve that minimum error. Data aligned to GRCh38 showed no change in phasing error with improvements in base quality scores. Further, we found that phasing errors were concentrated to 0.5% of OMIM genes with phased variant pairs when aligning to CHM13, compared to being spread across 25% of OMIM genes in GRCh38 and CHM13. **Discussion:** Overall, our results indicate that LRS accurately phases variants from singleton cases, with a significant improvement in phasing accuracy across all data types when aligning to CHM13. Importantly, alignment to CHM13 concentrates phasing errors to a small fraction of genes; potentially allowing for the classification of genes into high and low confidence categories for clinical reporting.

Session Title: Omics Technologies Poster Friday Session

Board 1156F: Phasing of RNA isoforms reveals haplotype-specific expression

Authors:

C. Paisie[1], P. Audano[1], M. Kostich[1], A. Ferraj[2], P. Hallast[1], F. Yilmaz[1], Human Genome Structural Variation Consortium, C. R. Beck[2]; [1]The Jackson Lab., Farmington, CT, [2]UConn Hlth.and The Jackson Lab., Farmington, CT

Abstract:

A single human gene often encodes multiple RNA isoforms, with human genes having an average of at least eight isoforms. Alternative isoforms from the same gene can have different functions, and these differences have been shown to play a role in the development of disease. Changes in isoform expression or structure are also sources of genetic variation within populations, and have been associated with cancer progression. With the development of telomere to telomere, phased human genome assemblies, we now have the opportunity to understand how heterozygous genetic variation can affect transcription levels and isoform structure from an individual genome. Phasing of isoforms to their related genetic haplotype is crucial for understanding the relationship between genetic variation and both expression and splicing phenotypes as well as the function of possible causal variants in disease. Previous isoform phasing approaches have relied upon the use of parental genotypes to assign haplotypes as either paternal or maternal, and these genotypes are not always available. Therefore, there is a need for phasing strategies irrespective of parental genotypes. We combined 12 long-read RNA-sequencing (Iso-Seq) samples with a phased callset including structural variants (SVs) and small variants we developed with the Human Genome Structural

Variation Consortium (HGSVC). We ran lima and isoseq refine and then mapped the reads to GRCh38-NoALT (GRCh38) and T2T-CHM13v2.0 (T2T-CHM13) with minimap2. We then ran WhatsHap haplotag utilizing the single nucleotide variants (SNVs) from the phased callset. Using our approach, we are able to assign reads to their original haplotype and found that 61-64% of isoforms could be phased per sample with 17-28% unique to a single haplotype. We further identified four of nine known imprinted genes in lymphoblastoid cell lines expressed in a single haplotype. We are currently working to determine the effects of structural variants and other genetic features on haplotype-specific isoforms, and how this understudied variant class may also represent a source of novel isoform generation. With modern genomic sequencing data and accompanying isoform sequences, we are investigating the effects of haplotype-specific gene expression and identifying features affecting isoform structure.

Session Title: Omics Technologies Poster Friday Session

Board 1157F: Polygenic Risk Scores as Knockoffs for Deep Learning Analysis of Exposome Data: Unveiling Nonlinear Disease Risk Factors

Authors:

A. Sriram, H. Park, J. Carcillo, K. Kernan; Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Unraveling the intricate interplay between genetic and environmental factors is crucial for understanding the phenotypic variance observed in complex diseases. While genetic influences have been extensively studied, identifying modifiable environmental risk factors, collectively termed the exposome, is essential for developing effective interventions and preventive strategies. However, accurately estimating the risk associated with exposome features poses significant methodological challenges due to the complex interplay between genetic effects and other exposures. To address these challenges, we propose deepEXPOKE, a novel deep learning framework that integrates biological knockoffs to accurately estimate the effect size of exposome features. Leveraging polygenic risk scores (PRS) as the first biologically-driven knockoffs, deepEXPOKE controls for not only statistical properties but also genetic correlations, enabling the disentanglement of exposome effects from potentially highly correlated genetic influences. Applying deepEXPOKE to three diseases of significant public health concern: sepsis, Alzheimer's disease and related dementias (ADRD), and coronary heart disease, we successfully validated previous findings from carefully designed cohort studies while unveiling new insights into the complex interplay between exposome features and disease risk. For instance, our analysis identified previously established risk factors, such as Apolipoprotein A-1, HDL

Cholesterol, and Triglycerides, as significant contributors to sepsis incidence and mortality. Additionally, deepEXPOKE revealed existing associations, such as glucose, cholesterol, and hypertension, with coronary heart disease, as well as novel associations for ADRD risk, independent of genetic predisposition. Altogether, deepEXPOKE represents a groundbreaking advancement in exposome data analysis, enabling accurate estimation of exposure risk while accounting for potentially confounding genetic effects. This innovative approach paves the way for more precise risk stratification, targeted interventions, and personalized prevention strategies for complex diseases.

Session Title: Omics Technologies Poster Friday Session

Board 1158F: Predicting Loss- and Gain-of-Function in Ion Channel Variants Using Machine Learning

Authors:

J. Safer[1], P. Rissom[2], P. Yanez Sarmiento[2], A. Brunklaus[3,4], D. Balaz[5], R. Castelli[6], C. W. Coley[7,8], C. Depienne[9], A. L. George[10], E. Kurganov[11], C. Marini[12], R. Steensbjerre Møller[13,14], A. Moroni[6], J. Pan[11], B. Santoro[15], B. Y. Renard[2], S. Iqbal[1,16,17,18], H. Heyne[2,19,20]; [1]Broad Inst. of MIT and Harvard, Ctr. for the Dev. of Therapeutics, Cambridge, MA, [2]Hasso Plattner Inst., Digital Engineering Faculty, Univ. of Potsdam, Potsdam, Germany, [3]Sch. of Hlth.and Wellbeing, Univ. of Glasgow, Glasgow, United Kingdom, [4]Paediatric NeuroSci.s Res. Group, Royal Hosp. for Children, Glasgow, United Kingdom, [5]Univ. of Tübingen, Tübingen, Germany, [6]Dept. of BioSci.s, Univ. of Milan, Milano, Italy, [7]Dept. of Chemical Engineering, Massachusetts Inst. of Technology, Cambridge, MA, [8]Dept. of Electrical Engineering and Computer Sci., Massachusetts Inst. of Technology, Cambridge, MA, [9]Inst. of Human Genetics, Univ. Hosp. Essen, Univ. of Duisburg-Essen, Essen, Germany, [10]Dept. of Pharmacology, Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, [11]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [12]Child Neurology and Psychiatric Unit, Pediatric Hosp. G. Salesi, Azienda Ospedaliero Univ.ria delle Marche, Ancona, Italy, [13]Dept. of Epilepsy Genetics and Personalized Med., Danish Epilepsy Ctr., Dianalund, Denmark, [14]Dept. of Regional Hlth.Res., Univ. of Southern Denmark, Odense, Denmark, [15]Dept. of NeuroSci., Mortimer B. Zuckerman Mind Brain Behavior Inst., Columbia Univ., New York, NY, [16]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [17]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [18]Cancer Data Sci., Dana-Farber/Harvard Cancer Ctr., Boston, MA, [19]Hasso Plattner Inst., Mount Sinai Sch. of Med., New York, NY, [20]Finnish Inst. for Molecular Med. (FIMM), Univ. of Helsinki, Helsinki, Finland

Abstract:

Voltage-gated ion channels comprise a family of proteins essential in the evolution of rapid signal propagation in organisms. Missense mutations in genes encoding ion channels are drivers of a wide variety of genetic disorders, including neurological, cardiac, and muscular channelopathies. Here, we sought to characterize and model the functional effect of ion channel protein variants, which is a demanding problem, using biologically interpretable features and features learned by protein language models (PLMs). We collected data across the largest set of ion channels to date, spanning 145 ion channels across 11 gene families, including Nav, Cav, Kv, Kir, and HCN. We collected 146,349 population variants from gnomAD, 7,714 disease-associated variants from ClinVar and HGMD, and 3,435 reported loss-of-function (LOF) and gain-of-function (GOF) variants from electrophysiology and phenotypic studies in the literature. We then annotated the variants with protein features including structural features (e.g., secondary structures, pockets, protein-protein interactions, mutant effects) and sequence features (e.g., post-translational modifications, disorder prediction, amino acid conservation). Combining data from multiple ion channels increased the statistical power for our study and we identified motifs (log OR = -2.85, p = 7e-6), intramembrane regions (log OR = -2.48, p = 5.47e-22), and annotated domains (log OR = -1.69, p = 5.40e-5) as top features enriched in pathogenic LOF variants, and repeats (log OR = 0.82, p = 3.55e-21), helices (log OR = 0.66, p = 4.93e-12), and aliphatic residues (log OR = 0.54, p = 1.73e-9) as top features enriched in pathogenic GOF variants. Next, we collected embeddings for the ion channel variants from three protein language models (PLMs), leveraging these models for a LOF, GOF, and neutral variant comparison. We developed a new Python tool named ema (https://github.com/pia-francesca/ema) for comparing variant embeddings across different PLMs. Our analysis of ion channel variants using ema demonstrated the variance in learning by different PLMs, such as ESM1b and ESM2. Furthermore, we identified correlations between PLM embeddings of LOF and GOF variants with some protein features, but not all, indicating that embeddings in PLMs can be complementary to biological features for training an AI model for ion channel variant function prediction. At the time of submission, we are training and tuning an AI model for ion channel functional effect prediction leveraging both biological and learned features by PLMs. We aim to make this model available to the community for interpreting variants across all voltage-gated ion channels.

Session Title: Omics Technologies Poster Friday Session

Board 1159F: Proteomic Analysis Identifies Novel Protein Quantitative Trait Loci and Causal Proteins in Type 2 Diabetes Among Individuals of African Ancestry ⭐

Authors:

O. Soremekun[1,2], Y-C. Park[2], M. Tutino[2], A. Kalungi[1], W. Rayner[2], M. Nyirenda[1], S. Fatumo[3], E. Zeggini[2]; [1]MRC/UVRI $ LSHTM, Entebbe, Uganda, [2]Helmholtz Zentrum, Munich, Germany, [3]MRC/UVRI & LSHTM Uganda Res. Unit, Entebbe, Uganda

Abstract:

The development of proteomic technology has made it possible to quantify circulating proteins on a large scale. However, individuals of African ancestry remain largely underrepresented in proteomic studies. This study seeks to bridge this gap by investigating the proteomic landscape of Type 2 Diabetes (T2D) in an African population.We measured 2873 proteins using the Olink proximity extension assay in plasma samples from 163 prediabetic/diabetic individuals (cases) and 362 controls. We then performed differential expression analysis to identify proteins differentially expressed between the two groups. Additionally, protein-wide association analysis was conducted to identify protein quantitative trait loci (pQTLs) influencing protein levels. Identified pQTLs were followed up with colocalization and Mendelian Randomization analyses to investigate their causal association with T2D.Differential expression analysis identified 91 differentially expressed proteins (DEPs) with 60 overregulated and 31 underregulated. Notably, EDIL3 was the most significantly upregulated ($P_{adj} = 7.5*10^{-14}$), while LPCAT2 was the most significantly downregulated ($P_{adj} = 5.0* 10^{-14}$). We identified novel African-specific pQTLs with 17% cis-pQTLs and 23% trans-pQTLs not previously reported. Over 70% of the identified pQTLs replicated in an independent cohort (UK Biobank). Colocalization analysis revealed a shared T2D risk variant with pQTL for SIRPA protein (PP4 = 95.5%). Mendelian randomization identified 20 proteins causally associated with T2D, including SPINT3, SPINK6, and PNLIPRP1 which increased T2D risk, while SPINK4, SHMT1 and others showed protective effects. These study reveals causal proteins implicated in the pathogenesis of T2D which may be leveraged for personalized medicine tailored to African ancestry Individuals. This study has also demonstrated the utility and potential discovery opportunities in the inclusion of African ancestry individuals in large scale proteomic studies.

Session Title: Omics Technologies Poster Friday Session

Board 1160F: QuadST: A Robust and Powerful Approach for Identifying Cell–Cell Interaction-Changed Genes on Spatially Resolved Transcriptomics

Authors:

X. Song[1], G-C. Yuan[2], P. Wang[2]; [1]Duke-NUS Med. Sch., SINGAPORE, Singapore, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Recent advances in spatially resolved transcriptomics (SRT) have opened up avenues for identifying cell-cell interactions on spatially coordinated cells. We present QuadST, a novel statistical method for the robust and powerful identification of cell-cell interaction changed genes in SRT. QuadST constructs anchor-neighbor cell type integrated data matrix, models the association between gene expression and the cell-cell distance dynamically at different quantile levels, and contrasts association signals at different distance quantiles to identify the interaction changed genes. It eliminates the need for estimating interaction cell pairs, boosting the robustness and study power. In simulation studies, QuadST demonstrates resilience against measurement errors of cell-cell distance, unmeasured confounding factors, and gene-gene correlations, and provides higher power than existing methods, especially when interacting cell pairs cannot be perfectly specified. Application of QuadST to two mouse cortex datasets highlights the enrichment of synapse-related genes among significant interaction-changed genes in excitatory neurons.

Session Title: Omics Technologies Poster Friday Session

Board 1161F: Rare variant associations with the plasma proteome in individuals of South Asian ancestry

Authors:

A. Williamson[1], J. Carrasco Zanini[1], d. van heel[2], C. Langenberg[1,3], on behalf of the Genes and Health Investigators and Industry Consortium; [1]Precision Hlth.care Univ. Res. Inst., Queen Mary Univ. of London, London, United Kingdom, [2]Blizard Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [3]Computational Med., Berlin Inst. of Hlth.at Charité-Univ̈.tsmedizin Berlin, Berlin, United Kingdom

Abstract:

Proteogenomic studies provide insight into the genetic regulation of protein levels and mechanisms underlying disease-associated loci. To date, studies have predominantly focused on European-ancestry individuals and studied the impact of common variation. Here, we study the specific genetic architecture of individuals participating in the Genes & Health study and address the systematic underrepresentation of individuals of South-Asian ancestry in sequencing and multi-omic studies.
We performed plasma proteomic profiling using the Olink HT platform targeting 5,416 unique proteins in 1071 adults of Bangladeshi or Pakistani ancestry from the Genes & Health study with plasma samples and whole exome sequencing. We use Regenie for exome-wide gene-level collapsing analyses of rare variants (minor allele frequency <0.1%), grouped based on their predicted pathogenicity, and conditional analyses to assess the impact of common variants on all identified rare variant associations in cis, i.e. linked to genes encoding the corresponding protein product. We performed systematic cross-ancestry comparisons with rare variant associations detected in European ancestry individuals of UK Biobank (N = 49,736).
We identified 64 significant gene-protein pairs ($P < 5.32 \times 10^{-10}$), with at least one significant mask, 55 of which were in cis and 9 in trans, i.e. a gene distant from protein-encoding gene. A total of 52 identified unique cis gene-protein pairs were independent of a common lead cis-pQTLs ($P < 9.22 \times 10^{-12}$). These common lead cis-pQTL signals were identified for all but 21 rare cis-gene-protein pairs.Twelve of the gene-protein pairs (8 cis, 4 trans), have not previously been reported in gene-collapsing tests of the same protein target measured by the same technology in UK Biobank ($P > 1 \times 10^{-6}$), despite a sample size ~47 times larger.
In summary, our proteogenomic study of Genes & Health participants demonstrates the value and efficiency of rare variant proteomic studies in South-Asian ancestry individuals, likely due to enrichment for high impact homozygous variants in populations with parental relatedness. Studies of diverse ancestries are needed to provide insights in populations currently underrepresented in multi-omic research.

Session Title: Omics Technologies Poster Friday Session

Board 1162F: Rare-variant drop-out in biobank-scale multi-sample genotyping of next-generation sequencing data

Authors:

N. Larson, H. Yang, S. K. McDonnell, J. A. Frank, D. E. Elbarougy, P. C. Harris; Mayo Clinic, Rochester, MN

Abstract:

Recent technological improvements and reductions in cost have facilitated comprehensive next-generation DNA sequencing of biobank-scale cohorts, permitting powerful research into the contributions of rare variants to the risk of common and rare diseases. Such research typically requires aggregating thousands of single-sample genomic VCF files (gVCFs) into a multi-sample project-level VCF (pVCF) for downstream analysis. This joint genotyping effort presents many computational challenges and requires both harmonization of allelic representation and filtering of potential genotyping errors, the latter of which necessitates compromise in sensitivity vs. specificity for rare variation Herein, we explore how modern joint genotyping of biobank-scale data impacts rare variant call drop-out using data from Project Generation (PG), a partnership between Regeneron Genetics Center and Mayo Clinic that has generated whole-exome sequencing (WES) on approximately 116,000 subjects. Single-sample gVCF files were generated by DeepVariant, which were then aggregated into a PG-wide pVCF using the joint genotyping software GLnexus. A contributing disease cohort to PG included N = 459 subjects with polycystic kidney disease (PKD), an inherited form of chronic kidney disease, for which targeted panel sequencing results were available (including segmentally duplicated *PKD1* gene). To first enumerate general trends in variant drop-out rates between gVCF and pVCF call-sets, we examined variant call distributions among the PG PKD cohort for WES capture regions on chromosome 22. When restricting to variant calls of sufficient quality (DP≥12, GQ≥20), the median number of gVCF variants among subjects in the PKD cohort was 594 (range: 515 to 781), of which 2.6% were short insertion/deletion variants (INDELs). When cross-referencing the gVCFs to the pVCF, 47/459 (10.2%) had ≥1 missing variant (range: 0-2), and 4/54 variants (7.4%) were INDELs. To assess the impact of missing rare-variant calls, we compared the PG variant calls to the available clinical panel sequencing results. Among 306 of 459 PKD patients with a causal variant identified by targeted sequencing, 210 (69%) were corroborated by results in the pVCF. Among the N = 96 initially discordant *PKD1* and *PKD2* deleterious variant carriers that were pVCF negative, 80 (83%) were subsequently resolved when alternatively cross-referencing against the individual gVCFs, of which 28/80 (35%) corresponded to >2bp INDELs. Consequently, analysis-specific tolerances for false positives and negatives should be considered when working with biobank-scale joint genotyping data, especially for complex regions of the genome.

Session Title: Omics Technologies Poster Friday Session

Board 1163F: RISE: Relative Impact of Splicing and Expression in RNA-seq studies

Authors:

Y-J. Lin, A. A. Alazali, Z. Hu, S. E. Brenner; Univ. of California, Berkeley, CA

Abstract:

RNA-seq has been widely used to quantify expression and splicing changes in transcriptomes. Although biological consequences arise from changes in both expression and splicing aspects, researchers usually use their impressions to choose only one aspect to analyze, potentially overlooking significant impacts of the other. Even if researchers investigate both, the measurement scales of expression and splicing are different, and thus, their impacts are incomparable. To compare the relative impact of expression and splicing, we have developed RISE.
RISE qualifies the relative impact of expression and splicing changes caused by the treatment. To place the impact of expression and splicing changes on the same scale to compare, we developed the Normalized Variation (NV) measure. NV is defined as the proportion of the between-group variation to the total variation. Finally, we assess whether expression NV (eNV) or splicing NV (sNV) is significantly larger to understand the comparative influence of expression versus splicing alternations in the transcriptome.
To validate our method, we performed RISE analysis on RNA-seq data from knockdown or overexpression experiments of 11 transcription and splicing factors. RISE effectively categorizes transcription and splicing factors by their relative impacts on expression and splicing. As an example application, we applied RISE to 4 studies involving proteins with complex or previously unknown roles in regulating transcriptomes to understand their functions. In summary, RISE enables researchers to systematically compare the relative impact of expression and splicing.

Session Title: Omics Technologies Poster Friday Session

Board 1164F: ROQ: a new measure for more accurately filtering mapped reads

Authors:

N. Park, M. Gu, J. Kim; Konkuk Univ., seoul, Korea, Republic of

Abstract:

Mapping reads to a reference genome is a critical step in utilizing next-generation sequencing data. Typically, mapped reads are filtered based on their mapping quality (MAPQ) score, which indicates alignment reliability. However, our simulation-based study, where simulated short reads from a human genome assembly were mapped back to the same genome, demonstrated that high MAPQ scores do not always correspond to accurate mapping of reads to their original genomic regions. The correlation between the MAPQ scores and the overlap proportion of the actual and mapped locations was only 0.6237. This highlights the need for a new measure for

more accurately filtering mapped reads. To this end, we introduced a new measure, read overlapping quality (ROQ), designed to quantitatively measure the proportion of overlapping of the genomic region where a read originated and the mapped region of the read, and developed a machine-learning model based on XGBoost to predict the ROQ score given various information of mapped reads. When evaluated using simulation data, our model achieved an MSE of 0.0006 and an R^2 of 0.9279 with a correlation of 0.9633 between the ROQ scores and the overlap proportion of the actual and mapped locations. These results suggest that ROQ can serve as a superior measure to MAPQ for filtering mapped reads. The use of ROQ will contribute to a better understanding of genomes by dramatically reducing false-positive interpretation of reads.

Session Title: Omics Technologies Poster Friday Session

Board 1165F: Sawfish: Improving long-read structural variant discovery and genotyping with local haplotype modeling

Authors:

C. T. Saunders, J. M. Holt, D. N. Baker, J. A. Lake, J. R. Belyeu, Z. Kronenberg, W. J. Rowell, M. A. Eberle; PacBio, Menlo Park, CA

Abstract:

We describe sawfish, a structural variant (SV) caller for mapped high-quality long reads. This method emphasizes assembly of local SV haplotypes and their utilization in downstream sample merging and genotyping steps, improving accuracy compared to variant-focused approaches in both individual and joint-genotyping contexts.
Assessing sawfish against the GIAB draft SV benchmark based on the T2T-HG002-Q100 diploid assembly shows substantial accuracy gains compared to pbsv and Sniffles2 on HiFi WGS 33x input, with a sawfish F1 score of 0.971 compared to 0.930 and 0.935 for pbsv and Sniffles2, respectively. This accuracy gain persists at lower depth, for example at 10x depth the sawfish F1 score is 0.937, compared to 0.857 and 0.882 for pbsv and Sniffles2. For SVs in the GIAB Challenging Medically Relevant Genes benchmark, sawfish has a combined false positive and false negative count of 4, compared to 19 and 15 with pbsv and Sniffles2, respectively.
Sawfish also has higher genotype concordance in the Platinum Pedigree (CEPH-1463). Joint-genotyping accuracy was assessed on 10 HiFi WGS samples comprising the 2nd and 3rd pedigree generations, where the known inheritance pattern enables genotype accuracy assessment. From high genotype-quality calls, sawfish yields 27,811 concordant and 4,414 discordant SV alleles (86.3% concordance), where concordant alleles respectively represent 7.8 Mb and 11.9 Mb of deleted and inserted sequence. This substantially improves concordant

allele count, length and percent concordance compared to the next most concordant method, Sniffles2, with 20,519 concordant and 7,645 discordant alleles (72.9% concordance), where concordant alleles represent 4.2 Mb and 5.6 Mb of deleted and inserted sequence.
As additional improvements, our assembly-focused approach allows all calls to be made with single-base precision, enabling breakpoint insertion and homology annotation for all SV types. Sawfish also assesses depth of large deletions and duplications to evaluate their consistency with its own expected GC-corrected depth model, improving precision for these large SV types. Through the combination of high genotyping accuracy, detailed breakpoint modeling, and joint assessment of breakpoint evidence with read depth, sawfish offers improved options for WGS sample analysis with high-quality long reads.

Session Title: Omics Technologies Poster Friday Session

Board 1166F: scExploreR: a flexible platform for democratized analysis of multimodal single-cell data

Authors:

W. Showers[1,2], J. Desai[2], S. R. Gipson[1], K. L. Engel[1,2], C. Smith[1,2], C. T. Jordan[1], A. Gillen[1,3]; [1]Div. of Hematology, Univ. of Colorado Sch. of Med., Aurora, CO, [2]RefinedSci., Aurora, CO, [3]Rocky Mountain Regional VA Med. Ctr., Aurora, CO

Abstract:

Single-cell sequencing data analysis is often inaccessible to biologists and clinicians due to the specific technical expertise required to analyze these data. Proficiency in either R or Python is required even for simple analysis tasks such as generating plots from data, resulting in a high barrier of entry to perform analysis. Biologists and clinicians frequently must collaborate with bioinformaticians to extract insights from these data, but the different “languages” spoken by these disciplines often leads to communication roadblocks that complicate effective analysis. Multiple browsers have been developed to allow the direct interrogation of single-cell data through a point-and-click interface accessible to non-programmers, but many of these browsers are limited in the depth of analysis that can be performed, or the flexibility of input data formats accepted. We developed scExploreR to address these limitations, extending the range of analysis tasks that can be performed by users without extensive programming experience. scExploreR, implemented as a packaged R shiny app that can be run locally or easily deployed for multiple users on a server, allows for real-time data interrogation via a live R backend. This provides a user-friendly interface for the creation of common single-cell visualizations and basic differential expression analysis. Compared to many other browsers, scExploreR gives users

greater control over plot production by leveraging extensive customization features. For example, users may facet plots by a metadata variable, modify data ordering, adjust legends, or customize colors, features that are not common or accessible in these browsers. Ultimately, this control, combined with the ability to export plots as vector graphics and high-resolution bitmaps, allows users to generate publication quality figures. Additionally, scExploreR takes advantage of our SCUBA package to directly support multimodal data stored in Seurat, SingleCellExperiment, anndata, or mudata objects without conversion and with identical plotting capabilities for all supported input formats. While scExplorer is designed to work with single-cell objects that have already undergone basic preprocessing and quality control, convenience functions are provided to load 10X Genomics Cell Ranger output directly into the browser. scExploreR allows non-programmers to interactively explore multimodal single-cell data that has been preprocessed by a bioinformatician, bridging the gap between these disciplines and democratizing single-cell multi-omics research by lowering barriers of entry involved in data analysis.

Session Title: Omics Technologies Poster Friday Session

Board 1167F: scraps: an end-to-end pipeline for measuring alternative polyadenylation at high resolution using single-cell RNA-seq

Authors:

R. Fu[1], K. Riemondy[2], W. Showers[3], R. Sheridan[2], J. Hesselberth[2], C. Jordan[2], A. Gillen[2]; [1]New York Genome Ctr., New York, NY, [2]Univ. of Colorado Denver - Anschutz Med. Campus, Aurora, CO, [3]RefinedSci., Boulder, CO

Abstract:

Alternative cleavage and polyadenylation (APA) contributes to the diversity of mRNA 3´ ends, affecting post-transcriptional regulation by including or excluding cis-regulatory elements in mRNAs, thereby altering their stability and translational efficiency. While APA analysis has been applied broadly in mixed populations of cells, the heterogeneity of APA among single cells has only recently begun to be explored. We developed an approach we termed scraps (Single Cell RNA PolyA Site Discovery), implemented as a user-friendly, scalable, and reproducible end-to-end Snakemake workflow, to identify polyadenylation sites at near-nucleotide resolution in single cells using TVN-primed single-cell RNA-seq (scRNA-seq) libraries. Our approach, which requires sequencing through the poly(T) homopolymer and into the adjacent cDNA sequence (e.g. Read 1 >100bp for 10X Genomics 3´ captures), is both unbiased relative to existing methods that do not capture the priming site directly while recovering more sites at higher

resolution. Optimal performance requires sequencing on a platform that can reliably read through homopolymer stretches (e.g. the Ultima Genomics UG100), but libraries sequenced with compatible read lengths on Illumina sequencers are also sufficient to yield insights despite the reduction in read quality observed following homopolymer stretches. For libraries sequenced with incompatible read lengths, we implement a fallback approach using standard alignments. In addition to enabling accurate mRNA 3´ end isoform level quantitation and direct measurement of APA events in single cells, scraps also enables assessment of internal priming events, which we demonstrate occur at a higher frequency during apoptotic 3´ RNA decay. We also provide R and python packages, scrapR and scrap-py, respectively, that integrate the results of the scraps pipeline with the popular Seurat and scanpy single-cell analysis packages, as well as single-cell APA modeling packages. Refinement and expanded application of these approaches will further clarify the role of APA and 3´ UTR regulatory code in single cells, as well as the effects of internal priming on expression measurements in scRNA-seq libraries.

Session Title: Omics Technologies Poster Friday Session

Board 1168F: Single Cell RNA-seq Cell Annotation with Generative Pre-trained Transformers (GPT) Models

Authors:

R. Chen[1], R. Luo[1], H. Cui[2], B. Wang[2]; [1]Univ. of California, Irvine, Irvine, CA, [2]Univ. of Toronto, Toronto, ON, Canada

Abstract:

Annotation of cell types from scRNA-seq data is effective but demanding, often laborious, and error-prone, especially for rare, understudied, or novel cell types. While single-cell omics technologies are increasingly prevalent, there is a need for more robust and effective tools for cell annotation to fully utilize the high-resolution single-cell datasets generated by the community.

The rise of machine learning techniques, especially Generative Pre-trained Transformers (GPTs), presents a promising avenue to tackle the complexity of these large datasets. By leveraging the pre-trained scGPT foundation model, we explore fine-tuning with domain-specific data to generate models and functionalities uniquely tailored for visual system research. Using the retina V1 cell atlas dataset, which includes a total of 2.9 million single-cell and single-nucleus RNA sequencing profiles and 111 distinct cell types, we fine-tuned the scGPT model to optimize cell annotation for the human retina.

As expected, high performance is observed with scGPT without fine-tuning at the low-resolution

cell class level, which includes labels incorporated in the initial foundation model training. Improved performance is achieved at low resolution, resulting in the model performing nearly perfectly. In contrast, poor performance is observed for high-resolution cell type annotation without fine-tuning, as expected. However, striking improvements are observed with fine-tuning at high resolution. For example, the accuracy of predicting bipolar cell types, which account for about 10% of the retina cell population and encompass 14 cell types, improved from an initial 38.8% to 99.6% upon fine-tuning. Similarly, improved performance is achieved with the fine-tuning of scArches and Geneformer models, indicating the general applicability of fine-tuning for deep learning models. It is worth noting that scGPT is significantly faster than the other two models.
Further testing of these models against diverse datasets, including those from different technologies (single-nucleus versus single-cell RNA sequencing), both diseased and healthy samples, and with or without cell type enrichment, is ongoing to systematically evaluate their performance across various realistic scenarios. We expect that the experiences gained from this project will not only improve the applicability of GPT models for cell type annotation within the eye but will also inform strategies for fine-tuning foundation models applicable across different tissues and organs.

Session Title: Omics Technologies Poster Friday Session

Board 1169F: Single-cell spatial RNA-sequencing on viral infections in human cerebral organoids reveal human-viral transcript interactions in Alzheimer's disease

Authors:

E. Lim[1], P. Dawes[1], N. J. Barton[1], A. J. Orszulak[1], L. F. Murray[1], S. M. Chigas[1], K. Tran[1], M. N. Olson[1], J. Sundstrom[1], H. Oh[2], Q. Wang[3], G. M. Church[2], D. M. Knipe[2], B. Readhead[3], Y. Chan[1]; [1]UMass Chan Med. Sch., Worcester, MA, [2]Harvard Med. Sch., Boston, MA, [3]Arizona State Univ., Tempe, AZ

Abstract:

Common viral infections e.g. by herpes simplex virus 1 (HSV-1) were associated with Alzheimer's disease (AD). Previously, we found that HSV-1 infection in 2D dissociated cells from cerebral organoids (dcOrgs) led to AD molecular pathology: increased intracellular Ab aggregates, hyperphosphorylation of tau-181, decreased secreted Ab42/40, neuronal loss and increased proportions of glia cells. We explored transcriptomics signatures that may be associated with molecular AD pathology by conducting multi-transcriptomics profiling on HSV-1 infected 2D

dcOrgs and 3D cerebral organoids (cOrgs). We generated bulk RNA sequencing from 39 samples, single-cell non-spatial RNA sequencing (scRNA-seq) from 6 samples using Parse Evercode and single-cell spatial RNA sequencing (sc-spatialRNA-seq) from 6 sections (36 cOrgs) using Stereo-seq. Despite the cell type heterogeneity in dcOrgs, we found high correlations across intra-condition replicates, with Pearson's r>0.99. Pseudo-bulk data from scRNA-seq showed similarly high correlations with r>0.99. Correlations across inter-condition samples were significantly lower at r=0.75-0.79. High correlations with r=0.92-0.95 were also observed for intra-condition Pseudobulk data from sc-spatialRNA-seq on 3D cOrgs; however inter-condition correlations did not differ significantly (r=0.9-0.98), indicating that HSV-1 infection in 3D cOrgs did not result in significant human transcript perturbations globally, unlike in 2D dcOrgs. Gene set enrichment analyses (GSEA) on bulk data showed that the differentially expressed genes were exclusively enriched for AD-associated genes in the GWAS Catalog (nominal P=0.039). GSEA on pseudo-bulk data from scRNA-seq revealed that the enrichment for AD-associated genes were driven by differentially expressed genes between HSV-1 negative (pseudo-uninfected) cells and HSV-1 infected (true-infected) cells, demonstrating that complex cell type interactions between neurons and glia cells contributed to differentially expression of AD-associated genes. Finally, we observed human transcript cell type clusters that were highly enriched for HSV-1 infected cells were found in both the scRNA-seq and sc-spatialRNA-seq datasets, and were enriched for neurons. UMAP analyses on viral transcripts from the sc-spatialRNA-seq data also revealed distinct clusters with varying proportions of viral transcripts across the transcriptional life cycle (intermediate-early, early, leaky late and true late) and shed insights into viral-human transcript interactions involved in neuroinflammation-induced AD pathology.

Session Title: Omics Technologies Poster Friday Session

Board 1170F: Single-cell Total-RNA Profiling Unveils Regulatory Hubs of Transcription Factors ⭐

Authors:

C. Zong, Y. Niu, J. Luo; Baylor Coll. of Med., Houston, TX

Abstract:

Recent development of RNA velocity uses master equations to establish the kinetics of the life cycle of RNAs from unspliced RNA to spliced RNA (i.e., mature RNA) to degradation. To feed this kinetic analysis, simultaneous measurement of unspliced RNA and spliced RNA in single cells is greatly desired. However, the majority of single-cell RNA-seq chemistry primarily captures

mature RNA species to measure gene expressions. Here, we develop a one-step total-RNA chemistry-based single-cell RNA-seq method: snapTotal-seq. We benchmark this method with multiple single-cell RNA-seq assays in their performance in kinetic analysis of cell cycle by RNA velocity. Next, with LASSO regression between transcription factors, we identify the critical regulatory hubs mediating the cell cycle dynamics. We also apply snapTotal-seq to profile the oncogene-induced senescence and identify the key regulatory hubs governing the entry of senescence. Furthermore, from the comparative analysis of unspliced RNA and spliced RNA, we identify a significant portion of genes whose expression changes occur in spliced RNA but not to the same degree in unspliced RNA, indicating these gene expression changes are mainly controlled by post-transcriptional regulation. Overall, we demonstrate that snapTotal-seq can provide enriched information about gene regulation, especially during the transition between cell states.

Session Title: Omics Technologies Poster Friday Session

Board 1171F: Spatial Characterization of Isoform Landscapes in the Human Brain

Authors:

H. Eden[1], L. Morina[1], S. Hicks[1], K. Maynard[2], K. Martinowich[2], W. Timp[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Lieber Inst. for Brain Dev., Baltimore, MD

Abstract:

Both recently dubbed method of the year by *Nature Methods*, spatial transcriptomics and long-read sequencing stand out as platforms for genetic exploration with unrivaled resolution. Spatial transcriptomics adds tissue organization context to RNA sequencing, revealing genes with spatially distinct expression patterns, and identifying neighborhoods within tissues. Long-read RNA/cDNA sequencing captures full-length molecules enabling direct quantification of RNA isoforms. Our work aims to amplify the utility each of these methods offer by applying them in tandem. Conventional spatial transcriptomics methods coupled to short read sequencing report gene level information and do not capture splice isoforms, limiting the ability to detect differential transcript usage within tissues. As phenotypic complexity is driven by the alternative splicing of RNA into functionally distinct messages, employing methods with the sensitivity to distinguish transcript isoforms promises to fine-tune our understanding of elaborate biological systems. We have taken advantage of the compatibility between Oxford Nanopore sequencing (ONT) and the 10X Visium spatial platform to generate spatially resolved RNA sequencing data with single-molecule resolution. The feasibility of this approach is facilitated by recent advances

in ONT technology which have improved sequencing accuracy such that we see a 5% increase in per base quality scores. This enables confident demultiplexing of spatial barcodes and UMI deduplication. As proof of concept, we generated a “spatial barnyard” by applying tissues from both mouse and human brains to a Visium slide array. After processing with 10X reagents, and sequencing the resulting mixed species cDNA using ONT, we demultiplexed the reads and assessed the frequency at which a barcode known to be covered by either tissue was represented by reads that mapped to the corresponding genome. We saw > 95% species specificity from demultiplexed reads, validating the use of this approach for further investigations. As a result, we have profiled transcriptomes of the dorsolateral prefrontal cortex from four post-mortem tissue samples to characterize transcript isoform expression patterns across anatomically and transcriptionally distinct layers of the human brain.

Session Title: Omics Technologies Poster Friday Session

Board 1172F: Speeding up Genetic Analysis with a Smaller Data Format

Authors:

A. Stine; NCBI, Bethesda, MD

Abstract:

The Sequence Read Archive (SRA) is a raw sequencing data archive storing submissions from across the taxonomic tree including 16PB of human data. Submission to the archive continues at a rapid pace with 5PB published in 2023. Balancing the goals to sustainably archive 32PB of genetic data while providing rapid access for users presents a constant challenge. To address these challenges the SRA Lite format was designed to calculate a single binary read quality score from per base quality scores. This reduces the size of SRA Lite files to about half the file size of the SRA format with full quality scores. SRA Lite is already available for ~90% of data in the archive and is frequently used for sequence data analysis within the archive. SRA Lite substantially reduces download and extraction time for researchers to access data while remaining a useful data storage format for a variety of genetic analysis.

Session Title: Omics Technologies Poster Friday Session

Board 1173F: Streamlined, multimodal, epigenetic measurements using CUT&RUN combined with nanopore sequencing

Authors:

P. Hook[1], B. Venters[2], A. Hickman[2], W. Timp[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]EpiCypher, Inc, Durham, NC

Abstract:

Introduction: Cleavage Under Targets and Release Using Nuclease (CUT&RUN) has become a popular method for measuring protein-DNA binding. CUT&RUN libraries are typically sequenced on Illumina, short-read platforms, tethering this assay to a particular sequencing technology. We sought to show that robust CUT&RUN profiles could be produced on Oxford Nanopore Technologies (ONT) nanopore sequencing devices.
Methods: We performed CUT&RUN on GM12878 and K562 cell lines using antibodies against five targets: IgG (control), H3K4me3 (active promoters), H3K27me3 (inactive chromatin), H3K36me3 (transcribed genes) and CTCF (transcription factor). Extracted DNA was prepared for Illumina and ONT sequencing using PCR with typical Illumina adapters. In addition, purified fragments were sequenced without amplification with ONT. Illumina libraries were sequenced on a NextSeq 2000 and ONT libraries were sequenced on MinION or PromethION flow cells.
Results: Reads from all libraries showed expected fragment size distributions indicative of nucleosome laddering. The majority of amplified reads had fragment lengths less than 200 base pairs. Unamplified-ONT libraries had longer read lengths, often exceeding 600 base pairs. All three sequencing approaches had high, pairwise, genome-wide correlation for all marks. Peak calling was optimized for ONT datasets and they recapitulated 80-90% of short-read peaks for narrow marks (H3K4me3 and CTCF) and 65-75% of short-read peaks for broad marks (H3K27me3 and H3K36me3). The longer length of unamplified-ONT libraries allowed for ~25% of reads to be successfully assigned to phased haplotypes, compared to just ~5% of Illumina reads and ~9% of amplified-ONT reads. In addition, the unamplified-ONT libraries allowed us to accurately measure 5-methylcytosine on the same single molecules bound by the target proteins.
Conclusion: By establishing CUT&RUN with nanopore sequencing, we have streamlined the benchtop measurement of protein-DNA interactions. The sequencing of unamplifed-ONT libraries transforms CUT&RUN into a multimodal, single-molecule assay without the need for complex molecular biology protocols. These advances will increase the accessibility of CUT&RUN and expand the biological questions that can be asked with this powerful assay.

Session Title: Omics Technologies Poster Friday Session

Board 1174F: Temporal and spatial dysregulation of immune and metabolic pathways of prefrontal cortex due to COVID-19

Authors:

K. Kumar; Columbia Univ., New York, NY

Abstract:

The COVID-19 pandemic has resulted in over 760 million cases and more than 6.9 million deaths globally. Neurological involvement may occur in over a third of COVID-19 cases, necessitating an understanding of the molecular mechanisms underlying SARS-CoV-2 neuropathogenesis. This study re-analyzes single nucleus RNA-sequencing data from the prefrontal cortex, including samples from control/non-viral individuals, COVID-19, and influenza infected patients. Predicted mapping of single nucleus data with spatial transcriptomics characterizes differential euclidean distances between cell type pairs and identifies intercellular LRIs. Prediction of significantly altered spatial distances between L2/3/4 neurons and microglial cells indicates disrupted cell-cell communication in COVID-19 patients compared to non-viral and influenza-infected. Key intracellular LRIs involving immune cells include *CD74-MIF* and *CXCR4-CXCL12*, which may drive neuroinflammation and contribute to neurological symptoms. Dysregulated pathways related to immune response include NF-κB and JAK-STAT signaling, with elevated expression of *IL6*, *TNF*, and *CXCL10* in SARS-CoV-2 patient samples. These cytokines may activate IDO-1, catalyzing the first step of Trp catabolism through the kynurenine pathway which produces several neurotoxic and immunosuppressive metabolites. Synaptic signaling and neurogenesis pathways show alterations in *NRXN1*, *NRG1*, and *BDNF*, further suggesting impaired neural communication and brain plasticity. Metabolic processes, including oxidative phosphorylation and glycolysis, may be disrupted with significant downregulation of *ATP5F1* and *PFKM* in astrocytes of SARS-CoV-2 infected patients compared to controls. Gene regulatory network analysis identifies hub transcriptional regulators *ST18*, *IRF7*, and *RELA* central to the immune response to SARS-CoV-2 infection. Circadian-related genes, including *PER1* and *CLOCK* with significant alterations in their pseudotime oscillation patterns, suggest temporal regulation of pathways in SARS-CoV-2 infected compared to control and influenza infected patients. Drug-gene interaction analysis reveals potential therapeutic targets among differentially expressed genes, including for dexamethasone and baricitinib.

Session Title: Omics Technologies Poster Friday Session

Board 1175F: The Long Road to Long Reads: Challenges to Implementing a Long Read Sequencing Service

Authors:

M. Kokosinski[1], B. Marosy[1], J. Gearhart[1], J. Paschall[1], B. Craig[2], M. Mawhinney[1], D. Mohr[3], M. Sheridan[3], P. Witmer[4], D. Muzny[5], A. Scott[6], J. Hosea[3], C. Montano[1], L. Morina[3], Q. Li[3], A. Klein[7],

M. Schatz[3], W. Timp[3], K. Doheny[8], All of Us Research Program; [1]Johns Hopkins Univ, Baltimore, MD, [2]Johns Hopkins Genomics, Baltimore, MD, [3]Johns Hopkins Univ., Baltimore, MD, [4]Johns Hopkins Univ SOM, Baltimore, MD, [5]Baylor Coll. Med., Houston, TX, [6]Johns Hopkins Sch. of Med., Baltimore, MD, [7]Johns Hopkins Med Sch, Baltimore, MD, [8]Ctr. for Inherited Disease Res. (CIDR), JHU, Baltimore, MD

Abstract:

Long read sequencing has been a steadily developing technology with several advantages over short read sequencing including resolving large and complex structural variants, tandem repeats, phasing of variants and native detection of base modifications. Implementation of Oxford Nanopore Technologies long read sequencing in a high throughput core facility has required standardization and optimization of protocols to achieve 30x coverage whole genomes with an N50 >25kb in a reproducible manner. Typically, long read sequencing protocols require new extractions to facilitate high molecular weight DNA and increased DNA inputs to ensure higher N50s. Studies which use previously extracted DNA require optimization of protocols in order to achieve similar N50s. Here we provide insight from implementing this technology utilizing biobanked DNA and optimizing protocols for longer N50s, utilizing 2-6 ug of input DNA, g-Tube shearing and size selection on the Blue Pippin prior to LSK114 library prep and sequencing on a single R10.4 flowcell with 3 loads over 72 hours. To date we have processed over 464 samples achieving an average of 32x coverage and N50 of 27.2kb. Challenges to production-scale work have included the time required for super-high-accuracy live basecalling as well as inconsistencies in total yield of R10.4 flowcells. A second high priority study is examining familial pancreatic cancer samples with an expected total of at least 400 long read WGS. A pilot using 66 cell line derived DNA samples (fresh extractions optimized for long reads) achieved an average of 60x coverage and N50s of 25.6kb. In order to achieve this higher depth, we adapted our methods by increasing DNA input to LSK114 library prep to at least 6 ug, creating 2 libraries per sample, and sequencing 2 R10.4 flowcells per sample. Currently, we are actively exploring adaptive sampling, to address clinical application for repeat expansion testing, methylation profiling, and elucidation of "unidentified" 2nd variants in individuals with cystic fibrosis, to inform treatment options.

Session Title: Omics Technologies Poster Friday Session

Board 1176F: Uncovering the impact of splicing variants in intractable pediatric epilepsies ⭐

Authors:

I. McSalley, H. Xie, S. Parthasarathy, S. Ruggiero, S. Ganesan, B. Kennedy, I. Helbig; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

The disruption of conventional splicing patterns in the transcription of mRNA is a well-established mechanism contributing to the cause of many genetic diseases. The continued advancement of *in silico* predictors of splice-altering variants has fueled both the expansion of the framework of genomic context underpinning the complexity of alternative splicing and the accelerating discovery of cryptic splicing variants found far beyond the canonical splice site. Validating these predicted splicing outcomes in clinical settings is challenging due to limited access to relevant tissue samples. This particularly impacts research areas where tissues are scarce, such as Neurodevelopmental Disorders (NDD). Here, leveraging whole-genome-sequencing (WGS) data from a cohort of 26 pediatric patients, receiving surgical interventions for drug-resistant epilepsies, we detect single nucleotide polymorphisms (SNPs), predicted by SpliceAI's deep learning model to yield alternative splicing events. To evaluate their effect, candidate variants were further interrogated with RNA-Seq data, produced from resection specimens, using Multivariate Analysis of Transcript Splicing (rMATS) to detect differential splicing events. After filtering for read depth > 10 and gnomAD v.4.1 population frequency < .0001, we identified 220,249 unique variants in 17,582 genes. Of these variants, 223 in 223 genes were evaluated with SpliceAI Delta Scores above 0.2. After filtering against an expert curated list of 593 genes associated with NDD, we selected 16 variants in 16 genes for evaluation of alternative splicing, using rMATS. We performed one-vs-all comparisons of differential splicing events, only considering events where differences in exon inclusion levels exceeded 5%, were supported by 20 or more reads, and in which the exons involved were in the immediate vicinity of the associated SNP. We identified three variants (*CACNA1D*, *MMADHC, SYT1*) with significant differential splicing events (FDR < .01). Using transcriptional data from tissue recovered from pediatric patients with intractable epilepsies who received surgical interventions, we uncover variants impacting ordinary splicing patterns in disease relevant genes. These variants provide insights into the impacts of alternative splicing in epilepsy and NDD.

Session Title: Omics Technologies Poster Friday Session

Board 1177F: Understanding the role of rare untranslated region (UTR) and long non-coding RNA variants on human plasma protein levels using ~50,000 UK Biobank whole-genome sequences

Authors:

K. Kundu[1], F. Hu[1], Q. Wang[2], O. Burren[1], R. Dhindsa[2], K. Carss[1], S. Petrovski[1]; [1]AstraZeneca, Cambridge, United Kingdom, [2]AstraZeneca, Waltham, MA

Abstract:

Evaluation of the contribution of rare variants to human phenotypes have thus far primarily focused on protein coding regions. Consequently, the role of rare non-protein coding variants in human traits and diseases remains largely unknown. The recent UK Biobank whole-genome sequencing (WGS) data of ~500,000 individuals, combined with Olink plasma proteomics measured in a subset of approximately 50,000 individuals, offers an unprecedented opportunity to explore the impact of rare non-protein coding variants on protein levels at scale.
In this study, we identify associations between rare non-protein coding variants in untranslated regions (UTRs, 5' and 3') and long non-coding RNAs (lncRNAs), and 2,923 plasma protein abundances measured in 49,736 UK Biobank participants. Qualifying variant models were defined by characterizing variants based on minor allele frequency (MAF) and deleteriousness predicted by CADD and UTRannotator.
To identify rare non-protein coding variants that impact protein levels, we performed variant - level, and collapsing analyses by aggregating the effect of rare variants on 2,923 plasma protein abundances separately for UTRs and lncRNAs, applying a framework we previously developed for protein coding variants [1]. Our variant-level analysis identified 2,511 and 2,338 rare genotype-protein associations for UTR and lncRNA, respectively and our gene-level collapsing analysis yielded 463 significant ($P < 1x10^{-8}$) UTR-protein associations, including 194 cis protein quantitative trait loci (pQTLs) and 269 trans pQTLs, involving 374 gene UTRs and 364 proteins. Among them 93% of cis pQTL and 51% of trans pQTLs decrease protein levels. For lncRNAs, we identified 463 significant ($P < 1x10^{-8}$) lncRNA-protein associations, including 98 pQTLs, involving 88 lncRNAs and 64 proteins (71% of pQTLs decreases protein levels).
Our study is the largest to date that explores the impact of rare non-protein coding variants on plasma protein levels. Identifying the subset of rare variants in non-protein-coding regions that have a functional effect remains challenging. Our work will improve our understanding and interpretation of non-coding variants. Ongoing work includes augmenting rare variant protein coding association analyses with our UTR pQTLs to improve our power to detect novel associations. This study will also offer valuable insights into protein and disease biology, and provides a framework that could be extended to other regulatory regions.

Dhindsa et al., Rare variant associations with plasma protein levels in the UK Biobank. Nature. 2023

Session Title: Omics Technologies Poster Friday Session

Board 1178F: Unmasking HMC3 Cell Lineage: Insights from Comparative RNA Sequencing Analysis

Authors:

K. Rahm, D. V. Hansen, M. Bailey; Brigham Young Univ., Provo, UT

Abstract:

Background: Recent advances in Alzheimer's research suggest that the brain's immune system plays a critical role in the development and progression of this devastating disease. Microglial cells are essential to the brain's immune system. These cells comprise ~10% of the cells in the brain and are the brain's first line of defense. While working with cell lines in laboratory settings is critical to understanding human disease, very few experimental microglial-cell lines exist. One cell line, HMC3, was recently developed as a promising experimental model to understand its role in Alzheimer's disease and other neurodegenerative diseases. It has been used in over 100 published papers in the last four years. RNAsequencing expression data, however, suggests that HMC3 cell lines may more accurately represent a different yet vital brain cell, rather than microglia.
Methods: We compiled gene expression data in both human brain cells and brain cell lines, and used relative gene expression comparisons to create a relational matrix of gene expression within certain cell or cell line types. We used a statistical package multiclassPairs in order to avoid the need for RNAsequencing batch correction across multiple studies. Briefly, multiclassPairs utilizes the relationship between two genes as opposed to single gene quantities, thus, allowing the comparison among multiple studies. We constructed a predictive model that performed well in predicting the cell of origin from various datasets.
Results: When applied to iMGL cells, an induced pluripotent stem cell-derived microglia line, the model predicted that these cells most closely resemble microglia. However, when applied to HMC3 cells, the model predicted that these cells are more closely related to astrocytes than microglia.
Conclusions: Contrary to rising belief, the HMC3 cell line is not a suitable human microglia model. We would recommend limiting over-interpretation of the HMC3 cell line results when applied to Alzheimer's Disease and other neurodegenerative disease research.

Session Title: Omics Technologies Poster Friday Session

Board 1179F: Unraveling the Testicular Seminoma Tumor Microenvironment by Single Cell RNA-sequencing

Authors:

K. Giannikou[1], J. T. Lafin[2], J. Sfakianos[3], B. Hopkins[4], H. Yu[5], C. Jamieson[6], J. F. Amatruda[1], A. Horowitz[5], A. Bagrodia[6]; [1]Cancer and Blood Disease Inst., Children's Hosp. Los Angeles, Los Angeles, CA, [2]Dept. of Urology, Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [3]Dept. of Urology, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]The Marc and Jennifer Lipschultz Precision Immunology Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY 10029, USA; Dept. of Oncological Sci., Tisch Cancer Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Dept. of Urology, Univ. of California San Diego, La Jolla, CA

Abstract:

Background/Aim: Seminomas is a malignant germ cell tumor (GCT) affecting male adults and is characterized by low somatic mutation burden, histological homogeneity, and high survival rate. It is currently unknown what drives tumorigenesis of this curable testicular cancer. Tumor/immune microenvironment (TME) has a vital role in cancer survival and progression with implications in cell death resistance and immune escape. Herein, we aimed to unravel the transcriptomic TME landscape of seminomas and better understand seminoma pathogenesis.**Methods:** We performed single cell RNA-sequencing (scRNA-seq; 10X-Illumina platform) on surgically resected pure seminomas from four patients. A total of ~24,000 single cells were analyzed using Seurat/Bioconductor R packages and standard means for data visualization. Key findings were validated by in situ hybridization RNA analysis (RNAscope) and Visium spatial transcriptomics (10X) on histological sections in an independent set of tumor samples. **Results:** T-cells and Natural-Killer cell populations were mostly predominantly enriched in this seminoma cohort, where tumor cells consisted of a distinct subpopulation with limited presence (1%). Other distinct cell populations were also identified for additional immune cell types e.g., B-cells, macrophages/monocytes and smooth muscle cells. ScRNA-seq revealed a unique set of signature marker genes (e.g., *POU5F1, KHDC3L, DPPA5, PCSK1N, NANOS3*) highly expressed in seminoma cells which are associated with pivotal cellular processes e.g., developmental pluripotency and germ cell development. Notably, this gene set was also highly differentially expressed (DE) in The Cancer Genome Atlas (TGCA) seminomas (n=73) compared to TCGA non-seminomas (n=38) (FDR<0.01). Amongst top marker genes, *KHDC3L*, also higher expressed in seminomas compared to normal testis, was validated by RNA-scope. Notably, KHDC3L demonstrated the strongest positive correlation with DPPA5 (Spearman/Pearson;

R2=0.8, p<0.0001), top #3 ranked DE gene as well as with other top 20 DE genes identified by scRNA-seq. Remarkably, similar strong correlations among other top DE genes were also found in the testicular TGCA cohort. Top DE genes including master transcription factors (e.g., *POU5F1, NANOG, SOX4, MYBL2*) are involved in the same protein pathway/networks likely suggesting synergic effect/involvement in the same regulatory circuitry.**Conclusions:** Our analyses revealed novel marker genes, uniquely expressed in seminomas which consist of a homogeneous cell population across all 4 tumors analyzed, clearly distinct from other cell types identified.

Session Title: Omics Technologies Poster Friday Session

Board 1180F: Unveiling novel insights into the pathophysiology of TANGO2 Deficiency Disorder using integrated functional and multiomics studies

Authors:

A. Kashiparekh[1,2], A. Powers[1,2], A-W. Mohsen[1,2], J. Vockley[2,1], L. Ghaloul-Gonzalez[1,2]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]UPMC Children's Hosp. of Pittsburgh, Pittsburgh, PA

Abstract:

TANGO2 deficiency disorder (TDD) is a multisystem disease caused by mutations in the *TANGO2* gene. Symptoms include metabolic crises with rhabdomyolysis, neurological regression, seizures, cardiomyopathy and cardiac arrhythmias. While mitochondrial dysfunction and impaired fatty acid oxidation are evident, the disease pathophysiology remains unknown. Supplementation with Vitamin B5, a precursor of coenzyme A (CoA), has shown alleviation in seizures and metabolic crises but the mechanism of action is unclear.
Co-immunoprecipitation (co-IP) studies were performed to identify physiologically relevant TANGO2 binding partners through limited proteomics. TDD patient-derived fibroblasts, with and without B5 treatment, were used to identify transcriptional and metabolomic networks. We performed a systematic integration of the multiomic profiles paired with functional validation to understand the function of TANGO2 and the pathophysiology of the disease.
Initial co-IP studies confirm TANGO2 protein pull down. Lipidomics data show alterations in ceramide, cardiolipin and sphingomyelin levels, pointing to potential causes of cardiomyopathies and neuropathies observed. We note an impairment in cholesterol biosynthesis correlating with previous studies potentiating a role of TANGO2 in maintaining lipid homeostasis. Integrative analyses demonstrate changes in mitochondrial and glycolytic ATP production in TDD, consistent with *in vitro* studies. We note a buildup of pyruvate and decreased long fatty acylcarnitines, both leading to decreased acetyl-CoA entering the TCA

cycle. This implies that B5 provides a partial rescue by increasing the free CoA pool, driving the conversion of pyruvate to acetyl-CoA to maintain TCA cycle activity. Calcium imbalance, increased ER stress, impaired Golgi-ER retrograde trafficking, and consequent increase in proteasomal activity are predicted. All correlate with functional Golgi/ER trafficking studies and apoptotic studies and can explain the arrhythmias observed. Gene expression changes including *SNCA* and *SH3GL2*, noted in several neurological disorders, are observed.
We postulate a primary role of TANGO2 in maintaining lipid homeostasis, calcium signalling and metabolic pathways. The loss of TANGO2 protein demonstrates a varied impact on different tissues due to several pathways involved, explaining many of the clinical symptoms observed. In addition, our data provide novel insights into how B5 may partially ameliorate phenotype in patients. This integrative approach offers a unique lens to understand TDD pathophysiology with the potential to guide targeted therapies.

Session Title: Omics Technologies Poster Friday Session

Board 1181F: Unveiling the complex landscape of 16p11.2 deletion: a multifaceted examination of phenotypes, genotypes, and transcriptomic patterns

Authors:

G. Li, Y. Jianle, C. Jihao, L. Qing, N. Wu; Peking Union Med. Coll. Hosp., Beijing, China

Abstract:

Background: The 16p11.2 BP4-BP5 duplication is recognized as the most frequently associated copy number variant with autism spectrum disorder (ASD), schizophrenia, and various comorbidities, including decreased body mass index (BMI) [PMID: 26629640]. This study aims to elucidate the intricate relationship between the 16p11.2 deletion and its associated phenotypes.
Methods: A total of 84 individuals carrying the 16p11.2 microdeletion, along with their relatives, were identified through Whole Exome Sequencing (WES) in conjunction with MassARRAY/SNP array or Whole Genome Sequencing (WGS).
Phenotypic Assessment: Utilizing a comprehensive phenotypic assessment, we constructed a multi-system phenotypic profile for individuals with the 16p11.2 microdeletion. Based on this detailed phenotyping, patients carrying the 16p11.2 microdeletion were classified into phenotypic subgroups.
Transcriptomic Analysis: Subsequently, we conducted a transcriptomic analysis on all patients who underwent RNA sequencing. A differential expression analysis was performed across the

phenotypic subgroups to identify patterns and variations in gene expression.
Results: Our findings reveal distinct phenotypic and transcriptomic patterns associated with the 16p11.2 deletion, providing a nuanced understanding of the genotype-phenotype relationship in this cohort.
Conclusion: This multifaceted examination underscores the heterogeneity of the 16p11.2 deletion syndrome and offers insights into the molecular underpinnings of associated phenotypes. The study contributes to the field by delineating the complex interplay between genetic variations and their phenotypic manifestations.

Session Title: Omics Technologies Poster Friday Session

Board 1182F: Unveiling the Hidden Rules: Enhancing NMD Prediction for Protein-Truncating Variants

Authors:

I. Egab[1], P. Orchard[2], J. Posey[3], R. Gibbs[4], C. Carvalho[5], J. Lupski[3], C. Shaw[6], S. Yi[7], L. Mestroni[8], M. Taylor[9], E. Boerwinkle[10], A. Reiner[11], P. De Vries[12], A. Morrison[13], S. Jagannathan[14], Z. Coban Akdemir[15]; [1]UTHlth.Houston, Houston, TX, [2]Univ. of Michigan, Ann Arbor, MI, [3]Baylor Coll. of Med., Houston, TX, [4]Baylor Coll. Med., Houston, TX, [5]Pacific Northwest Res. Inst., Seattle, WA, [6]BAYLOR Coll. OF MEDICINE, Houston, TX, [7]The Univ. of Texas at Austin, Austin, TX, [8]Univ Colorado AMC, Aurora, CO, [9]U of Colorado Anschutz Med. Campus, Aurora, CO, [10]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [11]Univ of Washington, Seattle, WA, [12]UTHlth.at Houston, Houston, TX, [13]Univ. of Texas Hlth.Sci. Cntr Houston, Houston, TX, [14]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [15]UTHlth.Sch. of Publ. Hlth., Houston, TX

Abstract:

Protein-truncating variants (PTVs) introduce premature termination codons (PTCs) into mRNA (PTC-variants), often resulting in loss-of-function (LoF) alleles due to nonsense-mediated decay (NMD). However, some PTC-variant bearing transcripts escape NMD, leading to truncated proteins with potential LoF, gain-of-function (GoF), or neutral effects. Accurately predicting whether a PTC variant will undergo NMD is challenging, as existing models, such as the canonical rule based on the exon junction complex (EJC), explain only about 50% of NMD outcomes. This highlights the need for additional predictive rules, potentially incorporating alternative signals. We propose identifying novel NMD prediction rules by analyzing large-scale RNA sequencing (RNA-Seq) and whole genome sequencing (WGS) datasets. We extracted ~13K high-quality heterozygous germline PTC-variants from 7,617 samples available from the NHLBI Trans-Omics for Precision Medicine (TOPMed) freeze 1.1 RNA-Seq dataset coupled with WGS

and mapped them to the GENCODE v26 canonical transcript set. For NMD-triggering variants, the expected proportion of wild-type allele to total expression is ~1, while for NMD-escape variants, it should be ~0.5. NMD efficiency was quantified based on allele-specific expression (ASE) values and analyzed against known and potential novel NMD rules, including canonical rule, 3'UTR length, PTC distance to start codon/stop codon/nearest exon junction, RNA-binding motifs, potential EJCs, and within 3'UTR sequences. Our analysis confirmed the canonical rule, showing higher NMD efficiency for PTC-variants upstream of the last EJC (p-value=4.4e-12). Other significant rules included increased NMD efficiency for PTC-variants with longer 3'UTRs (p-value=1.9e-3) and decreased efficiency for PTC-variants near the start codon (first 200 bp) (p-value=1.6e-4) and within the first 31% of a transcript (p-value=8.7e-5). PTC-variants with three downstream exon-exon junctions (p-value=7e-4) had increased NMD efficiency, whereas those further from the nearest exon-exon junction (≥162 bp) had decreased efficiency (p-value=7.2e-3). Using these rules, we built a random forest classification model to predict NMD outcomes for PTC-variants, achieving ~71% accuracy on a test set. This model classified ~42.6% of 165.3K high-quality PTC-variants from the TOPMed freeze 9 dataset as NMD-escaping and 57.4% as NMD-triggering. In sum, identifying novel NMD prediction rules and developing a more accurate NMD prediction model can enhance genetic and clinical research, leading to improved diagnosis and treatment strategies for diseases caused by PTC-variants.

Session Title: Omics Technologies Poster Friday Session

Board 1183F: Using transcriptomics to identify RNA biomarkers of autophagy status in primary human mononuclear cells

Authors:

E. Roberson, M. A. Rosana; Washington Univ., Saint Louis, MO

Abstract:

Autophagy is the process cells use to recycle old and damaged cellular components. The rate of autophagy and ability to increase or decrease autophagy (autophagic flux) varies throughout development, decreases with age, and may be a key mediator of multiple human disease processes. The current gold standard for measuring autophagy is calculating the ratio of the protein LC3 in the cytoplasm (LC3-I) to that accumulated in the autophagosome (LC3-II) using flow cytometry or western blotting. Knowing general levels of autophagy would benefit clinical trials and drug development, but lower costs and more easily implementable biomarkers are needed. We hypothesized that we could use RNA-Seq to identify genes that correlate with low and high autophagy. Here, we demonstrate that a subset of RNAs vary with autophagy status.

We used primary human mononuclear cells (**PBMCs**) with a kit for measuring LC3-II with flow cytometry using a fluorescent antibody and selective cell permeabilization. We needed to induce low and high autophagy to find transcripts correlated with those states. We used Bafilomycin A1 (**BafA1**) to prevent autophagy and Torin2 to induce high autophagy. After optimization, we performed poly(A) RNA-Seq on untreated PBMCs and those treated with BafA1 or Torin2. Both treatments triggered significant changes in thousands of transcripts. However, only 389 transcripts were differentially expressed in both treatments and had opposing fold changes. A suitable RNA biomarker would show an opposing effect between treatments and demonstrate a gradient of expression from low to high that overlaps with the expression range in untreated PBMCs. The top 50 transcripts, weighted by p-value, perfectly separated treated from untreated cells, and many demonstrated an appropriate expression gradient.
We used transcript-transcript correlation to help further nominate a minimum set. There were 323 transcripts that formed a total of 2,147 transcript pairs with a minimum 0.90 absolute correlation. The overall network formed more than one independent subnetwork. Using network statistics, such as weighted degree and page rank, we could rank members of each subnetwork to identify a core set of transcripts correlated with autophagy. Future work will include translating these transcript sets into RT-qPCR assays that are more scalable and cost-effective for monitoring autophagy broadly in clinical trials than flow cytometry or western blotting.

Session Title: Omics Technologies Poster Friday Session

Board 1184F: Whole genome sequencing in clinical psychiatry: *RFX3* complex rearrangement identified by long-read sequencing in a hospitalized adolescent with autism, intellectual disability and behavioral decompensation

Authors:

A. Ramalingam[1], M. Mortazavi[2], S. Batalov[1], J. Lenberg[1], C. Blucher[1], A. Omorodion[2,3], D. Helbling[1], L. Van Der Kraan[1], Z. Bezares-Orin[1], M. Bainbridge[1,4], J. Sebat[2,5,6], A. Besterman[1,2,3]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]UCSD Dept. of Psychiatry, La Jolla, CA, [3]Rady Children's Hosp. San Diego, San Diego, CA, [4]Codified Genomics LLC, Houston, TX, [5]Dept. of Cellular and Molecular Med. and Pediatrics, UCSD, La Jolla, CA, [6]Inst. for Genomic Med., UCSD, La Jolla, CA

Abstract:

In recent years, genome wide sequencing such as whole exome (WES) or whole genome sequencing (WGS) has been recommended as a first-tier test for individuals with unexplained developmental delay, intellectual disability, congenital anomalies, and epilepsy. Genetic testing is widely used for patients with neurodevelopmental disorders resulting in reported diagnostic yields of up to 30-40%; however, historically it is used in a limited scale in clinical psychiatry. Traditional WGS, using short-read sequencing (SRS) chemistries, has several technical limitations as well. Long-read sequencing (LRS), while not yet widely used for clinical testing, holds great promise as a comprehensive genetic test that can detect an even larger range of genetic variants. A 17-year-old male with a history of autism spectrum disorder (ASD), intellectual disability (ID), anxiety, irritability, gastroesophageal reflux, and amblyopia presented with worsening anxiety, irritability, intermittent suicidal thinking and decreased sleep over the past year. Patient's father had a history of generalized anxiety disorder, but there was no other psychiatric family history. No dysmorphic features were reported except macrocephaly. Chromosomal microarray and Fragile X testing were both negative. WGS was conducted on peripheral blood DNA on an Illumina NovaSeq 6000 platform. Analysis of the patient's WGS data revealed a ~14 kilobase genomic duplication of exon 5 in *RFX3* gene on chromosome 9. *RFX3* haploinsufficiency is associated with ASD and ID. Another, larger, ~291 kilobase duplication in the adjacent *GLIS3* gene (associated with neonatal diabetes) separated by 323 kilobases of intervening sequence was also detected. Analysis of parental samples revealed that the unaffected mother had mosaic duplication of this region, while the father had normal copies. However, it remained unclear if the duplication was in tandem or not and if it could result in loss-of-function of *RFX3*. To interrogate the changes in chromosome structure more clearly, LRS for proband and mother was obtained that revealed a complex rearrangement in the maternal copy, which consists of a duplication in conjunction with a deletion intersecting *RFX3* and *GLIS3*. This rearrangement results in partial deletion of *RFX3* and generation of a *GLIS3/RFX3* fusion gene, suggesting there is no functioning *RFX3* in the maternal copy, while maintaining a normal copy of *GLIS3* (of paternal origin). LRS allowed us to detect a complex rearrangement (that was missed by SRS) affecting an important brain-expressed gene with an associated disease that closely aligns with the phenotype of the patient with psychiatric disorder.

Session Title: Omics Technologies Poster Friday Session

Board 1185F: Detecting Circadian Rhythmicity by Single-Cell Multiomics

Authors:

Y. Jiang, **C. Tong**, J. Menet; Texas A&M Univ., College Station, TX

Abstract:

From bacteria to humans, most organisms showcase inherent 24-hour circadian rhythms. Best exemplified with the sleep-wake cycle, these rhythms are remarkably widespread, governing hormonal, metabolic, physiological, and behavioral oscillations. Circadian rhythms are generated by “molecular clocks” that drive the rhythmic expression of thousands of genes throughout the body. Here we generate single-cell RNA and ATAC multiomic data to simultaneously characterize gene expression and chromatin accessibility of ~33,000 mouse liver cells across different circadian timepoints at a resolution never achieved before. Our study yields several key insights, including: (i) inferring circadian rhythmicity in both discretized liver cell types and transient sub-lobule cell states, and identifying space-time RNA and ATAC profiles in a cell-type- and cell-state-specific manner; (ii) delving beyond mean cyclic patterns to characterize distributions by accounting for the gene expression stochasticity due to transcriptional bursting; (iii) interrogating multimodal circadian rhythmicity, encompassing both RNAs and DNA regulatory elements, while examining priming/lagging effects across different modalities; (iv) inferring gene regulatory logic involving target gene, transcription factor, and cis-regulatory elements that controls circadian rhythmicity. Altogether, our research constructs a comprehensive map of the transcriptomic and epigenomic landscapes that underpin the function and mechanism of the liver peripheral clocks.

Session Title: Pharmacogenomics Poster Friday Session

Board 7001F: Analysis and diagnostic interpretation of therapeutic factor VIII plasmid in hemophilia A using logistic and exponential regressions

Authors:

N. Sendil Kumar; The Charter Sch. of Wilmington, Wilmington, DE

Abstract:

Hemophilia A is a genetic disorder caused by a missing to defective *factor 8 (F8)*, affecting 1 in every 5,000 male births. *F8* loss leads to improper blood coagulation, resulting in various symptoms. Gene therapy is often implemented to treat patients, tasked with delivering a functional/therapeutic gene for the replacement of a defective/missing protein. A F8 plasmid injection is used as a prophylactic infusion of *F8* to prevent spontaneous bleeding. Forming a purpose of analyzing the effect the days after F8 plasmid injection on Anti-F8 antibodies and *F8* activities is crucial in providing the best healthcare for patients. Following a common

procedure of regression analysis, analyzing the correlation of the data, estimating a prime function, and evaluating the validity, usually with residual plots. Based on our analysis our procedure yielded data representing the relationship between days after F8 plasmid injection & *F8* activities as an exponential decay function, and the days after an F8 plasmid injection and Anti-F8 antibodies as a logistic growth function. Acquired from our data it was recognized/concluded that at a certain point, the Anti-F8 antibodies overwhelmed the *F8* activities, causing increased distress for hemophiliac patients. Evaluating the regression analysis, an optimum day for re-administration of F8 plasmid injection can be determined. Regression evaluation research helps more than 200,000 individuals globally benefit from therapeutic factor 8 plasmid injection and gene therapy, enlisting them for absolute freedom.

Session Title: Pharmacogenomics Poster Friday Session

Board 7002F: Analysis of genetic variation among pharmacogenes in Sardinian population

Authors:

M. Idda[1,2], M. Zoledziewska[2], S. Urru[3], G. McInnes[4], A. Bilotta[5], V. Nuvoli[5], V. Lodde[1], S. Orrù[6], D. Schlessinger[7], F. Cucca[1], M. Floris[1]; [1]Univ. of Sassari, Sassari, Italy, [2]IRGB-CNR, Cagliari, Italy, [3]Hosp. Pharmacy Unit, Trento Gen. Hosp., Trento, Italy, [4]BioMed. Informatics Training Program, Stanford Univ., Stanford, CA, [5]Dept. of BioMed. Sci., Sch. of Med., Univ. of Sassari, Sassari, Italy, [6]Med. Genetics, Dept. of Med. Sci. and Publ. Hlth., Univ. of Cagliari, Cagliari, Italy, [7]Lab. of Genetics and Genomics, Natl. Inst. on Aging, NIH, Baltimore, MD

Abstract:

Pharmacogenetics (PGx) identify genetic factors defining inter-individual differences in response to drug treatment maximizing efficacy while decreasing the risk of adverse events. Evaluating the prevalence of PGx variants involved in drug response, is a critical preparatory step for large-scale implementation of a personalised medicine program in a target population. Here, we profiled fourteen clinically relevant pharmacogenes in 1577 unrelated sequenced Sardinians, an ancient island population that accounts for genetic variation in Europe as a whole, and, at the same time is enriched in genetic variants that are very rare elsewhere. To this end, we used PGxPOP, a PGx allele caller based on the guidelines created by the Clinical Pharmacogenetics Implementation Consortium (CPIC), and PharmCAT, to identify the main phenotypes associated with the PGx alleles most represented in Sardinians. Briefly, our analysis revealed that 99.43% of Sardinian individuals might respond atypically to at least one drug, that on average each individual is expected to have an abnormal response to about 17 drugs, and that for 27 drugs

the fraction of the population at risk of atypical responses to therapy is more than 40%. Finally, we identified 174 pharmacogenetic variants for which the minor allele frequency was at least 10% higher among Sardinians as compared to other European populations, a fact that may contribute to substantial interpopulation variability in drug response phenotypes. This study provides baseline information for further large-scale pharmacogenomic investigations in the Sardinian population and underlines the importance of PGx characterisation of diverse populations, such as Sardinians.

Session Title: Pharmacogenomics Poster Friday Session

Board 7003F: Association of *SULT2A1* with abiraterone clearance in the Alliance A031201: Randomized phase III study of enzalutamide compared with enzalutamide plus abiraterone for metastatic castration-resistant prostate cancer

Authors:

N. Norton[1], N. B. Larson[2], G. D. Jenkins[2], J. H. Beumer[3], B. Langevin[4], J. Gobbaru[4], M. J. Morris[5], Y. Nakamura[6], M. Kubo[7], D. Kroetz[8], H-J. Zhu[9], P. H. O'Donnell[6], L. D. Lewis[10], D. L. Hertz[9]; [1]Mayo Clinic FL, Jacksonville, FL, [2]Alliance Statistics and Data Management, Mayo Clinic, Rochester, MN, [3]Univ. of Pittsburgh, Pittsburgh, PA, [4]Univ. of Maryland, Baltimore, MD, [5]Mem. Sloan Kettering Cancer Ctr., New York, NY, [6]Univ. of Chicago, Chicago, IL, [7]Riken Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan, [8]Univ California San Francisco, San Francisco, CA, [9]Univ. of Michigan Coll. of Pharmacy, Ann Arbor, MI, [10]The Geisel Sch. of Med. and Dartmouth-Hitchcock Med. Ctr., Lebanon, NH

Abstract:

Background: Enzalutamide and abiraterone acetate both target androgen receptor signaling and improve survival in metastatic castration-resistant prostate cancer (mCRPC). Identification of genetic variants associated with inter-individual clearance rates of these drugs may aid in improved dosing regimens and reduction of therapy-related toxicities. **Methods:** We performed multiple association studies of estimated enzalutamide and abiraterone clearance in participants in the Alliance for Clinical Trials in Oncology A031201 clinical trial. Genome-wide genotyping was performed on participants using the Illumina OmniExpress Exome array and imputed using the TOPMed Imputation Server. Pharmacogene metabolic phenotypes were estimated for each participant from array data using PyPGx and Stargazer. Associations of the metabolic activity groups for *CYP3A4*, *CYP3A5*, *CYP2C19* and *SLCO1B1* with enzalutamide

clearance (N=853) and activity groups of *CYP3A4*, *SLCO2B1* and *UGT1A4* with abiraterone clearance (N=393) were tested by linear regression. Targeted SNP associations were assessed for abiraterone clearance at loci proximal to major metabolizing genes *SULT2A1*, *HSD3B1* and *HSD3B2*. Finally, full genome-wide association studies (GWAS) were performed for both sets of clearance values. Associated variants were assessed as expressed quantitative trait loci (eQTLs) in liver tissue in the GTEx database and as protein (pQTLs) in liver tissue in an independent dataset of 288 individuals. **Results:** No significant associations were identified between the estimated metabolic phenotypes of candidate pharmacogenes and clearance of enzalutamide or abiraterone. SNPs in the *SULT2A1* 5' flanking region were associated with lower abiraterone clearance, with the top result corresponding to rs296373-A allele (minor allele frequency=0.15, beta = -0.457, p=3.2E-6). In the GTEx database, rs296373-A allele is an eQTL demonstrating lower gene expression in liver, (p=4.63E-06). rs2637125 and rs182421 in the *SULT2A1* 5' flanking region were significantly associated with both lower SULT2A1 protein expression (p=5.41E-09) and lower abiraterone clearance (p=3.2E-06). **Conclusions:** *SNP* rs296373 and other *SULT2A1* eQTLs are associated with lower gene and protein expression, leading to lower abiraterone clearance and suggesting that these patients could receive lower doses if other studies demonstrate an association with toxicity. **Support:** U10CA180821, U10CA180882, U24CA196171; **ClinicalTrials.gov Identifier:** NCT01949337

Session Title: Pharmacogenomics Poster Friday Session

Board 7004F: Deep mutational scan of the pharmacogene *CYP2D6* in *Saccharomyces cerevisiae* ⭐

Authors:

G. Ferra, S. Showman, E. Wong, R. Geck, C. Amorosi, K. Harris, M. Dunham; Univ. of Washington, Seattle, WA

Abstract:

Pharmacogenes are responsible for metabolizing exogenous substances such as food, pollutants, and drugs. Due to this, genetic variability in pharmacogenes can help determine proper dosing for certain drugs. For instance, if an individual has a decreased-function pharmacogene allele, the standard dose for a certain drug might not be effective and the individual may need a higher dose. *CYP2D6* is a pharmacogene responsible for metabolizing about 25% of common drugs, including SSRIs, beta-blockers, and opioids. Specifically, individuals with an ultra-rapid metabolizer *CYP2D6* allele are at a greater risk of opioid

overdose. However, many of the genetic variants of *CYP2D6* have unknown consequences, making it difficult to provide clear guidance to patients and clinicians. My project aims to measure the function of thousands of alleles, both known and yet to be discovered, to address this gap. To begin to understand the functional effects of different *CYP2D6* variants, we conducted a deep mutational scan (DMS) of *CYP2D6* using click-seq in *Saccharomyces cerevisiae*. Click-seq uses an activity-based probe to measure enzyme activity-after the activity-based probe reacts with enzymes present in the yeast cells expressing *CYP2D6* variants, we introduce a fluorophore that binds to the probe and fluoresces. Fluorescence intensity is then expected to be proportional to the activity of the targeted enzyme within the yeast cell. Using click-seq, we have obtained function scores for 6,698 of the possible 9,940 missense variants in *CYP2D6*. Previous assays were only able to characterize 9 missense variants as WT-like, decreased-function, and no function, and our scores largely agree with these previous inferences. Additionally, we have started to explore the functional effects of double mutants to determine if their effects can be inferred from our previous results on single mutants. Since most *CYP2D6* alleles in the human population are complex alleles, demonstrating that their activities can be predicted based on their constituent single mutations could significantly enhance clinical decision-making. Investigating how different *CYP2D6* variants function in yeast cells will bring us one step closer to being able to successfully use pharmacogenomics to help determine proper dosing for drugs.

Session Title: Pharmacogenomics Poster Friday Session

Board 7005F: Distinct HLA Class I Allele Clusters with Shared Peptide-binding Specificities Predispose to Co-trimoxazole-induced Severe Cutaneous Adverse Reactions ⭐

Authors:

E. Phillips[1,2], Y. Li[2], A. Mohammadali[3], R. Dodiuk-Gad[4,5], A. Drucker[5], E. Ergen[6], M. Goh[7], R. Micheletti[8], B. Kaffenberger[9], M. Marks[1], A. Yu[1], E. Williams[1], A. O'Connor[1], D. Ostrov[10], M. Martin-Pozo[1], S. Mallal[1], A. Gibson[2], H. Saeed[3]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Inst. for Immunology and Infectious Diseases, Murdoch, Australia, [3]Univ. of Illinois, Chicago, IL, [4]Technion Inst. of technology, Haifa, Israel, [5]Univ. of Toronto, Toronto, ON, Canada, [6]Univ. of Tennessee, Knoxville, TN, [7]Univ. of Melbourne, Melbourne, Australia, [8]Univ. of Pennsylvania, Philadelphia, PA, [9]Ohio State Univ., Columbus, OH, [10]Univ. of Florida, Gainsville, FL

Abstract:

Background: Co-trimoxazole (CoT) is a common antibiotic associated with severe cutaneous adverse drug reactions (SCAR) including Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) and drug reaction with eosinophilia and systemic symptoms (DRESS). HLA-B*13:01 has been associated with CoT SCAR in Southeast Asia however the generalizability is unknown.Objective: To examine HLA risk factors associated with co-trimoxazole SCAR in the United States (US).Methods: High-resolution HLA class I and II typing was performed on prospectively collected dermatologist-adjudicated US patients with CoT SCAR (n=56). HLA class I and II was imputed by SNP2HLA from expanded multi-ethnic genotyping array (Mega^{Ex})from population controls (n=94489). HLA alleles amongst CoT SCAR cases and controls were compared using Fisher's exact test. A multivariate logistic regression model assessed the contribution of specific risk HLA alleles. HLA alleles carried by CoT SCAR patients were clustered based on known peptide binding specificity (MHC NEtpan4.1). Results: HLA-B*44:03, HLA-B*38:01 and HLA-C*04:01 showed the strongest association with CoT SCAR (P<0.05, FDR), and independently contributed to risk of CoT SCAR in a multivariate model with HLA-B*44:03 showing the strongest association (P=1.07E-05, FDR). Cluster 1 and Cluster 2 HLA-B alleles that share peptide binding specificity with HLA-B*44:03 and HLA-B*38:01 respectively were associated with SCAR (P<0.001). HLA-B*13:01, shares peptide binding specificities with HLA-B*44:03 and is part of Cluster 1, however it was not carried in our CoT SCAR cases. Carriage of HLA Cluster 1, 2 alleles and HLA-C*04:01 accounted for 53/56 (95%) of risk in CoT SCAR patients. Based on the estimated CoT SCAR incidence of 0.9%, HLA-B*44:03 shows PPV and NPV of 2.3% and 99.3%, respectively, with 444 individuals needing to be screened to prevent one case of SCAR. Conclusions: Distinct HLA class I alleles, including HLA-B*44:03, HLA-B*38:01 and HLA-C*04:01,strongly associate with US CoT SCAR, explaining most cases. Clustering of HLA alleles with shared peptide binding specificities and risk of CoT SCAR may explain different HLA associations across diverse populations and give insights into mechanisms and models for HLA drug interactions.

Session Title: Pharmacogenomics Poster Friday Session

Board 7006F: Exploring allelic variation at structurally variant and medically relevant loci in the *All of Us* Research Program long-read genomes ⭐

Authors:

J. Wertz[1], W. T. Harvey[1], A. Radhakrishnan[1], S. McGee[1], I. Xu[2], K. Garimella[3], S. Huang[3], M. Danzi[2], J. D. Smith[1], All of Us Research Program, E. E. Eichler[1,4]; [1]Dept. of Genome Sci., Univ. of Washington Sch. of Med., Seattle, WA, [2]Dr. John T. Macdonald Fndn. Dept. of Human Genetics

and John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Howard Hughes Med. Inst., Seattle, WA

Abstract:

Accurate precision medicine requires genomic data from a diverse set of individuals to characterize uncommon or population-specific genotypes for disease risk. The *All of Us* (AoU) program aims to create an extensive collection of genetic and linked EHR data from 1 million Americans. As part of a pilot project, we recently assembled 1,027 haplotype-resolved African American genomes from low-pass PacBio HiFi long-read sequencing (LRS) data. We evaluated pharmacogenomic and tandem repeat disorder loci that are challenging to investigate using short-read sequencing (SRS) to assess the added value of LRS to the program's long-term objective.

CYP2D6 catalyzes the metabolism of >18 commonly prescribed drugs and is encoded by a gene with frequent structural variation (SV) affecting appropriate drug dosage, suggesting potentially actionable findings for many participants. We developed a k-mer-based approach to rapidly characterize SV haplotypes from Hifiasm-assembled *CYP2D6/7* phased contigs, computed novel CYP2D6 missense variants using Augustus, and compared functional predictions with Stargazer/PyPGx. In terms of SV diversity, we identified 86 *CYP2D7*, 19 *D6-D6-D7*, 19 *D6-D6*, 5 *D6*, and 2 *D7-D7* structures from 1,266 haplotypes and observed 15 haplotypes with apparent fusion, including 6 *D6-D7::D6*, 5 *D6-D7::D6-D7*, and 4 *D7::D6-D7* structures (3' to 5'). Applying star allele assignments to the non-SV haplotypes, we catalog 212 missense variants in predicted coding sequence beyond the assigned star allele, 2 of which have an AlphaMissense classification of likely pathogenic and 16 of ambiguous. When LRS coverage is low (~10x), correspondence with state-of-the-art SRS predictions is ~94%; however, when HiFi sequence coverage exceeds 25x, functional predictions are >99% consistent, but with the added value of discovery of hundreds of novel variants of potential functional effect.

We developed a graph-based approach to rapidly characterize repeat expansion loci in *FMR1*, *HTT*, and *C9ORF72* implicated in Fragile X syndrome, Huntington's disease (HD), and ALS. From 1,346-1,487 sequence-resolved haplotypes, we identify 3 in *FMR1* premutation range with 57, 73, and 89 repeats; 1 with 36 repeats in *HTT* indicating low-penetrance HD; and 1 with an intermediate *C9ORF72* length of 24 repeats. We identify *FMR1* haplotypes with atypical CGG repeat interruptions, including 1 with a CTG, 6 with 4 AGG, and 1 with 5 AGG interruptions, which we predict transmit stably despite near-premutation lengths.

This analysis highlights the importance of LRS data for research and potential return-of-results for biomedically relevant SV loci for underserved populations in the AoU biobank.

Session Title: Pharmacogenomics Poster Friday Session

Board 7008F: *FTH1* gene knockout induce immunomodulatory drug resistance in multiple myeloma cells through disruption of endoplasmic reticulum stress and calcium signaling pathways.

Authors:

A. Sharma, L. Pathangey, R. Fonseca, S. Swaminathan; Mayo clinic, Phoenix, AZ

Abstract:

Introduction: Immunomodulatory agents (IMiDs) are a cornerstone in the successful management of multiple myeloma. However, acquired IMiD resistance leading to disease relapses remains a major barrier. Hydrogen peroxide generation and oxidative stress are a key mechanism for IMiDs effectiveness in multiple myeloma. Iron plays a key role in the generation of oxidative stress and therefore cellular iron levels are carefully governed. Heavy chain ferritin (*FTH1* gene) is the major iron storage protein that tightly regulates cellular iron availability. Hence, the present study is targeted to investigate the role of *FTH1* in MM and IMiD resistance. **Methods:** We used CRISPR-cas9-mediated knockout of *FTH1* gene from two different multiple myeloma cells lines: IMiD sensitive KMS11 and IMiD resistant RPMI8226 cell lines. Effect of *FTH1* gene knockout were analyzed with a combination of experiments including RNA seq, flow cytometry, cell proliferation assay, RT-PCR, and western blot. The advanced data analysis tools were utilized to for data interpretation. **Results:** Initial analysis showed a positive correlation between Ferritin-H expression levels with IMiD resistance across MM cell lines. FTH1-KO reduced IMiD sensitivity in KMS11 cell line. However, FTH1-KO had no effect in altering the phenotype of IMiD resistant RPMI8226 cell line. RNA seq data indicated reduced activation of genes involved in endoplasmic stress mediated pathways and calcium signalling after *FTH1*-KO. KMS11-*FTH1*KO cells showed a significant decrease in intracellular ROS, labile iron, and mitochondrial superoxide levels compared to the WT-KMS11 cells. **Conclusion:** Our data provides evidence of a link between *FTH1* gene, intracellular labile iron levels, and immunomodulatory drug resistance in multiple myeloma cells. Moreover, *FTH1*KO results in disruption of endoplasmic reticulum stress and calcium signalling pathways.

Session Title: Pharmacogenomics Poster Friday Session

Board 7009F: Genetic predictors of statin effectiveness, adherence to treatment, and adverse clinical outcomes: longitudinal analysis of UK Biobank electronic health records

Authors:

D. Turkmen[1], X. Liang[1], D. Gill[2], J. Delgado[1], J. Masoli[1,3], L. Pilling[1], J. Bowden[1,4]; [1]Univ. of Exeter, Exeter, United Kingdom, [2]Imperial Coll. London, London, United Kingdom, [3]Dept. of Hlth.care for Older People, Royal Devon Univ. Hlth.care NHS Fndn. Trust, Exeter, United Kingdom, [4]Dept. of Genetics, Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom

Abstract:

Background
Low-density lipoprotein cholesterol (LDL-c) is a modifiable risk factor for cardiovascular disease. Statins are commonly prescribed to lower LDL-c, with clinical guidelines (e.g., the American Heart Association) recommending 30-50% reduction for ages 40-75, yet many individuals do not achieve this. We explored the impact of pharmacogenetic variants, polygenic scores (PGS), and adherence (proportion of days covered [PDC]: the number of prescribed tablets divided by the prescription length in days) on LDL-c reduction within 1 year of starting statin treatment using real-world data. We also investigated adverse long-term clinical outcomes.
Methods
We longitudinally analysed 60,000 UK Biobank participants of European genetic ancestry prescribed atorvastatin or simvastatin in primary care: 32,000 had LDL-c measurements before statin initiation (median 16 days IQR 28) and within a year of starting treatment (median 89 days IQR 125). PGS included 8 traits (e.g. LDL-c, BMI) and 28 diseases (e.g. Schizophrenia, stroke) derived by Genomics PLC. The long-term clinical outcomes (n=84) included all diseases lasting more than 3 months with prevalence >0.5% including cardiovascular conditions.
Results
Genetic predictors of LDL-c reduction following statin initiation included the pharmacogenetic variant rs4149056 (*SLCO1B1**5), and genetic liability to cardiovascular diseases. We found a U-shaped relationship between LDL-c reduction and the time to first follow-up, with a maximum reduction between weeks 5-8 of 26% (95%CI 23-29). The genetic predictors of reduced effectiveness were independent of time.Higher adherence (PDC) was associated with a larger LDL-c reduction (independent of follow-up time) and lower incident cardiovascular disease (e.g., coronary heart disease HR 0.74, 95%CI 0.66-0.83). We found preliminary evidence that higher PDC increased the risk of iron deficiency anaemia (HR 1.44, 95%CI 1.17-1.77) and incident cataracts (HR 1.33, 95%CI 1.17-1.53), though more work is needed to understand the potential adverse effects. Genetic predictors of PDC included liability to schizophrenia, which reduced PDC ($Coef_{top}$ 20% -1.94, 95%CI -2.69 to -1.19), whilst genetic liability to cardiovascular diseases increased PDC. The other PGS were not significant after multiple-testing correction.
Conclusion
Predictors of reduced statin effectiveness include pharmacogenetic variants, polygenic scores, and adherence. Tailoring statin therapy could optimise treatment efficacy and safety, leading to

better long-term clinical outcomes for patients. We plan to identify further genetic predictors of PDC through genome-wide associations.

Session Title: Pharmacogenomics Poster Friday Session

Board 7010F: Genetic susceptibility of *HLA* alleles to beta-lactam antibiotics allergy in the Taiwanese population

Authors:

I-C. Chen[1], W. Chih-Chun[2], S. Ching-Hui[2], L. Guan-Cheng[1], C. Yi-Ming[3]; [1]Dept. of Med. Res., Taichung Veterans Gen. Hosp., Taichung, Taiwan, [2]Dept. of Anesthesiology, Taichung Veterans Gen. Hosp., Taichung, Taiwan, [3]Div. of Allergy, Immunology and Rheumatology, Dept. of Med. Res., Taichung Veterans Gen. Hosp., Taichung, Taiwan

Abstract:

Background Penicillin and cephalosporin, collectively known as beta-lactam antibiotics, represent a widely utilized class of antibiotics today. The hypersensitivity reactions associated with beta-lactam antibiotics constitute a multifaceted phenomenon, with the genetic and immunological mechanisms involved yet to be fully elucidated. Genetic associations with beta-lactam antibiotic allergies are not yet recognized for prevention or diagnosis, particularly in East Asian populations. This study aimed to investigate whether specific HLA alleles are associated with an increased risk of adverse reactions to beta-lactam antibiotics in the Taiwanese population. **Methods** This retrospective, hospital-based case-control study encompassed 17,817 participants with penicillin exposure and 27,933 with cephalosporin exposure, drawn from the public medical center. All participants underwent genotyping and imputation for precise HLA typing. Logistic regression analyses were employed to investigate associations between HLA genotypes, comorbidities, and allergy risk, with additional subgroup analyses conducted to examine the relationship between HLA genotypes, comorbidities, and the severity of allergic reactions. **Results** Our study assigned 496 cases to the penicillin allergy group with 4,960 matched controls, while 278 individuals were identified as having cephalosporin allergy with 2,780 corresponding controls. The risk of penicillin allergy showed a significant increase in individuals with HLA-DPB1*05:01 (OR = 1.38, p=0.004) and HLA-DQB1*05:01 (OR = 1.64, p=0.03), while urticaria emerged as an independent risk factor (OR = 1.75, p<0.001). Our findings also revealed specific HLA alleles associated with an increased risk of cephalosporin allergy reactions, including HLA-B*55:02 (OR = 1.76, p=0.005), HLA-C*01:02 (OR = 1.36, p=0.018), and HLA-DQB1*06:09 (OR = 2.58, p<0.001). Moreover, the HLA-C*01:02 allele was significantly correlated with a heightened risk of severe cephalosporin allergy (OR = 2.33,

p=0.04). However, neither the HLA alleles nor the comorbidities had a significant relationship with the risk of severe penicillin-induced allergy. **Discussion** We identified HLA genotypes linked to the onset and severity of beta-lactam antibiotics allergy. Our findings establish a basis for precision prescription in future clinical applications.

Session Title: Pharmacogenomics Poster Friday Session

Board 7011F: Genetically proxied therapeutic inhibition of kidney function drug targets and type 2 diabetes in Africans: A Mendelian randomisation study ⭐

Authors:

A. Diawara[1], M. Diakite[1], S. O. Doumbia[1], O. Soremekun[2], S. Fatumo[2]; [1]Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, [2]MRC/UVRI & LSHTM Uganda Res. Unit, Entebbe, Uganda

Abstract:

Background: The risk of developing type 2 diabetes (T2D) in Africans with kidney disease treated by medications remains unclear. This study explores this potential interaction within the African genetic context. Methods: We employed Mendelian randomization, a genetic proxy approach, utilizing data from African participants. This assessed the causal link between genetically predicted inhibition of specific pathways and T2D risk. Findings: Our analysis revealed an association between genetically predicted inhibition of VEGFA and RHEB pathways and increased T2D risk in Africans (OR, 2.66; 95% CI 1.34-3.78, and OR, 2.25; 95% CI, 1.34-3.28, respectively). Conversely, no significant association was found with pathways involving SLC22A2 or CLDN14(OR, 0.95; 95% CI, 0.61-1.48 and OR, 1.56; 0.71-2.20, respectively). Implications: These findings suggest that some kidney disease medications targeting VEGFA and RHEB pathways might exacerbate T2D risk in African populations. This highlights the importance of a nuanced understanding of drug-disease interactions specific to the African context when selecting treatment options for kidney disease. Further research with wider and more diverse African populations is crucial to solidify these results and elucidate the underlying biological mechanisms.

Session Title: Pharmacogenomics Poster Friday Session

Board 7012F: Genetics of glucocorticoid-induced hyperglycemia

Authors:

V. Kawai[1], X. Ruan[1], M. Williams[2], G. Karakoc[3], E. McNeer[2], L. Choi[1], **J. Mosley**[4], M. Stein[4]; [1]Vanderbilt Univ Med Ctr., Nashville, TN, [2]Vanderbilt Univ., Nashville, TN, [3]Vanderbilt Univ. Med. center, Nashville, TN, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Background: Glucocorticoids (GCs) are potent, commonly prescribed anti-inflammatory drugs with large variability in response that is not well understood. GC response is known to have a genetic component, but the genetic determinants of this variability are largely unknown. We aimed to investigate whether and to what extent genetic factors are associated with GC-induced hyperglycemia.

Methods: To test our hypothesis, we selected a cohort of individuals who had received a trigger dose of 40 mg/d prednisone or more at age 18 years or older as part of clinical care, had glucose measurements before and after GC administration, and had genome-wide genotyping available in the Vanderbilt University Medical Center biobank. The highest glucose level recorded within 30 days after the trigger dose was extracted. Individuals with a diagnosis of diabetes before the trigger dose were excluded. Only individuals of European ancestry (defined by genetic principal components (PCs) and HapMap reference) were studied. The highest post-GC glucose level was natural log transformed and a genome-wide association analysis (GWAS) was performed using linear regression and an additive model adjusting by sex, age, GC dose, and 5 PCs. P<5.0E-8 was considered significant. A polygenic risk score (PRS) for type 2 diabetes (T2D) was built using summary data from the largest genetic study in individuals of European ancestry and calculated in the study cohort. A linear regression between the post-GC glucose and the standardized PRS-T2D was performed adjusting for the same covariates as in the GWAS; P<0.05 was considered significant. The median baseline glucose level before GC was included as a covariate in a sensitivity analysis.

Results: There were 1333 eligible individuals. Fifty-two percent were male, the median [IQR] age and trigger dose was 56 [43, 65] years and 60 [40, 200] mg/d, respectively. Blood glucose increased from a baseline of 116 mg/dL to 158-mg/dL after GC. The GWAS showed a significant intronic SNP in *ESRRG* (P=3.67E-08) and 6 independent (R2<0.1) nominal associations (P<1.0E-05) that mapped to *ZDHHC13*, *ALOX5AP* and *MYH14*. The PRS-T2D was significantly associated with higher glucose levels after GC (estimate=0.05, P=5.46E-09); the signal remained significant after including baseline glucose as covariate (estimate= 0.04, P=4.0E-04)

Conclusions: In individuals without T2D, glucose levels after a high dose of GC are associated with genetic variation in *ESRRG* and with genetic predisposition to T2D.

Session Title: Pharmacogenomics Poster Friday Session

Board 7013F: Germline homologous recombination deficiency influences the fitness impact of oncologic therapy on TP53 clonal hematopoiesis

Authors:

I. Chan[1], J. T. Baeten[1], L. A. Moukarzel[2], J. Liu[1], D. Tran[1], J. Tabs[1], G. Petrone[1], A. Carter[1], M. Patel[2], K. Stopsack[3], P. Pharoah[4], J. Brenton[5], B. Li[2], A. Schram[2], W. Abida[2], B. Weigelt[2], M. Mullen[1], C. Cruchaga[1], H. Scher[2], R. Levine[2], E. Papaemmanuil[2], K. Cadoo[6], D. C. Link[1], K. L. Bolton[1]; [1]Washington Univ. in St. Louis, Saint Louis, MO, [2]Mem. Sloan Kettering, New York City, NY, [3]Harvard T. H. Chan - Sch. of Publ. Hlth., Boston, MA, [4]Cedars-Sinai Med. Ctr., Los Angeles, CA, [5]Univ. of Cambridge, Cambridge, United Kingdom, [6]St. James's Hosp., Dublin, Ireland

Abstract:

Poly (ADP-ribose) polymerase inhibitors (PARPi) have shifted treatment paradigms for tumors harboring defects in the homologous recombination repair (HRR) pathway. Concerns have emerged regarding increased incidence of therapy-related myeloid neoplasia (tMN) after PARPi use. These observations are potentially confounded by variables including exposure to other cytotoxic therapies and germline HRR mutations. *TP53* is most commonly mutated in tMN and considered a hallmark of the disease. With age, hematopoietic stem and progenitor cells (HSPCs) accumulate somatic mutations that result in their clonal expansion (ie. clonal hematopoiesis, CH) including pathogenic mutations in *TP53*. As *TP53* CH is the origin of most tMNs, studying the effects of PARPi therapy on CH would characterize the extent PARPi increase the risk of tMN independent of other confounding variables. Initially, exploring 10,156 solid tumor patients with complete treatment histories, revealed increased *TP53* CH following PARPi therapy (OR 4.4; $p = 2e^{-4}$). Adjusting for cumulative dosage of other cytotoxic therapies attenuated the association (OR 1.5; p = 0.07). Peripheral blood samples from individuals receiving PARPi (n = 100), carboplatin (n = 142) and matched healthy controls (n = 176) sequenced at an average depth of 18,946x with a custom CH panel, showed elevated *TP53* CH during carboplatin ($p = 1e^{-1}$) and PARPi ($p = 1e^{-12}$) treatment, with carboplatin resulting in greater increased growth ($p = 2e^{-4}$). Interestingly, individuals with germline HRR mutations showed reduced growth of *TP53* CH during PARPi (p = 0.04) therapy. To validate, a *TP53*-mutant CH murine model was developed by transplanting a mixture of *Trp53*(R172H$^{+/-}$) mutant bone marrow cells with wild-type cells into irradiated recipients. Peripheral blood and bone marrow *Trp53*-mutant leukocytes and HSPCs were increased post-cisplatin ($p = 1e^{-4}$; $p = 1e^{-4}$)

and to a lesser extent post-PARPi (p = 0.01; p = 0.04). To model the effect of germline HRR, chimeras containing heterozygous *Brca1*-mutated HSPCs competing with *Brca1*-mutated HSPCs with *Trp53*(R172H$^{+/-}$) were generated. Consistent with human data, no observable *TP53*-mutated cells were observed after either treatment. In summary, *TP53* mutations provide a fitness advantage to HSPCs following cytotoxic therapy. Comparatively, exposure to PARPi confers decreased selective pressure on *TP53* CH. Interestingly, the fitness advantage of *TP53*-mutated HSPCs is mitigated with germline HRR mutations. PARPi may therefore have less of an impact on tMN and show synergistic effects with homologous recombination deficiency in blocking the competitive advantage of *TP53* CH during genotoxic stress.

Session Title: Pharmacogenomics Poster Friday Session

Board 7014F: Incorporating transcriptional signatures to improve genetic prioritization of drug indications.

Authors:

A. Ko; Children's Natl. Res. Inst., Washington, DC

Abstract:

Drug development is highly costly and time-consuming, and most new therapeutics will fail at the clinical trial stage. A drug target with strong human genetic evidence has been shown to have a higher rate of advancing through trials. Therefore, many tools now leverage genetic information for drug indication. For instance, the recently developed genetic priority score (GPS) utilizes eight genetic features to prioritize targets with a higher chance of success. To provide additional mechanistic insight into genetic-guided drug indications, we investigated how incorporating experimental functional genomic data can improve prediction and infer directionality. The Connectivity Map (CMap) project profiled transcriptomic changes after a perturbation, such as drug treatments or CRISPR knockouts. CMap measured or inferred the expression of nearly 13,000 genes in multiple cells and enables querying of transcriptional signatures across many conditions. We first compared the expression changes of 2,517 genes with higher GPS (>=1.5) to the rest of the genes in 11 non-cancer cell lines. The average moderated Z-scores (MODZs) differ significantly between the two groups ($P=2.28x10^{-5}$), suggesting that expressions of high GPS genes are more likely to respond to perturbagens. GPS can also incorporate direction of effect (GPS-DOE). So, we next compared the predicted GPS direction of effects to the transcriptional changes in CMap. We observed that genes with either negative or positive GPS-DOE do not show any directional biases at the mRNA level compared to the background distribution with no significant difference in mean MODZs (P>0.05). This

finding indicates that the genetic features in GPS-DOE might not fully capture the directionality of drug responses, and the addition of functional genomic data can provide orthogonal evidence to infer the direction of drug modulation. Therefore, we applied machine learning algorithms to construct a new score, GPS-TS (transcriptional signature), to utilize the rich expression profiles from CMap. In summary, we demonstrate the combination of human genetic features and perturbed transcriptomic profiles can aid in identifying drug targets more effectively.

Session Title: Pharmacogenomics Poster Friday Session

Board 7015F: Integrating WES and array data for improved pharmacogenetic typing: A comparative study with WGS

Authors:

C. Lee, S. Jeong, M. Kho, S. Lee; Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract:

In pharmacogenetic analysis, the use of whole genome sequencing (WGS) is considered the gold standard due to its comprehensive coverage of both coding and critical non-coding regions of the genome. However, the high cost of WGS compels the use of alternative methods like whole exome sequencing (WES) and array data. WES, while cost-effective, primarily captures exonic regions, leading to significant discrepancies in star-allele calling due to missed intronic and other non-coding variants that are crucial for accurate pharmacogenotyping. To overcome these limitations, this study developed a pipeline integrating WES with Human Reference Consortium (HRC)-imputed array data. This integrated approach aims to enhance concordance with WGS, thereby improving the reliability of pharmacogenotyping in resource-limited settings. We processed and merged 469,835 WES samples with 487,409 HRC-imputed array samples, all from UK Biobank, focusing on a subset of 1,000 samples common across both datasets alongside WGS samples. To merge the two data types, we first converted array data from BGEN to VCF format using Plink2. Subsequent harmonization included updating VCF headers, standardizing chromosome formatting, standardizing IDs, and reordering samples. We then lifted the array data from hg19 to hg38 using chain files via LiftOver. Once the WES and array VCFs were compatible, we integrated, sorted, and split the resulting multi-sample VCF into individual single-sample VCFs using bcftools. We conducted Aldy4 analyses on each individual sample to assess pharmacogenetic implications. Our results show that the WES+array approach significantly enhances concordance with WGS, exemplified by notable improvements in pharmacogenes such as *CYP1A1* (from 75.1% to 94.3%), *CYP2C19* (from 61.1% to

97.5%), *CYP2E1* (from 56.4% to 86.1%), *NAT1* (from 55.3% to 92.1%), *UGT2B7* (from 28.5% to 98.7%), and *VKORC1* (from 8.2% to 90.9%). This enhanced detection underscores the utility of integrating WES with array data, suggesting significant benefits for pharmacogenetic analyses in both clinical and research settings.

Session Title: Pharmacogenomics Poster Friday Session

Board 7016F: Polygenic prediction of immune checkpoint inhibitor-mediated hepatotoxicity in non-small cell lung cancer patients ⭐

Authors:

P. Middha[1], R. Thummalapalli[2], Z. Quandt[1], Princess Margaret Lung Group, M. A. Gubens[1], C. M. Lovly[3], G. Liu[4], M. Aldrich[3], A. J. Schoenfeld[2], E. Ziv[5]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Mem. Sloan Kettering Cancer Ctr., New York, NY, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Princess Margaret Cancer Ctr., Toronto, ON, Canada, [5]UCSF, San Francisco, CA

Abstract:

Background: Immune checkpoint inhibitors (ICIs), including anti-PD1/PD-L1s and anti-CTLA4s, have revolutionized cancer care. Despite these breakthroughs, treatment with ICIs is frequently accompanied by immune-related adverse events (irAEs). ICI-mediated hepatoxicity (ICI-H) occurs in 1-22% of patients, depending on the type of therapy, and can lead to treatment discontinuation. There are no established predictive factors for ICI-H. We investigated the association between genetic predisposition to autoimmune diseases in the general population and ICI-H in non-small cell lung cancer (NSCLC) patients treated with ICIs. **Methods**: Our study comprises 1,349 NSCLC patients who received ICIs across four institutions between 2009 and 2022. Genotyping was performed using the Affymetrix Precision Medicine Array and imputed to the 1000 Genomes Reference panel. Clinical data, including detailed information on irAEs, were extracted by manual chart review. Using Cox proportional hazards models, we evaluated the association between a previously published standardized PRS_{AID} (PMID: 35393596) with ICI-H (all-grade and severe). All models were adjusted for age at cancer diagnosis, sex, type of ICI therapy, recruitment site, lung cancer histology, and 5 ancestry-informative principal components. Cumulative incidence of ICI-H (all-grade and severe) was assessed by categories of PRS percentiles. **Results**: Among 1,349 ICI-treated NSCLC patients, 71 (5.3%) had all-grade ICI-H and 31 (2.3%) had severe ICI-H. The PRS_{AID} predicts all-grade ICI-H with a hazard ratio (HR) per standard deviation (SD) of 1.43 (p=0.01) and severe ICI-H (HR per SD=1.53, p=0.03). The association was stronger in individuals with high genetic risk (top 20th percentile compared to bottom 80th percentile of the PRS_{AID}) (all-grade ICI-H: HR=2.11, p=0.007; severe ICI-H: HR=2.73,

p=0.008). Individuals in the high genetic risk group had increased cumulative events of all-grade and severe ICI-H (6% all-grade ICI-H among high-risk compared to 3% among low/moderate risk, log-rank p=0.01; 4% severe ICI-H among high-risk compared to 1.4% among low/moderate risk, log-rank p=0.006). **Conclusions**: An integrated PRS for autoimmune disease can predict all-grade and severe ICI-H in patients with NSCLC. These findings suggest that PRS_{AID} may help facilitate clinical decision-making, particularly in settings in which there remains clinical uncertainty around the magnitude of predicted ICI benefit. It could also be a tool for risk stratification in clinical trials to monitor and potentially prevent irAEs.

Session Title: Pharmacogenomics Poster Friday Session

Board 7017F: Polygenic risk of depression is correlated with treatment resistance, treatment nonresponse and antidepressant class specific response across ancestrally diverse populations

Authors:

S. Lapinska[1], A. Pimplaskar[2], Y. Ding[3], C. Frydman-Gani[1], K. Hou[2], V. Venkateswaran[1], K. Boulier[1], L. Olde Loohuis[4], B. Pasaniuc[1]; [1]UCLA, Los Angeles, CA, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]Dana-Farber Cancer Inst., Boston, MA, [4]Univ. of California Los Angeles, Los Angeles, CA

Abstract:

Treatment resistant depression (TRD), defined as limited or no response to at least two antidepressants, occurs in approximately one third of patients diagnosed with major depressive disorder (MDD) and approximately half of patients are treatment non-responders (TND), defined as no response to at least one antidepressant. Studies conducted on patients of European ancestry demonstrated that there is a genetic overlap between TRD and genetic risk for MDD. Utilizing ancestrally diverse biobanks within medical systems provides an opportunity for studies understanding treatment response to MDD across diverse individuals. Utilizing UCLA ATLAS Community Health Study (ATLAS) and All of Us Research Program (AoU), we aim to test for associations between a polygenic score for major depression (MDD-PGS) and TRD, stratified by antidepressant class response and genetic ancestry.
For 14,381 MDD patients in ATLAS, 6.6% were classified as having TRD and 31.0% were TND. In AoU, out of 42,872 MDD patients, 13.5% were TRD and 24.5% were TND. Within the European American (EA) ancestry individuals in ATLAS, the MDD-PGS is increased in patients with TRD (OR: 1.16 [1.05, 1.27], p = 2.24e-3) and TND (OR: 1.10 [1.04, 1.17], p = 1.33e-3) compared to treatment responders. The same direction of effect was observed in AoU with TRD (OR: 1.20

[1.14, 1.26], p = 1.46e-14) patients and TND (OR: 1.11 [1.06, 1.15], p = 5.82e-7). We observed no association in other GIA groupings likely due to limited sample size. In our class specific analysis, MDD patients of EA ancestry that respond to SSRI or Tricyclic/Tetracyclic treatment have lower MDD-PGS than both TND and TRD patients while response to Serotonin Modulators has lower MDD-PGS than TRD. Individuals of Hispanic Latin American ancestry with higher genetic loading for MDD have lower odds of responding to treatment with SSRI (as compared to both TND (OR: 0.86 [0.75, 0.98], p = 2.68e-02) and TRD (OR: 0.76 [0.59, 0.97], p = 2.94e-02)) and Atypical Agents when compared to TRD (OR: 0.58 [0.38, 0.86], p = 7.75E-03). Elevated MDD-PGS in African American ancestry individuals showed lower odds of responding to SNRI treatment when compared to TRD (OR: 0.78 [0.60, 0.94], p = 1.37e-02).Thus, we found significantly elevated MDD-PGS in individuals of EA ancestry in both TRD and TND, consistent with a hypothesis of increased genetic risk for MDD in TRD and TND cases. Polygenic risk scores for MDD may be a useful biomarker in identifying drivers of response to antidepressants in MDD patients, but to ensure the translation of findings, genomic studies would benefit from greater participant diversity to better understand the effect of genetic ancestry on drug response.

Session Title: Pharmacogenomics Poster Friday Session

Board 7019F: Rapid and scalable personalized therapeutic screening in patient-derived organoid models of genetic disease ⭐

Authors:

S. Younger, J. Means, C. Barrett, T. Pastinen; Childrens Mercy Kansas City, Kansas City, MO

Abstract:

Genetic conditions are often associated with clinical indications for which multiple symptomatic treatments are available. However, the complexity of genetic backgrounds underlying impacted patient populations can result in unpredictable variability in therapeutic efficacy across individuals. The process of identifying the optimal therapeutic approach for any given individual can be a major bottleneck, leading to a prolonged therapeutic odyssey. A critical barrier in translational sciences is the lack of practical laboratory technologies with the capacity to inform evidence-based treatments in a timely manner. The ability to accurately predict drug efficacy in a particular individual would: i) improve the quality of life for the individual by reducing the time to effective treatment, ii) reduce financial burdens on patients/families by reducing the frequency of clinical visits, and iii) minimize preventable patient volume on healthcare systems. To achieve this goal, we have leveraged our institute's unparalleled pediatric genetic data repository, "Genomic Answers For Kids", to guide the development of a scalable functional

precision medicine platform that enables the rapid and accurate prediction of patient-specific drug efficacy. Requiring only a standard blood draw, we isolate peripheral blood mononuclear cells, reprogram those cells into induced pluripotent stem cells (iPSCs), and differentiate iPSCs into patient-specific disease-relevant organoids in a 96-well format. From blood draw to mature organoids, this process requires 6-8 weeks. Seizures are among the most commonly occurring clinical presentations observed in our rare disease patient population. For proof-of-concept we have evaluated phenotypic responses to antiseizure medications in patient-derived brain organoids using a plate reader-based fluorometric assay capable of quantifying neuronal activity. We consistently observe that organoids derived from patients impacted by seizures exhibit hyperactive neuronal activity relative to unaffected individuals. Moreover, treatment of organoids with panels of antiseizure compounds reveals differential and patient-specific response profiles (changes in neuronal activity). Notably, responsiveness to antiseizure compounds has no correlation or predictable relationship with the mechanism of action for the compounds evaluated, further emphasizing the advantage of our empirical approach. This work has provided the foundation for a functional precision medicine platform that is being further developed to predict patient-specific responses to a broad range of symptomatic treatments across diverse cell systems.

Session Title: Pharmacogenomics Poster Friday Session

Board 7020F: Reproduction of *HLA-A*32:01* and discovery of two loci near *IL-7* and *MAP2K6* associated with vancomycin drug reaction with eosinophilia and systemic symptoms in an electronic health record-based study

Authors:

M. Krantz[1], J. Qian[1], C. Bejan[1], S. Zhang[1], Y. Xu[1], W-Q. Wei[1], B. Yoon[1], J. Trubiano[2], E. Phillips[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Austin Hlth., Melbourne, Australia

Abstract:

RATIONALE: Vancomycin is the most common antibiotic associated with drug reaction with eosinophilia and systemic symptoms (DRESS). Vancomycin DRESS is strongly associated with HLA-A*32:01, however, only 20% of those who carry this allele develop DRESS suggesting the importance of genetic and other risk factors outside of the major histocompatibility complex (MHC).
METHODS: High-resolution HLA class I and II typing was performed on vancomycin DRESS cases

adjudicated and prospectively accrued from Vanderbilt University Medical Center (VUMC) and Austin Health. Vancomycin-tolerant controls, including an *HLA-A*32:01*+ subset (n=113) were identified using the VUMC BioVU (de-identified EHR linked to a DNA biobank). HLA class I and II alleles were imputed from MEGA^{EX} array typing using SNP2HLA. GWAS was performed using vancomycin DRESS cases (n=78) and vancomycin tolerant controls (n=2,702). Conditional GWAS was also performed using *HLA-A*32:01*+ vancomycin DRESS cases and *HLA-A*32:01*+ vancomycin-tolerant controls.
RESULTS: Vancomycin DRESS cases (n=78) were identified with a median age of 55 years, 57% female, and 77% *HLA-A*32:01*+. For vancomycin tolerant controls, 2,702 cases were identified with a median age of 50 years, 41% female, and 6% *HLA-A*32:01*+. Carriage of *HLA-A*32:01* was the most significant factor associated with the development of vancomycin DRESS (pc<.00001). In the conditional GWAS analysis, two significant loci outside of the MHC predicted *HLA-A*32:01*+ vancomycin tolerance. These were identified at chromosomes 8 and 17, with the nearest genes being *IL-7* and *MAP2K6*, respectively.
CONCLUSIONS: *HLA-A*32:01* remains the strongest risk factor associated with vancomycin DRESS. Two loci outside of the MHC, nearest genes *IL-7* and *MAP2K6*, potentially explain *HLA-A*32:01*+ vancomycin tolerance.

Session Title: Pharmacogenomics Poster Friday Session

Board 7021F: StarPhase: Comprehensive Phase-Aware Pharmacogenomic Diplotyper for Long-Read Sequencing Data

Authors:

J. Holt, J. Harting, X. Chen, D. Baker, N. Gonzaludo, Z. Kronenberg, C. Saunders, M. Eberle; PacBio, Menlo Park, CA

Abstract:

Introduction: Pharmacogenomics (PGx) is critically important to precision medicine, informing the use of medications at an individual level, improving both safety and efficacy. PGx diplotyping relies on the ability to both accurately detect genomic variation and phase that variation onto distinct haplotypes, commonly referred to as “star (*) alleles”. PacBio HiFi sequencing provides long reads with highly accurate base-calling, enabling variant calling and phasing for both targeted and whole-genome sequencing approaches.
Methods: We developed StarPhase, a phase-aware tool for generating comprehensive PGx diplotype calls from PacBio HiFi sequencing datasets. StarPhase accepts both phased and unphased variant calls from a HiFi sequencing pipeline (e.g., DeepVariant followed by HiPhase)

as well as an aligned BAM file to produce PGx diplotypes for 21 genes, including the complex genes HLA-A, HLA-B, and CYP2D6. In contrast to existing tools, StarPhase correctly handles genes that are fully phased as well as ambiguity from unphased variants.
Results: We compared StarPhase diplotype calls to those from other PGx diplotyping tools (PharmCAT, HiFiHLA, Pangu, and Cyrius) and to known diplotypes from GeT-RM. For simple PGx genes, StarPhase has a 98.24% concordance with PharmCAT, and all discrepancies were explained via manual inspection as either differences in reporting or corrections that StarPhase made relative to PharmCat. For HLA-A and HLA-B, all StarPhase results were 100% concordant with the assembly-based results of HiFiHLA. Finally, StarPhase calls for CYP2D6 were 100% concordant for whole genome sequencing datasets, and 96% concordant for targeted sequencing. All discrepancies in the targeted sequencing were explained through either ambiguity in GeT-RM, errors in the comparator tool, or low coverage of hybrid alleles in the raw data (likely due to reduced capture rate). We further demonstrated the utility of StarPhase by applying it to CEPH pedigree 1463, consisting of 27 whole genome sequencing PacBio HiFi datasets from four generations. In this analysis, all diplotype calls across all 21 genes were inherited consistent with the pedigree.

Session Title: Pharmacogenomics Poster Friday Session

Board 7022F: Timely Integration of Pharmacogenomics Information into the Electronic Health Records to Deliver a Clinical Decision Support Tool for Clinicians- an Agile Solution at the Veterans Health Administration Medical Centers

Authors:

R. Relova[1], A. Callahan[2], S. Patel[3], P. Joshi[4], A. Goetschius[2], A. Gant-Curtis[2], K. Teague[2], C. Zanetti[3]; [1]US Dept. of VA, Palo Alto, CA, [2]US Dept. of VA, Washington, DC, [3]US Dept. of VA, Charleston, SC, [4]US Dept. of VA, San Francisco, CA

Abstract:

Introduction- Pharmacogenomics or PGx informs how an individual's genes affect one's response to medications. This field combines pharmacology- the science of drugs, and genomics- the study of genes and their functions, to develop effective, safe medications that can be prescribed based on a person's genetic makeup. Many currently prescribed medications utilize the one-size-fits-all approach, but they don't work the same way for everyone. The Veterans Health Administration is America's largest integrated health care system, providing

care to 9 million Veteran patients each year. To date, 7,318 VHA clinicians actively order PGx tests averaging approximately 3,000 PGx reports per month across 136 VHA medical facilities. Purpose- The deployment of PGx testing at the VHA medical centers underscored the profound necessity for integration of PGx reports into electronic health records or EHR. The purpose is to demonstrate how implementing an ancillary novel application, called Data Arch, augments the capabilities of the EHR by enabling the clinicians to access PGx-guided medication and dosing recommendations at point-of-prescribing.
Summary- The PGx report serves as an essential clinical decision support or CDS tool. It includes about 70 commonly prescribed medications. It incorporates curated patient-specific drug-gene interpretations and clinical actionability of a patient's medications based on Clinical Pharmacogenetics Implementation Consortium and US Food and Drug Administration guidelines. The lack of interoperability between the laboratory and EHR presents a prodigious hindrance to leverage PGx as a CDS tool in patient care. Currently, PGx reports are manually uploaded as PDFs, which takes 6-9 days to become available. Clinicians must log-in another resource to retrieve the PGx reports. Data Arch, a VHA Innovation platform, transforms and transmits PGx data from an external source into the EHR. The PGx report note is easily accessible to the provider. Clinical alerts will fire not only for current, but also for future impacted medications. Conclusion- Pharmacogenomics is an emerging, highly actionable precision medicine approach. PGx has the potential to impact the incidence of adverse drug events, a significant cause of morbidity and mortality. The dissemination of patient-specific PGx information at point-of-prescribing adds value in clinical decision-making. As part of the successful end-to-end implementation in clinical settings, relevant PGx-guided reports can be made promptly available to clinicians in a manner that seamlessly integrates into the clinical workflow of the EHR through an ancillary IT infrastructure.

Session Title: Pharmacogenomics Poster Friday Session

Board 7023F: Using whole genome sequencing to call star alleles expands coverage and sensitivity for clinical pharmacogenomics decision-making

Authors:

D. Shapiro[1], D. Belliston[1], R. Al Rabadi[1], T. Greer[1], Z. Jodon[1], K. Ward[2], C. Hansen[1], A. Hansen[1]; [1]Geneial, Houston, TX, [2]Slopes Bio, Salt Lake City, UT

Abstract:

Clinical decision support (CDS) plays a crucial role in implementing pharmacogenomics (PGx) in precision medicine. Whole-genome sequencing (WGS) has become more broadly used for clinical purposes, yet remains underutilized for clinical PGx reporting due to costs, privacy concerns, and the complexity of data management and interpretation. Our objective is to integrate WGS-based PGx decision-making by using the Geneial privacy-preserving software platform to support physician-initiated, patient-gated encrypted querying of relevant drug recommendations. Toward this end, we report the implementation of a WGS-based PGx allele calling and annotation pipeline and its evaluation against an existing PGx clinical assay.
Using 20 de-identified individuals from a partner lab previously tested in duplicate with the Thermofisher TaqMan OpenArray PGx Express Panel, we performed WGS on the Singular Genomics G4 platform at 30X coverage. Star-allele calling was performed using the PyPGx Python package, with statistical phasing estimated by Beagle, all wrapped in an automated workflow on the Geneial platform. We then investigated concordance of the PyPGx-WGS workflow with variants/star-alleles called by the TaqMan assay in 12 genes with strong published PGx evidence. Overall star-allele concordance was 88%; 71% of calls were confirmed by PyPGx as the "primary" prioritized diplotypes and an additional 17% were confirmed as secondary alleles. Importantly, WGS outperforms TaqMan in cases where rarer variation is present or haplotype phasing is more complex. Two samples showed discordant diplotypes due to a lack of phasing information from the TaqMan assay, two more samples showed discordance due to potential allelic dropout in the TaqMan assay, and most of the remaining discordant calls (n=21) were due to the WGS pipeline calling additional rare variants not previously covered. Lastly, evaluation of predicted phenotypic concordance suggested that in five cases (25%), calling rarer star-alleles with WGS reclassified the individual's phenotype based on their updated diplotype.
Through these efforts, we seek to lay a foundation for whole-genome data re-use and secure, private data exchange between clinicians and patients. We have developed a proof of concept (for research use only) data exchange model within the Geneial web and mobile app. In this model, clinicians can query for PGx-guided drug recommendations, while patients retain control over authorizing or denying data release. Geneial remains blinded to the contents of these sensitive data queries and results via the integration of privacy-enhancing technologies.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7025F: A Family Study: A previously unreported heterozygous variant in *DCHS2* is associated with CAKUT.

Authors:

M. Seid[1], A. Johnson[2], J. Turocy[3]; [1]Univ. of Washington, Seattle, WA, [2]Kaiser Fndn. Hosp., Modesto, CA, [3]Kaiser Fndn. Hosp., Modesto CA, Ripon, CA

Abstract:

A number of variants in *DCHS2* have been reported in individuals with Congenital Abnormalities of the Kidney and Urinary Tract (CAKUT) (Nicolaou 2016). As evidenced by our recent family study, a previously unreported heterozygous variant in *DCHS2* is associated with lower urinary tract obstruction (LUTO) of varying severity in different members of the same family.A 22-year-old patient, G3 P0020, presented at 12w3d gestation for first trimester screening and anatomy ultrasound (NT, hCG, PAPP-A). The screening result was negative for trisomies 13,18, and 21. However, the anatomy ultrasound revealed a fetal bladder diameter of 9 mm,consistent with megacystis. In other respects, the patient presented with normal first trimester anatomy. The patient declined prenatal diagnosis.Subsequently, a detailed fetal anatomy ultrasound at 19w0d gestation showed normal anatomy in a male fetus. The megacystis resolved. The fetal kidneys and bladder were sonographically normal. The patient delivered a healthy male infant who has had normal growth and development.Due to the patient's prior history of fetal megacystis, an early anatomy ultrasound was scheduled for her subsequent pregnancy at age 27. Her male fetus, at 15w4d gestation,showed signs of severe LUTO. The bladder filled most of the abdomen with severe bilateraldilation of the renal collecting systems and extremely thin renal cortices SFU grade 4. There was severe oligohydramnios and fetal hydrops with ascites, pleural and pericardial effusions, and significant body edema. Fetal demise occurred at 18w0d gestation and a D and E was performed. Array cgh revealed a normal male fetus 46,XY with no CNV of clinical significance.Whole exome sequencing done by Prevention Genetics showed a maternally inherited previously unreported variant in *DCHS2* c.2854G>A.The patient's first child, with the resolved megacystis, is now age 6 years. He has a 50%chance of having the familial *DCHS2* variant. He had a normal ultrasound of the kidneys and bladder. Molecular genetic testing is pending at the time of this report. If testing confirms that the has the variant as expected, it will be further supportive evidence that this *DCHS2* variantc.2854G>A is likely pathogenic. The patient is currently pregnant. Detailed ultrasound at 20weeks showed a female fetus with normal kidneys, bladder and otherwise normal anatomy.Genetic testing for the familial variant was deferred until after birth.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7026F: A method of identifying risk genes using ultra-low coverage whole-genome sequencing - insights into embryo aneuploidy

Authors:

J. Xing[1], S. Sun[1], M. Aboelenain[1], D. Ariad[2], M. Haywood[3], C. Wageman[3], M. Duke[1], A. Bag[1], M. Viotti[4,5], M. Katz-Jaffe[3], R. McCoy[2], K. Schindler[1]; [1]Rutgers, the State Univ. of New Jersey, Piscataway, NJ, [2]Johns Hopkins Univ., Baltimore, MD, [3]CCRM Genetics, Lone Tree, CO, [4]Zouves Fndn. for Reproductive Med., Foster City, CA, [5]Kindlabs, Kindbody, New York, NY

Abstract:

Aneuploidy frequently arises during human meiosis and is the primary cause of early miscarriage and *in vitro* fertilization (IVF) failure. Individuals undergoing IVF exhibit significant variability in aneuploidy rates, although the exact genetic causes of the variability in aneuploid egg production remain unclear. Preimplantation genetic testing for aneuploidy (PGT-A) using next-generation sequencing is a standard test for identifying and selecting IVF-derived euploid embryos. The wealth of embryo aneuploidy data and ultra-low coverage whole-genome sequencing (ulc-WGS) data from PGT-A has potential for discovering variants in parental genomes that are associated with aneuploidy risk in their embryos. Using ulc-WGS data from ~10,000 PGT-A biopsies, we imputed genotype likelihoods of genetic variants in embryo genomes. We then used the imputed variants and embryo aneuploidy calls to perform a genome-wide association study of aneuploidy incidence. Finally, we carried out functional evaluation of the identified candidate gene in a mouse oocyte system. We identified one locus on chromosome 3 that is significantly associated with meiotic aneuploidy risk. One candidate gene, *CCDC66,* encompassed by this locus, is involved in chromosome segregation during meiosis. Using mouse oocytes, we showed that CCDC66 regulates meiotic progression and chromosome segregation fidelity, especially in older mice. Our work extended the research utility of PGT-A ulc-WGS data by allowing robust association testing and improved the understanding of the genetic contribution to maternal meiotic aneuploidy risk. Importantly, we introduce a generalizable method that has potential to be leveraged for similar association studies that use ulc-WGS data.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7027F: Analysis of maternal-fetal genetic incompatibility in preeclampsia

Authors:

C. Cao[1], J. Hu[2], M. Maher[1], B. Keating[3], R. M. Burwick[4], S. Karumanchi[5], R. Saxena[6], K. Gray[7]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Dept. of Surgery, New York Univ., New York, NY, [4]Div. of Maternal Fetal Med., San Gabriel Valley Perinatal Med. Group, Pomona Valley Hosp. Med. Ctr., Pomona, CA, [5]Dept. of Med., Cedars-Sinai Med. Ctr., Santa Monica, CA, [6]Mass Gen. Hosp., Harvard Med. Sch., Boston, MA, [7]Univ. of Washington, Seattle, WA

Abstract:

Recent studies highlight the role of donor-recipient non-HLA incompatibility in transplant survival. Similarly, pregnancy success requires the tolerance of the maternal immune system to the semi-allogeneic fetal genome. Whether maternal and fetal genetic incompatibility contributes to pregnancy complications like preeclampsia is largely unknown. To examine the role of maternal-fetal genetic incompatibility in preeclampsia, we adopted a genome-wide approach for assessing donor-recipient genetic mismatch in transplant rejection. Whole exome sequencing data in maternal-fetal pairs of preeclampsia cases (n=44) were compared to whole genome sequencing data in healthy maternal-child pairs from 1000 Genomes Project (n=607). Proteins with trans-membrane or secreted annotations were selected from the Swiss-Prot database and mapped to 16,361 transcripts. After excluding MHC region and sex chromosomes, 28,998 coding SNPs in transcripts encoding trans-membrane or secreted proteins were selected for mismatch analysis and functionally classified into synonymous, missense, and loss-of-function SNPs using Ensembl Variant Effect Predictor. SNP mismatch was defined as the child having one different allele from the mother. For each maternal-child pair, mismatch counts were calculated separately for synonymous, missense, and loss-of-function SNPs. A total of 3.4% and 3.7% of SNPs were mismatched between mother-child pairs in preeclampsia and control groups, respectively. Mismatches in synonymous and missense SNPs were similar between cases and controls. However, the preeclampsia cases had increased number of mismatches in the loss-of-function SNPs compared to controls (n=40 ± 8 vs. n=16 ± 4). Replication in the TOPMed BCC-PREG cohort (928 case/ 1007 control pairs) is ongoing. If validated, these SNPs could be potential targets for further exploration in relation to immune tolerance and preeclampsia etiology. In conclusion, this preliminary analysis demonstrates the feasibility of maternal-fetal genetic mismatch analysis using methodology from graft survival prediction. Future analyses include incorporation of HLA mismatch analysis and functional investigation of mismatched SNPs. These investigations have the potential to provide significant insights into genetic interaction that affect pregnancy health.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7028F: Calculating Maternal Polygenic Risk Scores from Non Invasive Prenatal Testing Data

Authors:

S. Kim, V. Corey, m. chavez, l. qasim, c. deciu, a. henry, b. salome, t. sasha, M. Mehan, e. de Feo; Illumina, Inc., San Diego, CA

Abstract:

Polygenic Risk Scores (PRS) have enabled quantification of genetic risk for many common and complex traits. Early detection of genetic risk may improve patient care and health management. In addition, the ability to calculate PRS from pre-existing high throughput standard of care workflows could allow for quicker adoption and evidence generation. One potential avenue is leveraging prenatal screening by cell-free DNA (cfDNA) data to calculate maternal PRS as part of pregnancy management. Here we developed a method to estimate maternal PRS using low-coverage (0.25x) genome sequencing data from prenatal screening by cfDNA data intended to screen for fetal chromosomal aneuploidies.

A prospective study was conducted where 455 consented patients that performed prenatal screening by cfDNA as part of their standard of care were randomly selected. cfDNA and genomic DNA (gDNA) were isolated from the plasma and buffy coat of the blood drawn from pregnant women, respectively. cfDNA was processed by an adaptation of the VeriSeq™ NIPT Solution v2 while gDNA was processed by TruSeq™ PCR-Free library Prep. cfDNA was sequenced at ~0.25x coverage while gDNA was sequenced at ~15x coverage. All sequence data was analysed using DRAGEN™ v4.0 and imputed by DRAGEN Imputation v4.2.4. PRS was then calculated by utilizing imputed genotypes against >100 PRS panels from the Polygenic Score Catalogue where phenotypes include diabetes, hypertension, and breast cancer.

There was high correlation (average = ~0.9 across different PRS panels and panel sizes) between PRS from prenatal screening by cfDNA data and PRS from genome sequence data of the buffy coat. However, estimating maternal PRS from prenatal screening by cfDNA data is imperfect because the sequence data contains genetic information from both mom and fetus. This is illustrated by higher PRS correlations (mean ~=0.92) in samples with lower fetal fraction compared to samples with higher fetal fraction (mean correlation =~0.88).

This proof-of-concept study illustrates that maternal PRS can be calculated using low-coverage prenatal screening by cfDNA sequence data with high accuracy. Maternal PRS may be used to improve patient and pregnancy care as genetic data from prenatal screening by cfDNA can be applied to many other PRS panels that are designed to find genetic risk factors for specific

phenotypes. For example, there is evidence that genetic risk for hypertension or type II diabetes have been implicated to increase patient's risk to pre-eclampsia and gestational diabetes, respectively. Identifying patients as early as 10 weeks gestation that are potentially high risk for either of these conditions could benefit patient care.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7029F: Differential RNA editing in the placenta identifies genes relevant to preterm birth

Authors:

M. Huang, M. Horii, Q. Jiang, K. M. Fisch; UC San Diego, La Jolla, CA

Abstract:

Introduction: Preterm birth is a birth that occurs prior to 37 weeks of gestation and is a leading cause of neonatal mortality. The role that adenosine to inosine (A-to-I) RNA editing by *ADAR1* plays in preterm birth is an unexplored molecular mechanism. Understanding the relationship between A-to-I RNA editing and preterm birth may provide molecular biomarkers.
Methods: We performed A-to-I RNA editing analysis on 53 normotensive human placentas (N = 33 preterm, N = 20 term). Variants were called using the GATK Best Practices Workflow, annotated with VEP using hg38, and filtered for A-to-I sites in REDIPortal. We implemented a quality filter of variants with depth > 50, quality > 50, genotype quality > 20, quality by depth > 5, mapping quality > 40, rank sum test relative positioning > -3, rank sum test for mapping qualities > -10, and fisher strand sequencing bias < 200. We kept editing sites present in at least 5% samples, with a variant allele frequency < 0.9 and determined differential editing at each site using a log likelihood ratio test. Pathway enrichment analysis was performed using gProfiler. We performed clinical enrichment analysis with a Fisher's Exact Test in maftools. MicroRNA binding was predicted with TargetScan.
Results: We obtained 2563 RNA editing sites across 70 genes, with a median of 48 RNA edits per sample. Editing sites are enriched in pathways such as blood vessel development (p.adj = 0.0125) and vasculature development (p.adj = 0.0185). Pairwise clinical enrichment of RNA editing sites finds 23 genes with a FDR < 0.3, while differentially editing analysis finds 25 significantly edited sites (p.adj < 0.05) within 21 genes. Interestingly, differential editing sites are enriched for *VEGF* receptor activity (p.adj = 7.36E-03), which is a potential marker for preterm birth. Notably, *CTSB* is clinically enriched for preterm birth (fdr = 0.0381) and is edited in 2.0% of preterm placentas, but not edited in term placentas (p.adj = 0.0407). *CTSB* expression was found to be elevated in the urine of preterm babies. *FLT1* is edited in 5.9% of term and 1.6% of

preterm placentas (p.adj = 0.00409). Expression of *FLT1* in maternal serum was found to be reduced during preterm birth. hsa-miR-429 is predicted to bind *FLT1*, and is known to regulate expression of *VEGF*.
Conclusion: This study has found several sites in key genes that are differentially edited between preterm and term placentas. Future studies will incorporate clinical characteristics to subclassify the disease. Investigating the role of these RNA editing sites and the microRNAs that bind them using in vitro models could help elucidate the mechanism behind preterm birth and are underway in our lab.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7030F: EHR phenotypes associated with genetic testing and diagnosisin the NICU ⭐

Authors:

B. Schuler, J. Sucre, D. Ruderfer, L. Bastarache; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Genetic disorders contribute to morbidity and mortality and have a high prevalence in the neonatal intensive care unit (NICU). Understanding genetic testing practices and phenotypes predictive of genetic diagnoses can improve testing utilization.
The NICU database (NICUdb) contains 418 manually curated EHR features (surgical codes, congenital anomalies, medications, maternal data, etc.) of 22,837 NICU patients seen between 2005-21. We integrated all germline genetic testing data extracted from the EHR into the NICUdb, including the timing and types of tests and their results. We performed Bonferroni corrected logistic regression to identify EHR features (n=269) associated with receiving a genetic test and a diagnosis (controlled for estimated gestational age, birth weight, and sex).
135 EHR features were significantly associated with receiving a genetic test and 93 were significantly associated with receiving a genetic diagnosis. 82 EHR features were significantly associated with both receiving a test and a diagnosis including heart transplantation (test odds ratio [OR]=55.6, $p=9.6e^{-6}$; diagnosis OR=4.9, p=0.01) and complete atrioventricular canal defect (test OR=14.8, $p=3.5e^{-37}$; diagnosis OR=20.9, $p=7.2e^{-72}$). EHR features only associated with receiving a genetic test and not a diagnosis (n=53) included congenital diaphragmatic hernia (test OR=3.9, $p=5.3e^{-25}$ vs. diagnosis OR=0.81, p=1.0) and seizures (test OR=2.0, $p=7.7e^{-15}$ vs. diagnosis OR=1.3, p=1.0). Nine EHR features were associated with a diagnosis but not testing including: inborn errors in metabolism (testing OR=6.3, p=0.08 vs. diagnosis OR=16.2, $p=2.5e^{-6}$) and nephrectomy (testing OR=5.4, p=1.0 vs. diagnosis OR=27.6, $p=6.3e^{-4}$).

Features associated with testing likely represent phenotypes that increase pretest suspicion of a diagnosis. Features associated with a genetic diagnosis in the absence of testing may reflect alternative diagnostic approaches (e.g. newborn screening). Features associated with a test but not a diagnosis likely reflect clinical presentations in which genetic testing may have lower utility. Understanding genetic testing practices and phenotypes predictive of genetic diagnoses can inform test utilization and provide insight into pretest suspicion of an underlying genetic disorder.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7031F: Exome sequencing uncovers recurrent and novel genetic causes of spermatogenic failure in a large cohort of infertile men

Authors:

A. Riera-Escamilla[1,2], H. Castillo-Madeen[1], A. Akbari[1,2], C. Peaslee[1,2], K. Vigh-Conrad[1], GEMINI consortium, K. I. Aston[3], D. F. Conrad[1,2,3]; [1]Div. of Genetics, Oregon Natl. Primate Res. Ctr., Oregon Hlth.and Sci. Univ., Beaverton, OR, [2]Ctr. for Embryonic Cell & Gene Therapy, Oregon Hlth.and Sci. Univ., Beaverton, OR, [3]Andrology and IVF Lab., Dept. of Surgery (Urology), Univ. of Utah Sch. of Med., Salt Lake City, UT

Abstract:

Infertility affects 15% of couples, with male factors accounting for half of these cases. Despite the known impact of genetics on male infertility, clinical diagnostics typically involve only karyotype analysis and Y-Chromosome deletions. These methods leave the genetic cause unidentified in 40% of cases. To provide novel insights into male infertility etiology, we performed exome sequencing on: i) 1,097 infertile men with azoospermia; ii) 377 men with reduced sperm counts, ranging from extreme oligozoospermia (EO; <2 million total sperm count) to severe (SO; 2-10 million) and moderate oligozoospermia (MO; 10-29 million). The control group comprised 633 fertile men with normozoospermia. We prioritized likely damaging variants compatible with recessive Mendelian disease models using an analysis pipeline based on population sampling probability. Genes with loss of function mutations found in controls, identical genotypes, variants affecting the same protein domain in both infertile and fertile men and variants classified as benign or likely benign according to ACMG were excluded as potentially causative. The resulting variants were crossed against two lists: 444 autosomal recessive/sex-linked genes previously reported to be mutated in isolated spermatogenic failure and 1,406 genes linked to reproductive defects in knockout (KO) mice but not previously linked to human male infertility. The analysis of the diagnostic yield based on ACMG-Clingen guidelines

is ongoing. We found 281 infertile men carrying recessive damaging variants in 210 genes previously linked to human spermatogenic failure, representing a detection rate of 19.06%, whereas only 35 controls were carriers (5.53%) (p-value = 2.6e-15). The detection rate increased to 27.41% (404 cases) when including patients without variants in known human infertility genes but carrying variants in genes linked to reproductive defects in KO mice. From these analyses, we discovered 98 genes mutated in 123 patients as potential novel targets for human infertility. Notably, in 87/98 genes, no variants were found in controls. The detection rate was higher in patients with the most severe forms of male infertility, nearing 28% in cases with azoospermia or EO, while it was 17-23% in cases with SO and MO. Moreover, we observed a higher detection rate in patients with testis histology of maturation arrest (41.37%) compared to those with Sertoli cell-only or hypospermatogenesis (25.8% and 27.08%, respectively, p-value = 0.008). In summary, our study shows how incorporating extensive cohorts and knowledge of mouse phenotypes adds value to the understanding of the genetic landscape underlying male infertility.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7032F: Expansion of the phenotype spectrum due to a splice variant in *WDR44*-gene - severe prenatal phenotype in multiple male fetuses in a Finnish family

Authors:

E. Salminen[1,2], E. Niemelä[3], M. Ataei[4], G. Saldo Rubio[3], P. Bauer[4], A. Bertoli-Avella[4], A-K. Anttonen[1,3]; [1]Dept. of Clinical Genetics, HUS Diagnostic Ctr., Helsinki Univ. Hosp., Helsinki, Finland, [2]Dept. of Med. and Clinical Genetics, Univ. of Helsinki, Helsinki, Finland, [3]Lab. of Genetics, HUS Diagnostic Ctr., Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, [4]Centogene GmbH, Rostock, Germany

Abstract:

Background: Recently, 11 male patients from nine families were reported who had missense variants (n=8) and a nonsense variant (n=1) in *WDR44*-gene (Accogli A et al. 2024). Reported patient phenotypic spectrum included developmental delay/ intellectual disability, hypotonia, distinct craniofacial features and variable brain, renal, cardiac and musculoskeletal abnormalities. Accogli et al. demonstrated phenotype modelling in zebrafish and were able to show that *WDR44* variants associated with more severe disease had impaired ciliogenesis initiation and ciliary signaling. *WDR44*-gene is intolerant to loss-of-function variants in gnomAD v2.1 database, pLI 1 and LOEUF 0.195.

Objective: To study the genetic basis for recurrent severe prenatal phenotype in male fetuses in a Finnish family.
Methods: During the years, genetic investigations (karyotype, chromosomal microarray, exome sequencing and genome sequencing) were performed for a Finnish family in diagnostic genetic laboratories. RNA-analysis was performed to study the effect of the identified splice variant in fetal fibroblasts.
Results: The couple had altogether nine spontaneous pregnancies, with multiple early miscarriages. In three pregnancies, abnormal ultrasound findings were noted in 1st trimester including fetal hydrops and later oligo/anhydramnion. Evaluation of the fetuses after later miscarriages (H14, H16) and one termination of the pregnancy (H17) noted fetal structural abnormalities including bifid thumb in 2/3 fetuses. Karyotype and chromosomal microarray analyses were normal.
Exome sequencing for two male fetuses and parents and genome sequencing for two male fetuses and parents did not identify any causal variants in known genes. Research analysis of genome data identified a maternally inherited hemizygous splice acceptor variant c.1975-1G>C in *WDR44* (NM_019045.4) gene in two affected male fetuses and the same variant was observed as hemizygous in the exome sequencing data of the 3rd male fetus. Sanger sequencing of the 4th male fetus with DNA available demonstrated this variant as hemizygous. Targeted Sanger sequencing of the maternal grandparents did not identify the variant suggesting a *de novo* variant in the mother. Splice prediction tools predicted the consensus splice site to be abolished with the creation of a new cryptic acceptor splice site within exon 15, and a predicted 8 bp deletion. PCR amplification of cDNA from fetal skin fibroblasts and Sanger sequencing identified an 8 bp deletion as predicted leading to a frameshift and premature termination codon.
Our findings expand the phenotype spectrum due to *WDR44*-variants in humans to a severe prenatal phenotype.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7033F: Functional assessment of *PASD1* gene variants in male infertility

Authors:

C. Pombar[1], G. Atkins[1], R. Hvasta[1], W. Walker[1], N. Ullah[2], A. Malcher[3], C. A. Castro[1], K. Hwang[4], K. Orwig[1], P. N. Schlegel[5], M. Olszewska[3], M. Kurpisz[3], S. Sherman[6], M. Khan[2], A. Yatsenko[1]; [1]Magee Womens Res. Inst., Pittsburgh, PA, [2]Khyber Med. Univ., Peshawar,

Pakistan, [3]Polish Academy of Sci., Poznan, Poland, [4]Magee Women's Hosp., Pittsburgh, PA, [5]Weill Cornell Med., New York, NY, [6]Infertility Ctr. of St. Louis, St. Louis, MO

Abstract:

Infertility is growing health issue, which affects roughly 17.4% of people worldwide. One of the most severe forms of infertility is azoospermia, the complete lack of sperm in the ejaculate. In cases of non-obstructive azoospermia approximately 40% have no known cause and are believed to have an underlying genetic cause. One possible candidate gene is *PASD1*, an X-linked gene known to interact with the circadian rhythm complex, CLOCK-BMAL1, that is exclusively expressed in the testis. We examined our cohort for promising hemizygous variants in *PASD1*. DNA was isolated from whole blood and underwent whole exome sequencing. Variants were analyzed via Fabric Genomics software. Two models were developed for characterization of variant effects in mice and human cells. Initial mouse knock-out model was generated using CRISPR/Cas9 on embryos, which were implanted in female mice. Sanger sequencing to confirm genotype in mice was used to validate breed testing pairs. Total RNA was converted to cDNA, and RT-qPCR was used to confirm *PASD1* RNA expression in human cell line of interest. Immunocytochemistry was used to confirm protein expression of PASD1 in the cells.
From our cohort, we identified one significant loss of function variant, a novel frameshift, p.Gly468SerfsTer105. We first attempted to generate a knock-out model in mice to determine if the gene has an effect on male fertility in mice. Hemizygous mice with frameshift variants were fertile, and testis histology appeared normal. After determining loss of the gene had no impact on mouse fertility, we switched our focus to in vitro assays in a human cell model. SW-480 is a cancer cell line that endogenously expresses *PASD1*. Using primers designed to identify the mRNA of *PASD1*, we confirmed significant expression of *PASD1* in the SW-480 cells. Anti-PASD1 antibodies were used to confirm protein expression in cells. Initial attempts to generate a knock-out mouse model for *Pasd1* were unsuccessful, suggesting that the function of the gene is not critical to male fertility in mice. In parallel, phylogenomic data suggests that mouse and rat sequences for the gene strongly diverge from other mammals. Given this divergence, human cell models that naturally express the gene are the most optimal models for evaluating its function. Currently, we are developing a knock-in cell line using CRISPR/Cas9 and identifying localization of key proteins in the CLOCK-BMAL1 pathway. Additionally, we are planning to evaluate differential RNA expression of certain genes in the mutant cells. Following characterization of this loss-of-function model, we will attempt to knock-in additional *PASD1* variants we have identified to evaluate possible impact.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7034F: Genome-wide identification of epigenomic elements in human embryonic palatal cells

Authors:

T. Carter, L. Moat, S. Yerukala Sathipati; Marshfield Clinic, Marshfield, WI

Abstract:

Cleft palate, a common craniofacial birth defect with a prevalence of 1 in every 2,000 births, arises from abnormal development of the secondary palate. Population-based and family studies suggest that the etiology of cleft palate has a strong genetic component. However, knowledge of the molecular interaction pathways that contribute to human palate development and are disrupted in cleft palate is incomplete. To identify genomic enhancer regions potentially involved in regulating gene expression in the developing palate, genome-wide profiling of the binding of six histone proteins (H3K4me1, H3K4me3, K3K9me3, H3K27ac, H3K36me3, and H3K36me3) and of IgG (used as a control) to chromatin in a human embryonic palatal mesenchyme (HEPM) cell line was performed using CUT&RUN assays. Profiles were analyzed using the ChromHMM algorithm, which integrates global histone binding data to segment the genome into chromatin states that represent functional regulatory elements. Eighteen chromatin states that could be broadly categorized into enhancers (n = 207,849), transcription start sites and flanking sites (n = 46,689), transcription regions (n = 43,396), and repressed regions (n = 220,297) were identified. To identify cell-specific, active enhancers, the subset of enhancers that were annotated as active (n = 66,213) were compared for overlap with 827,518 enhancers that were previously reported in a single genome annotation based on the combined analysis of data from over 100 cell and tissue types and are therefore active in a broad range of cell types. This comparison identified 17,599 active enhancers that were specific to HEPM cells. The identification of HEPM cell-specific enhancers facilitates further investigation to identify the binding sites of transcription factors that regulate gene expression during palate development and the genes targeted by the enhancers and to uncover the mechanisms through which disruption of enhancer regions by genetic variants contribute to cleft palate.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7035F: Genome-Wide Meta-Analysis of Single Variant Associations in Uterine Fibroid Samples

Authors:

J. Jaworski[1,2], A. Williams[3], J. Hellwege[4,5], J. Kim[4], E. Jasper[4,6], J. Denny[7,3], T. Edwards[1,2], D. Velez Edwards[1,6,8]; [1]Vanderbilt Univ. Med. Ctr., Vanderbilt Univ., Nashville, TN, [2]Div. of Epidemiology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Natl. Human Genomic Res. Inst., Bethesda, MD, [4]Vanderbilt Genetics Inst., Vanderbilt Univ., Nashville, TN, [5]Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Div. of Quantitative Sci., Dept. of Obstetrics and Gynecology, Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]All of Us Res. Program, NIH, Bethesda, MD, [8]Dept. of BioMed. Informatics, Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Uterine fibroids (UF) are benign smooth muscle growths affecting up to 75% of women by menopause. African American (AA) women have higher prevalence, earlier onset, and larger and more numerous fibroids than European American (EA) women. Genome-wide association studies have found numerous genetic loci associated with fibroids risk, however most of those studies were of European ancestry. Sequencing studies of fibroids to date have focused on somatic sequence changes. We performed whole genomic sequencing (WGS) on AA and EA samples to determine the impact of rare variants on fibroid risk across populations. Samples from a de-identified electronic health record system linked to a DNA biorepository (BioVU) were selected using a previously described algorithm. A total of 548 AA (271 cases/ 277 controls) and 659 EA (371 cases/ 288 controls) samples were sequenced using the Illumina Dragen Germline application (v.3.10.4) on the Illumina Basespace sequence hub and joint called separately using the Dragen PopGen application (v.3.10.4). Sample and variant quality control was performed in PLINK2. The Ensembl Variant Effect Predictor was used for all annotation. Single variant and gene-based testing were performed using RVTESTS. Models were adjusted for age and 10 principal components, while limited to variants with a minimum minor allele count (MAC) of 20. Meta-analysis of the single variant association results was performed with the *All of Us* Research Program (*All Of Us*) UF WGS dataset (version 7) using METAL. The ancestry specific and cross-ancestry meta-analysis was done with the *All Of Us* AA (4588 cases/5589 controls) and EA (15,132 cases/10,519 controls) summary statistics from their WGS project. For the BioVU AA samples, the single variant association analysis showed nominally significant loci in 4p12 (p-value $1.5 \times 10^{-8)}$ and 20q11.22 (p-value 1.4×10^{-7}). In the African ancestry meta-analysis with the *All Of Us* dataset, the top locus was at 11p13 in the *WT1* gene (p-value 1.0×10^{-7}). The loci at 4p12 did not replicate. The EA meta-analysis also showed a significant peak at the same 11p13 locus (p-value 8.9×10^{-12}) as the AA ancestry. There was an additional locus at 6q25.2 in the gene *SYNE1* (p-value 6.9×10^{-10}). The cross-ancestry meta-analysis confirmed the 11p13 and 6q25.2 peaks with significant p values of 8.9×10^{-17} and 1.3×10^{-13} respectively. Although the cross-ancestry meta-analysis did not show any novel associations, it did replicate the findings previously reported of *WT1* and *SYNE1* association with increased risk of UF in both AA and EA

ancestry. Future directions include investigating the role of rare variants in gene-based meta-analysis and non-coding regions.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7036F: High frequency of *de novo* variants and parental DNA damage response gene variants in couples with recurrent pregnancy loss

Authors:

M. Aminbeidokhti[1], H. Cakmak[1], E. Jaswa[1], A. Nagasuri[1], M. Baldwin[2], A. Edelman[2], E. Pollard[3], M. Snyder[3], M. Sirota[1], C. Tise[3], J. Bernstein[3], M. Stephenson[4], R. B. Lathi[3], S. Yatsenko[3], A. Rajkovic[1]; [1]Univ. of California San Francisco, San Francisco, CA, [2]Oregon Hlth.& Sci. Univ., Portland, OR, [3]Stanford Univ., Stanford, CA, [4]Univ. of Illinois at Chicago, Chicago, IL

Abstract:

Recurrent pregnancy loss (RPL), defined as two or more miscarriages, impacts approximately 5% of couples. We enrolled 68 RPL couples and performed genome sequencing on euploid DNA samples obtained from pregnancy losses and on blood samples from both parents. We observed a high rate of de novo variants in POC samples. We hypothesized that de novo variants are significant contributors to RPL that may be due to advanced parental age or a hidden germline mosaicism. We also studied the possibility that a subset of RPL patients carry pathogenic variants in genes involved in a DNA damage response, that predisposes to genomic instability during gametogenesis leading to a high frequency of de novo variants and fetal loss. For detailed analysis we selected two sets of variants: pathogenic/likely pathogenic associated with known autosomal dominant and X-linked conditions and variants of unclear clinical significance present in a heterozygous state in POC samples, but never reported in a database of healthy individuals (GnomADv4.1.0). In addition, we performed deep-coverage exome sequencing and digital droplet PCR (ddPCR) on parental DNA extracted from urine, blood, and semen to rule out parental germline mosaicism.
We identified 35 de novo variants in 41.1% (28/68) POC samples, including 7.35% (5/68) and 1.47% (1/68) of POC samples with two and three de novo variants, respectively. There was no significant difference between parental ages of the participants with de novo variants and those who did not have a de novo variant in their POC sample. Parental germline mosaicism testing in de novo cases, yielded positive results in 2.9% (2/68) of families. Parental samples were examined for variants in 234 genes involved in DNA repair and/or damage response (DDR). In RPL cohort, the proportion of parents with pathogenic/likely pathogenic variants was much

higher (15.8%) than the observed 5.6% rate in the general population. We hypothesize that heterozygous variants in DDR parental genes may be a significant contributing factor of RPL and lethality.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7037F: Highly variable expressivity of a CNV deletion involving *TBX4* in three siblings with lethal lung developmental disorder and their mildly affected mother and grandfather.

Authors:

P. Stankiewicz[1], Q. Ain[1], T. Gambin[2], D. Gail[3], N. Cortes-Santiago[4], S. Raskin[5], P. Szafranski[1]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Inst. of Computer Sci., Warsaw Univ. of Technology, Warsaw, Poland, [3]Dept. of Lab. Med. and Pathology, Univ. of Washington Sch. of Med., Seattle, WA, [4]Dept. of Pathology & Immunology, Baylor Coll. of Med., Houston, TX, [5]Laboratorio Genetika, Curitiba, Brazil

Abstract:

In addition to pulmonary arterial hypertension (PAH) and ischiocoxopodopatellar syndrome (small patella) with or without PAH, single nucleotide variants (SNVs) and copy-number variant (CNV) deletions involving the dosage sensitive (pLI score = 1.0) T-box transcription factor *TBX4* gene have been associated with lethal lung developmental disorders (LLDDs) in neonates, including acinar dysplasia (AcDys), congenital alveolar dysplasia (CAD), and other unspecified primary pulmonary hypoplasias. In contrast to missense SNVs, CNV deletions encompassing *TBX4* have been found exclusively as *de novo* events. To address variable expressivity of *TBX4* SNVs and CNV deletions, we previously proposed a complex compound inheritance model with hypomorphic regulatory SNVs in the *TBX4* lung-specific super-enhancer acting in *trans* (PMID: 30639323). Here, we present a multigenerational family with three neonate siblings with clinical features of LLDD who died within 35-66 days after birth. Histopathological evaluation in the lung biopsy of one neonate showed pronounced alveolar simplification with mild hypertensive changes of the pulmonary arteries consistent with CAD/pulmonary hypoplasia. Intra-alvaeolar accumulation of proteinaceous debris and superimposed bronchopneumonia were also present. Whole genome sequencing identified a heterozygous ~ 108 kb CNV deletion encompassing *TBX4*. This deletion was also detected in the other two siblings and their mother who has a history of a cyanotic episode at age 9 months diagnosed as pneumonia, three other pneumonia episodes at the age 5, 10, and 21 years, and a thick upper airway secretion on a daily basis. Unexpectedly, we also found this non-mosaic

deletion in the mother's father who only had a surgically corrected genu valgum. Segregation studies revealed that the father transmitted different *TBX4* alleles to his three children, suggesting transmission of hypomorphic variant(s) on different chromosome(s). In search of these variants, we computationally analyzed the candidate genomic regions involved in human lung development in the affected and unaffected family members. Our data further demonstrate complex compound inheritance of LLDDs.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7038F: Human data, machine learning, and mouse models provide evidence that *WNT4* plays a critical role in heart and diaphragm development. ⭐

Authors:

A. Hernandez-Garcia; Baylor Coll. of Med., Houston, TX

Abstract:

WNT4 is a secreted protein that plays a pivotal role in the regulation of cell fate and embryogenesis. Autosomal dominant missense variants in *WNT4* have been shown to cause Mullerian aplasia, hyperandrogenism, and occasionally, unilateral renal agenesis in multiple individuals. These findings align with observations from female *Wnt4* null mice, which exhibit masculinized genitalia, absence of the Mullerian duct, ectopically activated testosterone biosynthesis, and atretic kidneys. In contrast, only a single consanguineous kindred has been identified in which autosomal recessive variants in *WNT*4 have been implicated in the development of SERKAL syndrome, a distinct genetic disorder characterized by 46,XX sex reversal and dysgenesis of the kidneys, adrenals, and lungs. Within this kindred, one of the affected fetuses had a ventricular septal defect and pulmonary artery stenosis, while another had a diaphragmatic hernia. These structural birth defects have not been previously described in *Wnt4* null mice. Furthermore, using a machine learning algorithm we demonstrated that *WNT4* was highly similar to genes known to play a role in endocardial-to-mesenchymal transition (EndMT), a key process in the formation of the endocardial cushion, and genes known to cause congenital diaphragmatic hernia. We subsequently demonstrated in histological analyses that a subset of *Wnt4* null embryos have perimembranous and muscular ventricular septal defects. Using a combination of histological analyses, diaphragmatic microdissection, and micro-CT analyses, we also demonstrated that *Wnt4* null embryos and newborn mice have abnormal thinning of the diaphragmatic musculature and retrosternal anterior and posterior sac hernias. These findings are similar to the cardiac and diaphragmatic defects we have previously

identified in SOX7 deficient mice. Hence, our findings not only suggest that WNT4 deficiency can cause cardiac defects and diaphragmatic hernia in mice and humans with SERKAL syndrome, but also provide additional evidence in support for our proposed model of mammalian septal development in which the expression of WNT4 is positively regulated by SOX7 in the endocardium and induces EndMT by acting as a paracrine factor to upregulate BMP2 expression in the myocardium. Moreover, these results also suggest that a similar molecular mechanism may play a key role in diaphragm development.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7039F: Identification of fetal nucleated red blood cells suitable for cell-based non-invasive prenatal genetic testing using single-cell multi-omics sequencing.

Authors:

N. Ito[1], Y. Okazaki[1], K. Taniguchi[1,2], T. Fujii[1], H. Ogata-Kawata[1], K. Hata[1,2], H. Sago[1], K. Nakabayashi[1]; [1]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [2]Gunma Univ. Graduate Sch. of Med., Gunma, Japan

Abstract:

Non-invasive prenatal genetic testing (NIPT) using cell-free DNA is the standard method for current prenatal screening tests. Since fetal cells in maternal blood represent a pure source of fetal genomic DNA, they are expected to be ideal for NIPT if successfully isolated. We have previously isolated fetal nucleated red blood cells (fNRBCs) from maternal blood and conducted whole-genome amplification and sequencing for the isolated cells (Ito N. *et al.* 2021). Our results at that time suggested low and partial recoveries of genomic DNA likely due to enucleation and/or apoptosis in most isolated cells. We hypothesized that cells in the primitive stage of NRBC maturation maintain intact genomic DNA suitable for cell-based NIPT. To identify such primitive cells among NRBC candidate cell population based on the expression patterns of differentiation stage-specific genes (Zhao, Y. *et al.* 2019), we conducted genome and transcriptome sequencing (G&T-seq). Using umbilical cord blood as a model, we isolated 165 single NRBC candidate cells from four samples and also 12 single lymphocytes from one sample, and obtained their transcriptomic and genomic data. In the multi-dimensional scaling (MDS) analysis of transcriptomic data, as expected, we were able to distinguish lymphocytes and NRBC candidates. Among 165 NRBC candidates, five cells formed a cluster distinct from that of the other 160 cells, and the cluster was located between the lymphocyte cluster and the cluster of 160 cells along the dimension-1 axis of MDS. Five cells were confirmed to be primitive NRBCs

based on the expression levels of primitive marker genes such as *GYPC*, *CD47* and *CD36*, and the majority of the other 160 cells were classified as mature cells based on their gene expression patterns. The numbers of expressed genes were found to be high in 12 lymphocytes (average 4774) and low in mature RBC cells (average 259). The numbers of expressed genes in the five primitive NRBCs were intermediate between them, ranging from 875 to 2073. In the genomic data analysis, two out of five primitive cells showed high library yields and mapping rates comparable with those of lymphocyte cells, suggesting the intactness of their nuclear genome. The data obtained here supported our above-mentioned hypothesis and highlight the importance of identifying primitive cells among NRBC candidates to develop fetal cell-based NIPT as a practical method. We plan to investigate the relationship between the progression of RBC maturation and the intactness of genomic DNA among NRBC candidates isolated from maternal blood at various gestational weeks.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7040F: Identifying temporal postnatal modifiers of Hirschsprung Disease using transcriptomic, chromatin accessibility and cellular composition changes in the mouse gut

Authors:

H. Chubaryov[1], A. Chakravarti[2]; [1]NYU, New York, NY, [2]NYU Grossman Sch. of Med., New York, NY

Abstract:

Introduction: In embryonic gut development, enteric neural crest cell (ENCC) migration, differentiation, and proliferation leads to the establishment of the Enteric Nervous System (ENS). Genetic defects that perturb this developmental program lead to Hirschsprung Disease (HSCR), arising from the partial or total absence of the ENS. Despite successful surgical repair, 30-50% of HSCR patients continue with persistent gastrointestinal symptoms. We hypothesize that postnatal genetic factors modify HSCR and its progression**. Methods:** To understand temporal trends, wildtype mouse colon tissue was collected at E12.5, E14.5, P0, P07, P21, and adult for bulk RNA-seq and ATAC-seq, as well as P0 and P21 for snRNA-seq. Additionally, we collected P21 bulk RNA-seq data for five wildtype and mutant genotypes of *Ret/Ednrb*, the two major genes for HSCR. **Results:** Our analyses identified 5 significant and novel findings. First, we discovered large temporal changes throughout development with 81% (RNA-seq) and 85% (ATAC-seq) of PCA variance attributable to time, largely between embryonic and postnatal timepoints. Second, chromatin accessibility data reveals a bimodality of peak behavior over time, either ubiquitously open or timepoint specific. Third, we identified 9,959 genes and

24,158 enhancers temporally variable, with several clusters peaking in postnatal development, indicating tremendous restructuring of the gut genetic program after birth (P0). Fourth, the snRNA-seq data revealed that the changes at P0 and P21 were from fibroblasts, which increase across development, and neurons and glia of the ENS, which decrease. Finally, by hierarchically clustering the P21 bulk RNA-seq across *Ret/Ednrb* genotypes we identified the postnatal cell type(s) affected by HSCR. One cluster was enriched in neurons and glial cells and associated with neurotransmitter transport, while a second was enriched in neurons and glial as well as in the fibroblasts and associated with ECM matrix organization. This result indicates the potential involvement of non-cell autonomous effects later in the disease. **Conclusion:** We have identified the genes, enhancers, and cell types that are involved in gut development and can now identify which postnatal developmental changes affect HSCR disease progression.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7041F: Incidence and sex ratio in children born with tetralogy of Fallot in the Czech Republic

Authors:

A. Sipek Sr[1,2,3,4], A. Sipek Jr[1,5,2], J. Klaschka[6], M. Maly[6,7]; [1]Dept. of Med. Genetics, Thomayer Univ. Hosp., Prague, Czech Republic, [2]Inst. of Med. Genetics, 3rd Faculty of Med., Charles Univ., Prague, Czech Republic, [3]Dept. of Med. Genetics, Pronatal Sanatorium, Prague, Czech Republic, [4]GENNET, Prague, Czech Republic, [5]Inst. of Med. Biology and Genetics, 1st Faculty of Med., Charles Univ. and Genaral Univ. Hosp., Prague, Czech Republic, [6]Inst. of Computer Sci. of the Czech Academy of Sci., Prague, Czech Republic, [7]Natl. Inst. of Publ. Hlth., Prague, Czech Republic

Abstract:

Objectives: Tetralogy of Fallot is a severe congenital heart defect that occurs as an isolated anomaly or part of complex genetic syndromes. We performed a retrospective epidemiological analysis of the frequency and sex ratio of children born with tetralogy of Fallot in the Czech Republic in the period 1994-2020.
Methodology: The work uses data from the National Registry of Congenital Anomalies which is a part of the National Registry of Reproductive Health run by the Institute of Health Information and Statistics of the Czech Republic. This registration is population-wide, compulsory by the national law, and does not require informed consent. All cases of diagnosis Q213 (in ICD-10 classification) registered in births were included. The analysis focused on the changes in the annual incidences and the sex ratio. The statistical analysis was performed by the Fisher's exact

test.
Results: In the period 1994-2020, more than 2.8 million children were born in the Czech Republic. Of this total number, there were more than 65,000 children born with a congenital heart defect. In the case of tetralogy of Fallot, there were a total of 849 cases, of which 476 diagnoses were recorded in boys and 373 diagnoses in girls. In relative numbers, it was 2.78 per 10,000 live births in total, 3.75 for boys, and 1.77 per 10,000 live births for girls. In terms of percentage representation, there were 69.1% affected boys and 30.9 % affected girls.
Conclusions: Congenital heart defects are among the most common congenital defects in childhood. In our work, we analyzed the incidence of the tetralogy of Fallot in the Czech Republic. The incidence of this defect has been increasing since 1994, the highest values are in the period 2000-2006. Then the frequency of this defect in births decreases. Furthermore, we found more affected boys compared to girls with this defect.
Supported by MH CZ - DRO (Thomayer University Hospital - TUH, 00064190)

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7042F: Multi-omics framework discovers novel biomarker and genetic signatures for women's reproductive health

Authors:

S. Stankovic, M. Garg, A. Tsakiroglou, L. Middleton, S. Petrovski, D. Vitsios; AstraZeneca, Cambridge, United Kingdom

Abstract:

Underfunding and lack of research in women’s health has contributed to poor understanding of underlying biology and lack of reliable non-invasive biomarkers. This often results in misdiagnosis, delayed diagnosis, and/or incorrect treatments for conditions like endometriosis, polycystic ovary syndrome (PCOS) and infertility, which have a wide range of symptoms and causes. Large-scale population studies, such as UK Biobank (UKB), contain rich phenotypic data which can help us uncover the biomarker signatures of such underdiagnosed conditions and predict novel undiagnosed cases. We have trained MILTON, an ensemble machine learning (ML) tool, to infer disease-specific biomarker signatures and identify potential missing cases based on shared biomarker profiles with already diagnosed cases. We utilized 67 baseline and 49 reproductive health-related biomarkers available across UKB (n=500K), as well as UKB metabolomics (n=300K) and proteomics (n=50K) data to train MILTON. Reproductive health-related features, such as, age at menopause, copper levels and pelvic pain, considerably boosted performance (AUC =0.71 for endometriosis compared to 0.59, when using only 67

baseline biomarkers), followed by proteomics data. Disease prevalence in UKB increased by ~2-fold, after inclusion of MILTON-predicted cases for the studied disease cohorts. Features, such as, age, BMI, waist circumference and HDL cholesterol were consistently important for both PCOS and endometriosis. Notably, age at first live birth, pelvic pain, and length of menstrual cycle were primarily important for all endometriosis subtypes, but not PCOS, while the number of spontaneous miscarriages, apolipoprotein A and copper were important for infertility and PCOS, but not endometriosis.
Next, we ran phenome-wide association analysis on MILTON augmented case-control cohorts and identified 12 significant ($p<5x10-8$) gene-disease associations across the three phenotypes, not previously identified in baseline PheWAS, 9 of which are novel. We found that novel gene signals are also enriched for several phenotypes of metabolic, cardiovascular, immune and neoplastic origin. Finally, we further triaged all nominally significant ($p<0.05$) baseline PheWAS hits using a comprehensive ML gene-prioritisation tool (Mantis-ML 2.0) to highlight genes with strong support from both methods. Top signals were enriched for pathways involved in hormone synthesis, immune response and mitochondrial function. Our approach demonstrates the value of leveraging multi-omics data with ML to increase our understanding of complex reproductive conditions.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7043F: NGS-based genetic analysis of embryos for chromosomal abnormalities in Nepal

Authors:

S. Khanal, S. Thapa, G. Joshi, N. Thakur, A. Kunwar; Kathmandu Ctr. for Genomics and Res. Lab., Lalitpur, Nepal

Abstract:

Chromosomal abnormalities in embryos pose significant risks to reproductive and developmental outcomes. Preimplantation Genetic Screening (PGS) offers a means to detect such abnormalities before embryo implantation, thereby enhancing the selection of viable embryos. This study aimed to assess the prevalence of chromosomal abnormalities and their impact on embryo viability through genetic analysis of embryos from patients in Nepal.Genetic analysis using Next-Generation Sequencing (NGS) was conducted on embryos from 108 patients, each assigned unique Patient IDs and Embryo IDs. Chromosomal abnormalities were categorized based on impacted chromosomes, distinguishing euploid, aneuploid, and mosaic embryos.Analysis of 446 embryos revealed 114 euploid, 122 aneuploid, and 178 mosaic

embryos. Chromosomes 16, 22, 20, 19, and 21 exhibited the highest combined gains and losses among the embryos, indicating significant impact.The prevalence of aneuploidy and mosaicism underscores the importance of genetic screening in embryo selection. The identification of chromosomal abnormalities, particularly on chromosomes 16, 22, 20, 19, and 21, highlights potential areas of concern for embryo viability and developmental outcomes. This study provides critical insights into the genetic landscape of embryos in Nepal, emphasizing the significance of PGS in selecting viable embryos for assisted reproductive technologies. These findings inform clinical practices aimed at improving pregnancy success rates and reducing the risk of genetic disorders in offspring.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7044F: Pleiotropic effect of *de novo* variants in structural congenital anomalies and neurodevelopment disorder

Authors:

W. Gao[1], J. Hagen[2], L. Qiao[1], G. Zhong[3], C. Shu[1], Y. Shen[3], W. Chung[4]; [1]Columbia Univ. Irving Med. Ctr., New York, NY, [2]Columabia Univ. Irving Med. Ctr., New York, NY, [3]Columbia Univ., New York, NY, [4]Columbia Univ., Boston, MA

Abstract:

Structural congenital anomalies like congenital diaphragmatic hernia (CDH), orofacial cleft and disorders of sexual development have heterogenous genetic causes. Previous studies suggest that hundreds of genes can contribute to risk of these conditions through *de novo* variants. Additionally, many genes involved in core developmental pathways may have pleiotropic effects and contribute to the risk of multiple congenital anomalies and neurodevelopmental disorders. We aimed to study pleiotropic effects of *de novo* variants across these conditions. We obtained genome data across congenital anomalies from the NIH Gabriella Miller Kids First (GMKF) Pediatric Research Program, including 1078 congenital diaphragmatic hernia (CDH) trios, 204 orofacial cleft birth defect trios, 107 nonsyndromic craniosynostosis, and 185 esophageal atresia and tracheoesophageal fistula (EA/TEF) trios. We called de novo coding variants from probands. We also obtained published de novo coding variants from studies of congenital heart disease (The Pediatric Cardiac Genetics Consortium, N~3000), neurodevelopmental disorders (N~30,000), orofacial cleft (N~750), autism (N=22660) and syndromic craniosynostosis (N~520). Consistent with previous studies, there was a significant enrichment of de novo protein truncating variants and missense variants in all GMKF conditions, especially among constrained genes (ad hoc defined as gnomAD pLI≥0.5) that are intolerant of loss of function variants.

Enrichment rates are larger in the syndromic group compared to nonsyndormic group for each condition. PARs (population attributable risk) of missense and LGD variants in syndromic cases are larger than those of nonsyndromic cases in CDH, autism and CHD. To estimate pleiotropic effects, we applied a Poisson test of damaging variants (PTV and missense variants with CADD phred-scale score≥20), with expected mean counts estimated by a background mutation model, in the combined data of all the conditions. We identified 227 genes with exome-wide significance (p-value<2.5e-6). Among these genes, 157 have at least one damaging variants in at least two conditions, including 149 estimated NDD genes and eight putative novel risk genes. One notable gene is *ZFHX3* (p=1.2e-7), with damaging variants in individuals with NDD, CHD, cleft, nonsyndromic craniosynostosis and CDH. These results provide new insight for rare congenital anomalies.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7045F: Pre-implantation genetic testing by polygenic risk scores selects different embryos across score construction methods with randomness

Authors:

S. Namba[1,2,3], M. Akiyama[4], H. Hamanoue[5], K. Kato[2], M. Kawashima[6], I. Kushima[7], K. Matsuda[1], M. Nakatochi[7], S. Ogishima[8], K. Sonehara[1,2,3], K. Suzuki[2], A. Takata[3], G. Tamiya[8,3], C. Tanikawa[1], K. Yamamoto[2], N. Yamamoto[2], The BioBank Japan Project, N. Ozaki[7], Y. Okada[1,2,3]; [1]The Univ. of Tokyo, Tokyo, Japan, [2]Osaka Univ., Suita, Japan, [3]RIKEN Ctr., Yokohama, Japan, [4]Kyushu Univ., Fukuoka, Japan, [5]Yokohama City Univ. Hosp., Yokohama, Japan, [6]ROIS-DS, Kashiwa, Japan, [7]Nagoya Univ., Nagoya, Japan, [8]Tohoku Univ., Sendai, Japan

Abstract:

Private enterprises have marketed preimplantation genetic testing by polygenic risk scores (PRS) to select embryos with 'desirable' potential. Despite researchers' critiques (e.g., small gains and potential eugenic usage), public expectations were high, showing that the public remains unconvinced by the existing concerns. Here, we question the testing's validity; selected embryos can differ across PRS methods and replicates of the same method. To empirically assess this concern, we randomly generated 500 virtual mate pairs using publicly available biobank resources and simulated the recombination process to generate ten embryo genotypes per mate pair. We then used six popular methods to construct PRS for adult height, a well-known target of mating preferences. Strikingly, in any combination of PRS methods, the same embryo was selected as the top-ranked one with only a 30.0% chance (median; range, 20.4-41.6%). The

top-ranked embryo by a particular method was lowest-ranked by at least one other method with a median 5.9% chance (3.4-8.0%), and vice versa (median 5.2% [4.0-7.8%]), implying that the unselected embryos would be selected if other PRS methods were used. We observed consistent results when repeating the simulation for type 2 diabetes, one of the commercial targets of this testing. The rank inconsistency of embryos was invariant under various degrees of assortative mating, suggesting that the inconsistency was independent of mating patterns. As the growing GWAS sample size improves PRS prediction accuracy, we tested whether it can solve the rank inconsistency. We simulated saturated summary statistics with a sample size larger than any studies to date (N=1×10^7), but the top-ranked embryos were still different across PRS methods (median 24.0 [18.6-34.8%]), indicating that the inconsistency cannot be solved in the current or near-future sample sizes. Furthermore, even simply repeating the same method produced a different embryo ranking, illustrating that PRS construction is a probabilistic process with fluctuation. In conclusion, the selected embryo substantially relied on the choice of PRS method and the randomness in PRS construction, raising serious ethical concerns about selecting embryos and consequently discarding other embryos in an unreliable way.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7046F: Reassessment of Variants of Unknown Significance (VUS) in cases of undiagnosed fetal anomalies

Authors:

F. Hasegawa[1], K. Taniguchi[2], K. Ozawa[1], S. Wada[1], H. Sago[3], K. Hata[4]; [1]Ctr. for Maternal-Fetal-Neonatal and Reproductive Med., Natl. Ctr. for Child Hlth.and D, Setagaya-ku, Japan, [2]NDept. of Molecular and Cellular Biology, Gumma Universit, Maebashi, Japan, [3]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [4]Dept. of Human Molecular Genetics, Gunma Univ. Graduate Sch. of Med., Maebashi, Japan

Abstract:

Reevaluation of variants of unknown significance (VUS) using whole-exome sequencing (WES) is necessary to improve the diagnostic rate for genetic disorders. We performed WES in 133 cases presenting with fetal morphological abnormalities and reanalyzed 83 cases (66.2%) in which no clear pathogenic variants were found, using three methods. 1) Loss-of-function observed/expected upper bound fraction (LOEUF) was used to reevaluate the pathogenicity of disease-related gene variants that had been determined to be VUS; polymorphisms in genes with low LOEUF scores are considered more likely to cause haploinsufficiency. Consequently, we could extract candidate pathogenic variants of several genes that could cause X-linked recessive

disease in several cases. 2) In cases where clinical symptoms strongly suggested certain diseases but did not lead to definitive pathogenic variants, we reanalyzed the relevant genes for those diseases using SpliceAI. We identified a pathogenic variant causing a splice abnormality in the intron of a gene related to Meckel-Gruber syndrome. 3) For the 23 patients whose relevant genes were unknown because they were asymptomatic, we first targeted 2,971 genes with a LOEUF score of less than 0.35, which were estimated to have low genetic variant tolerance. Subsequently, the variants possessed by these genes (including intronic variants) were re-evaluated using splice AI. In one case of severe fetal edema with multiple malformations, a variant that could cause a splice abnormality in a known etiologic gene for X-linked recessive disease (and the possibility that this was a severe case of the disease) were strongly suspected.The weighting of pathogenicity by LOEUF and the prediction of splice abnormality were considered useful in reevaluating VUS.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7047F: Towards better reproductive health in Qatar: a custom genotyping array for expanded genetic carrier screening

Authors:

R. Messai Badji, Y. Al-Sarraj, T. Fadl, W. Al-Muftah; Qatar Precision Hlth.Inst., Doha, Qatar

Abstract:

Background: The prevalence of consanguineous marriages in Qatar contributes to a higher incidence of genetic disorders, underscoring the need for targeted genetic screening programs to prevent problematic pregnancies and reduce the healthcare burden of genetic diseases.
Objective: This proposal outlines the development and implementation of an Expanded Reproductive Genetic Screening (ERGS) program in Qatar, utilizing a custom-designed SNP genotyping microarray, the QChip. The focus is on assessing the feasibility and acceptability of this screening within the Qatari population.
Methods: The QChip is developed using the extensive genomic data available from the Qatar Genome Program (QGP). It is specifically tailored to detect genetic mutations prevalent in the Qatari population. The design process integrates prevalent pathogenic variants identified in the QGP dataset, or prevalent in GnomadV4 to enhance its applicability across different populations.
QChip Design: The QChip utilizes a selective approach, focusing on specific genetic markers rather than the entire genome. This design facilitates a more efficient and cost-effective screening process, essential for large-scale implementation. The array targets key genes and

variants that are critical for reproductive genetic health in the Qatari population.
Pilot Study Plan: The proposal details a pilot study to evaluate the effectiveness of the QChip in a clinical setting. The study will assess the technical performance of the array, its integration into healthcare workflows, and its acceptance by the population. This initial phase is crucial for determining the practical aspects of deploying the QChip as part of a national screening program.
Conclusion: The development of the QChip represents a significant advancement in genetic screening tailored to the needs of the Qatari population. By focusing on specific, relevant genetic markers, the QChip offers a practical solution for early detection of genetic risks associated with reproductive health.
Future Directions: The outcomes of the pilot study will inform the larger-scale deployment of the QChip within national health services, aiming to establish a robust, culturally acceptable genetic screening program in Qatar.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Friday Session

Board 7048F: Unraveling the mutational landscape of Placenta Accreta Spectrum (PAS) disorder

Authors:

R-C. Robertson[1], E. Zhang[1], C. Mason[1], H. Chavez[2], M. Huang[1], K. Fisch[1]; [1]UC San Diego, La Jolla, CA, [2]Univ. of California, San Diego, Riverside, CA

Abstract:

Placenta accreta spectrum disorders (PAS) are characterized by invasive placentation and can lead to obstetric complications, such as maternal hemorrhage. The pathology observed in PAS is similar to that seen in tumors, such as cell invasiveness and altered angiogenesis, motivating the use of cutting-edge genomics techniques to molecularly profile PAS. To define the mutational landscape of PAS, we obtained placental tissue, fetal cord blood, and maternal blood from 10 PAS pregnancies (n=6 accreta, 2 increta, 2 percreta) from the UCSD Center for Perinatal Discovery Biobank and extracted DNA using the QIAsymphony DSP DNA Mini Kit. We performed 30x whole genome sequencing on N=30 samples from 10 pregnancies on a NovaSeqX Plus (Illumina). We called somatic mutations unique to the placenta relative to cord blood using GATK Best Practices with Mutect2, using the hg38 reference genome and bcbio-nextgen. We further filtered out maternal germline variants to remove maternal blood contamination and annotated variants with Variant Effect Predictor. We calculated mutational burden, performed mutational signature analysis and gene set enrichment analysis. Of the 10 pregnancies, 802

genes had somatic mutations with a median of 62 mutations per placenta. Of these mutations, 25 were found to be recurrent in at least 2 placentas. In 9/10 cases, damaging mutations were discovered in genes associated with fetal developmental abnormalities. Significant somatic interactions were found between 12 genes ($p<0.05$); for example, *DNHD1* (Gln4250Lys) was mutated in 2/10 cases and showed significant somatic interactions with *ARID1A* (Lys2189Asn). Of the 10 cases, 4 harbored at least one nonsense mutation, 8 in total. Interestingly, one case carried a nonstop mutation in *F2* (Ter623Tyr), which encodes a coagulation factor involved in tissue repair, angiogenesis, and vascular integrity maintenance. These mutations were also enriched in Regulation of Insulin-like Growth Factor transport and uptake by Insulin-like Growth Factor Binding Proteins (adj. P = 0.032). Lastly, mutational signatures that represent patterns of somatic mutations that are correlated with mutational processes identified *SBSA*, *SBSB*, *SBSC*, *SBSCD* and *SBS4-like* mutational signatures in PAS placental tissue. *SBS4-like* (tobacco smoke mutagenesis) signature was positively correlated with mutational burden ($R^2 = 0.71$), suggesting an effect of maternal environmental exposures on placental DNA damage. Together, these results shed light on the genetic pathologies of PAS, which we will further investigate with functional studies in iPSC-derived trophoblast cell lines.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4002F: A Bayesian Rare Variant Association Analysis Method Incorporating Functional Variant Information ⭐

Authors:

X. Sun[1], S. Han[2,1], L. Sloofman[3], F. Satterstrom[4], G. Wang[5], L. Liang[1], N. Knoblauch[1], S. Zhao[6], Y. Liu[1], W. Sheng[2], H. T. Nguyen[7], X. He[1]; [1]The Univ. of Chicago, Chicago, IL, [2]Marquette Univ., Milwaukee, WI, [3]Seaver Autism Ctr., New York, NY, [4]Broad Inst., Cambridge, MA, [5]Columbia Univ., New York, NY, [6]Dartmouth Coll., Lebanon, NH, [7]Virginia Commonwealth Univ., Richmond, VA

Abstract:

GWAS have identified numerous variants for complex traits. The common variants found by GWAS, however, have small effect sizes, and exhibit linkage disequilibrium (LD), making it difficult to find causal variants. Large whole-exome/genome sequencing datasets enable rare variant association studies. However, single-variant tests lack power due to low frequencies. While many gene-based methods have been developed, they often make unrealistic assumptions, e.g. the commonly used burden tests implicitly assume all rare variants in a gene

have the same risks. As a result, the power of these methods is often limited in practice. We developed a Bayesian method: MIxture model based Rare variant Analysis on GEnes (MIRAGE), for rare variant analysis in the matched case-control studies or transmitted variant data from trio studies. MIRAGE captures the heterogeneity of variant effects by treating all variants of a gene as a mixture of risk and non-risk variants. The prior probabilities of variants being risk variants are modeled as a function of variant annotations such as allele frequencies and deleteriousness scores. We have performed extensive simulations to assess MIRAGE. MIRAGE has substantially higher power in identifying causal genes than existing tests such as burden test, SKAT, and ACAT, and showed a calibrated Bayesian FDR control. We applied MIRAGE to WES data from 7,291 ASD-affected child-parent trios, from the Autism Sequencing Consortium and the SPARK study. While current methods detect no significant risk genes, MIRAGE identified 7 candidates. The top genes found by MIRAGE are significantly enriched with gene sets commonly linked to autism, e.g. constrained genes or synaptic genes. The candidate genes from MIRAGE included both known risk genes such as PLXNA1 and NAV3, and novel and promising genes such as MYH10. In conclusion, MIRAGE is a novel and powerful Bayesian method for rare variant association analysis. it is accessible at https://xinhe-lab.github.io/mirage.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4003F: A multi-ancestry reference panel from 10,721 individuals to improve the genotype imputation for individuals of East Asian- and Polynesian-Ancestry

Authors:

X. Wang[1], B. Dinh[1], X. Sheng[1], E. Young[2], M. Leask[3], K. Taparra[4], L. K. Stamp[5], N. Dalbeth[6], R. Murphy[6], J. de Zoysa[6], T. Naseri[7,8], M. Reupena[9], S. Viali[10,11,12], S. E. Castel[13], T. R. Merriman[3,14], J. Carlson[15], D. E. Weeks[16], L. Wilkens[17], L. Le Marchand[18], I. Hall[10], C. A. Haiman[1], N. Stitziel[19], C. W. K. Chiang[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Washington Univ. in St. Louis, St Louis, MO, [3]Univ. of Otago, Dunedin, New Zealand, [4]Stanford Univ., Stanford, CA, [5]Univ. of Otago, Christchurch, Christchurch Central City, New Zealand, [6]Univ. of Auckland, Auckland, New Zealand, [7]Brown Univ., Providence, RI, [8]Naseri & Associates Publ. Hlth.Consultancy Firm & Family Hlth.Clinic, Apia, Samoa, [9]Lutia i Puava ae Mapu i Fagalele, Apia, Samoa, [10]Yale Univ., New Haven, CT, [11]Oceanian Univ. of Med., Apia, Samoa, [12]Natl. Univ. of Samoa, Apia, Samoa, [13]Variant Bio, Seattle, WA, [14]Univ. of Alabama at Birmingham, Birmingham, AL, [15]Univ. of Pittsburgh, Pittsburgh, PA, [16]Univ of Pittsburgh, Pittsburgh, PA, [17]Univ. of Hawai'i, Honolulu, Honolulu, HI, [18]Univ Hawaii, Honolulu, HI, [19]Washington Univ. Sch. of Med., Saint Louis, MO

Abstract:

Genotype imputation lays the foundation for robust association testing. The latest version 3 release of the reference panel from TOPMed improved imputation accuracy dramatically for certain under-represented populations, such as African Americans and Latinos. However, the imputation accuracy for other global populations remains low. Here, we present an imputation reference panel based on whole genome sequences (WGS) of 10,721 individuals from the Multiethnic Cohort (MEC; 1,270 African Americans, 1,065 Native Hawaiians, 3,106 Japanese Americans, 4,141 Latinos, and 1,139 non-Hispanic whites), consisting of 82 million (M) SNPs and 5 million indels. We evaluated the imputation accuracy of the MEC reference panel by imputing the genotype of 100 held-out individuals, compared to their genotypes from WGS data as the "ground truth". Despite containing < 10% as many samples and < 20% as many variants as TOPMed, the MEC panel outperforms the TOPMed panel in accuracy for Native Hawaiians (MEC-NH) and Japanese (MEC-JA). For example, the mean Pearson's R^2 between imputed dosages and the sequenced genotypes for variants with 1-5% allele frequencies (AF) in MEC-JA improved from 0.78 when imputing with TOPMed to 0.90 when imputing with the MEC reference. We further boosted the imputation accuracy and genomic coverage of imputed variants by combining the results from TOPMed and MEC together via meta-imputation. For example, we found a 28% and 63% increase in high-quality ($R^2 > 0.8$) imputed variants with 1-5% AF by meta-imputation for MEC-NH and MEC-JA (2.87M and 2.44M variants), respectively, compared to TOPMed imputation alone (2.24M and 1.49M variants). These observations were replicated in additional genome-wide array datasets of 3,119 Sāmoans from the Sāmoan Adiposity Study and 4,176 Japanese from BioBank Japan, demonstrating the generalized utility of the MEC reference panel. Finally, we showed that meta-imputation with the MEC reference panel would improve imputation accuracy for variants enriched in Māori from Aotearoa/New Zealand, but rare elsewhere (< 0.1% AF in gnomAD). For instance, there is an increase of 21% in the number of high-quality Polynesian-enriched variants (1-5% AF in Māori; $R^2 > 0.8$) by meta-imputation (0.188M variants) in MEC-NH compared to TOPMed imputation alone (0.155M variants). Our results suggest that the multi-ancestry MEC reference panel, particularly when combined with TOPMed through meta-imputation, will improve the power of genetic mapping studies and identify population-enriched associations in understudied populations, such as those of East Asian and Polynesian ancestry.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4004F: A New Approach for Discovery of Functional Links Connecting Non-Coding Regulatory Variants to Gene Targets

Authors:

H. Farooq, K. Chronis, J. Liang; Univ. of Illinois Chicago, Chicago, IL

Abstract:

Genome-wide association studies (GWAS) have linked thousands of genetic variants to various complex traits or diseases. However, most identified variants have weak individual effects, are correlated with nearby polymorphisms due to linkage disequilibrium (LD), and are located in non-coding cis-regulatory elements (CREs). These characteristics complicate the assessment of each variant's direct impact on tissue specific gene expression and phenotype. To address this challenge, we have developed a novel algorithm that leverages polymer folding and 3D chromatin interactions to prioritize and identify putative causal variants and their target genes. From the millions of eQTL-Gene pairs identified by GTEx in human somatic tissues, we classify only ~10-20% as putative functional eQTL-Gene pairs supported by phenotypic associations confirmed through CRISPR deletion experiments. Our findings show that unlike most variants, functional eQTL-Gene pairs predominantly reside within the same topologically associating domain (TAD) and have strong associations with cell-type specific cis-regulatory elements (CREs), enriched for binding sites of tissue-specific transcription factors. Unlike most approaches that rely on linear distance or other chromatin features (histone code, accessibility), our algorithm emphasizes the importance of physical interactions and 3D chromatin folding in gene regulation, as the identified eQTL-Gene pairs are all among the small fraction of physical chromatin interactions sufficient for chromatin locus folding. Overall, our algorithm reduces false positive associations between DNA variants and genes identified by eQTL analysis and uncovers novel variant-gene pair associations. These findings suggest a mechanism where a small number of regulatory variants control tissue specific gene expression via their physical association with target genes confined within the same TAD. Our approach provides new insights into the molecular mechanisms driving GWAS phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4005F: A novel GWAS single-cell eQTL colocalization test improves power and resolves potentially disease-mediating cell types

Authors:

J. Mitchel[1], P. V. Kharchenko[2], S. Sunyaev[1]; [1]Harvard Med. Sch., Boston, MA, [2]Altos Labs, San Diego, CA

Abstract:

Expression quantitative trait loci (eQTLs) can aid in the interpretation GWAS findings by linking non-coding SNPs to genes and identifying the cell types where their effect takes place. A test to identify if the same SNP drives both a GWAS and eQTL locus is referred to as a "colocalization test". Surprisingly, systematic application of colocalization tests fail to identify evidence of shared effects at the majority of GWAS loci when using bulk tissue eQTLs (a phenomenon termed "missing regulation"; Connally et al. 2023 eLife). In a promising direction, recent studies have identified many new colocalizations by using single-cell expression data to map eQTLs in separate cell types. However, current single-cell approaches may be underpowered in small cell populations and depend on accurate cell clustering, which may be unreliable for cell subtypes/states. To improve power and robustness of single-cell colocalization tests, we developed a cluster-free method we call scJLIM. Specifically, we make use of mixed models that test if a SNP's effect on expression is dependent on a cell's PC coordinates. By fitting models for all SNPs in a locus and computing estimated effects at each cell's PC coordinates, our method can test each individual cell for colocalization with a GWAS locus (multiple testing corrections applied). We generated simulated GWAS data and cell-type-specific single-cell eQTLs for scenarios where the same causal SNP drives both traits. Compared to bulk colocalization, our approach demonstrated increased overall power with comparable calibration. Our approach also accurately identified the ground true causal cell populations. Notably, power is especially higher in scenarios where the causal cell population is small and other nearby eQTLs exist in non-causal cell types. We have further applied scJLIM to blood and brain scRNA-seq datasets, identifying new colocalizations not identified with bulk analyses. Additionally, we have characterized properties of SNPs (e.g., functional elements) and genes (mendelian/conserved genes) that uniquely colocalize in single-cell data. Overall, we developed a novel GWAS-eQTL colocalization method for single-cell expression data that shows improved power over existing methods and can identify fine-resolution cell populations that are potentially mediating the genetic effects on a given trait.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4006F: A robust and efficient framework in transcriptome-wide association studies using cross-tissue learner

Authors:

M. Billah, K. Zhang; Michigan Technological Univ., Houghton, MI

Abstract:

The advent of transcriptome-wide association studies (TWAS) has revolutionized the perspective of classical genome-wide association studies (GWAS). TWAS leverages the relationship between gene expression and genotypes from the reference panels (e.g., GTEx) and uses imputed gene expression to identify susceptible genetic variants from GWAS. While methods have been developed based on the expression data from a single tissue/study, methods based on the gene expression data from multiple tissues/studies have garnered significant attention due to their enhanced statistical power. However, existing methods based on the gene expression data from multiple tissues/studies often lose valuable information when integrating outcomes from individual tissue/study and require prolonged computational time to yield the results. In this paper, we introduce a novel cross-tissue/study learner (CTL) approach to address these limitations. By incorporating the heterogeneous nature of human gene expression values across different tissues into its framework, CTL learns from the single-tissue learners (STLs) by leveraging the weights of the least-performing STL and modifying losses for all other STLs. This method efficiently combines gene expression values from the diverse human tissues in association with genetic variation and quantifies gene-trait associations comprehensively. We conducted extensive simulations and compared our results with UTMOST (unified test for molecular signatures), a prominent method in the field. Although both methods controlled the type I error rates at a significant threshold, our approach outperformed UTMOST by achieving higher statistical power across different simulation settings. Additionally, the CTL approach requires less than half the computational time compared to the UTMOST method. We expect that the CTL method will be pivotal in the intricate domain of heterogeneous cross-tissue/study gene expression analysis, discovering additional causal genes through the integration of gene-trait associations within genetic studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4007F: A scalable workflow for submicron-resolution spatial transcriptomics data from Seq-Scope

Authors:

W. Cheng[1], Y. Kim[1], C-S. Cho[1], Y. Hwang[1], Y. Si[1], A. Park[1], M. Schrank[1], J-E. Hsu[1], A. Anacleto[1], J. Xi[1], M. Kim[1], E. Pedersen[1], O. Koues[2], W. Thomas[1], C. Lee[3], J. Lee[1], H. Kang[4]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan Advanced Genomics Core, Ann Arbor, MI, [3]Harvard Med. Sch., Brookline, MA, [4]Univ Michigan, Ann Arbor, Ann Arbor, MI

Abstract:

Spatial transcriptomics technologies have revolutionized gene expression studies but face challenges such as low resolution, limited gene coverage, and high costs. Seq-Scope mitigates these challenges by repurposing the Illumina sequencing platform to provide submicrometer resolution, high capture efficiency, and precise histopathological annotation at a significantly reduced cost. This approach allows profiling multiple tissue sections within an area of 7 mm × 7 mm or larger using a single Illumina NovaSeq 6000 flow cell.
To streamline the application of Seq-Scope and address issues related to procedural complexity, scalability, and reproducibility, we introduce NovaScope (https://seqscope.github.io/NovaScope), a Snakemake-based pipeline, to process high-resolution spatial transcriptomics data from SeqScope efficiently. It transforms raw sequencing files into spatial digital gene expression (SGE) matrices at submicron resolution. Key functionalities include generating spatial barcode maps, matching spatial barcodes with sequencing files, aligning these files, and producing SGE matrices, with providing quality control metrics for alignment and visualizations of barcode distribution at each step. In addition, NovaScope's modular design allows precise alignment of histological images with transcriptome data, conversion of the SGE matrix for compatibility with our segmentation-free method (FICTURE), and organization of the SGE matrix into customizable hexagonal grids.
To facilitate user adoption, NovaScope includes three example datasets from mouse liver tissue along with detailed documentation. A Docker image is also provided for straightforward installation and execution with Docker or Singularity, allowing users to reproduce example results and analyze their datasets.
We also provide an exemplary downstream analysis (https://seqscope.github.io/NovaScope-exemplary-downstream-analysis/) for the SGE matrix generated by NovaScope. This analysis uses Latent Dirichlet allocation and Seurat to identify spatial factors for the example datasets independently, which effectively identify metabolically zonated hepatocytes and various cellular states, including damaged or antiviral response hepatocytes, hepatic stellate cells, macrophages/myofibroblasts, cholangiocytes, and red blood cells. Then, FICTURE enhances this analysis by projecting such spatial factors onto the histological space with pixel-level precision.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4008F: A Statistical Framework for Partitioning the Effects of Alleles Mediated by Parental Genetic Nurture and Direct Genetic Effects while Controlling for Assortative Mating

Authors:

A. Perry[1], A. Akbari[1], A. Durvasula[2], C-E. Fischer[3], N. Patterson[4], D. Reich[5]; [1]Harvard Univ., Cambridge, MA, [2]Univ. of Southern California, Los Angeles, CA, [3]Univ. of Tromsø, Tromsø, Norway, [4]Broad Inst MIT & Harvard, Cambridge, MA, [5]Harvard Univ., Boston, MA

Abstract:

Understanding the etiology of disease is crucial to treatment development. The strength of genetic and environmental contribution to pathology often is ambiguous, due in part to statistical difficulties in separating these causal pathways. One such issue is that a subject's alleles are not present only in the subject, but also in their relatives. The same alleles present in a subject may thus impact the traits of relatives who are in the subject's environment, and the relatives' traits may impact the subject's likelihood of developing a disease. For example, genetic variants in pregnant mothers are known to impact maternal blood pressure, which impacts birth weight of the neonate. These "parental genetic nurture" effects confound gene-trait associations because the alleles which impact parental traits have a 50% chance of being transmitted to the embryo during conception. For example, if we perform a statistical analysis to detect alleles correlated to "birth weight", we will find blood-pressuring raising alleles correlate to birth weight despite the fact these alleles only influence birth weight when expressed in a pregnant mother. Furthermore, these genetic associations are even more strongly confounded if parents with similar alleles tend to have kids together, a process known as assortative mating. A barrier to disentangling the effects of child and parental alleles is the lack of genotyped mothers, fathers, and children (family "trios"). We show that all existing methods of estimating the degree to which an allele's impact on disease is mediated by parents vs. children are underpowered at presently-available trio sample sizes. We propose a method for differentiating between child-mediated allelic effects, parental genetic nurture, and assortative mating. We describe applying our method to a dataset of dozens of phenotypes in over 20,000 genotyped trios: The Norwegian Mother, Father, and Child Cohort Study. We anticipate our investigations will shed light on the extent to which gene-trait associations are mediated by distinct causal pathways, increasing the utility of gene-trait associations for subsequent clinical investigation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4009F: Activating Transcription Factor 3 Associated with Nausea in Gastroparesis Patients - a Whole Genome Sequencing Study

Authors:

H. Bai, S. Smieszek, J. Carlin, C. Xiao, G. Birznieks, C. Polymeropoulos, M. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Abstract:

Gastroparesis is a severe medical condition characterized by delayed gastric emptying and symptoms of nausea, vomiting, bloating, fullness after meals, and abdominal pain. Thus far, little is known about the genetic risk factors underlying gastroparesis nausea. We explored the association of genetic variants with nausea severity in a clinical gastroparesis study.Whole genome sequencing data was obtained from adults with diabetic or idiopathic gastroparesis. Nausea score was assessed from a daily symptom dairy with a range from 0 mild to 5 severe. GWAS using linear model was conducted on variants with MAF ≥0.01 with adjustment of the subject's age, sex, BMI, etiology, and the first 3 PCs.Variant clusters from multiple genes are significantly associated with baseline nausea score at p=1e-5 level. Top variants are within *ATF3*, *PP2D1*, *CNTN4*, *BMP2*, *ABL1*, *CTNNA3* and *KIF6*. Interestingly, upon comparing our results with a transcriptomic analysis, we found that *ATF3* and *BMP2* were overlapping with genes differentially expressed in both diabetic and idiopathic gastroparesis when compared with controls. *ABL1* was, however, differentially expressed in diabetic cases whereas *CNTN4* was differentially expressed in idiopathic controls. We expanded the analysis with single gene tests which shows that *ATF3* is significant especially for MAF>0.05 variants (SKAT p=0.02). The top variant in *ATF3* is rs79151334 (beta= 0.68, p=5.76e-6). Additionally, we report new associations of *CTNNA3*, *PP2D1*, and *KIF6* as well as a cluster of variants between *CDH2* and *MIR302F*, with baseline nausea score. Among those, ATF3 modulates the immune response, atherogenesis, cell cycle, apoptosis, and glucose homeostasis, and both *CTNNA3* and *CDH2* encode cadherin related proteins, suggesting this type of junction could play a role in the underlying mechanism. Here we identified potential nausea related variants involved in immunological responses and metabolism that could provide further understanding of the pathophysiology to be tested in future investigations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4010F: Advances in genetic relatedness software and data simulation for scalable, high throughput analysis

Authors:

G. Evans[1], **J. Below**[2], C. Huff[3], J. Baker[1], H-H. chen[1], L. Petty[4], A. Petty[4]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ Med Ctr., Nashville, TN, [3]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Computational methods that enable genetic relatedness estimation - often in the absence of known familial pedigree structure - have become increasingly critical for basic and clinical genomics research especially in the context of large and regionally enrolled biobank datasets in which relatedness may be higher due to non-random sampling. Robustly estimating pairwise relatedness and reconstructing pedigrees in large and diverse datasets requires scalable and interoperable methods that have been benchmarked against diverse and recently admixed genetic datasets. To this end, we developed an integrative computational pipeline to improve upon software tools that leverage identity-by-descent (IBD) data to characterize relatedness between individuals at biobank scale. We introduce a composite relationship estimation algorithm that uses both genome-wide IBD estimates from the program PRIMUS and shared segment length-/quantity-based estimates from the program ERSA to build more accurate probability vectors for each pairwise relationship in a family. We also introduce a robust genetic data simulation program that builds IBD- and shared-segment based datasets. Our software is faster, highly parallel in execution, and flexible to founder data sources - offering generation of admixed simulated datasets - while boasting substantially lower computational overhead, with decreases in RAM & CPU usage and output file size.
We present preliminary validation of our software across varied ancestral genomic data (from 1000 Genomes Project), family size, and pedigree missingness (select family members removed from the data input) in both PRIMUS and ERSA as well as our composite algorithm, COMPADRE. We estimated familial relatedness in 1500 simulated pedigrees spanning both 20- and 40-person families and 0% to 20% missingness. In initial simulations derived from the EUR superpopulation, our composite algorithm increased both total count of correct relationship estimations per pedigree and statistical probability of the selected relationship being correct (as a proxy measure for relationship confidence) by ~20% and ~2%, respectively. We anticipate an even more substantial delta in relationship estimation performance in additional superpopulation analyses with heightened admixture. We demonstrate initial utility of our pipeline to assist in both developing new and expanding existing software tools, as well as our own improvements to such tools in the relatedness field and their downstream effects of improved power and reduced bias in disease gene discovery applications.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4011F: Allele stacking for cohort-scale analysis of long-read sequencing to identify functional non-coding variants

Authors:

D. Yang; Albert Einstein Coll. of Med., Bronx, NY

Abstract:

Long read sequencing (LRS) now allows us to generate DNA sequence and DNA methylation information on the same strands of DNA. This permits genome-wide identification of functional non-coding variants throughout the human genome. With recent initiatives to expand the catalogue of LRS human reference genomes, such as the Human Pangenome Reference and those assembled by the CoLoRSDB consortium, there is a growing need for LRS-tailored analysis tools which can identify candidate functional non-coding genetic variants at the cohort scale. We are leveraging large LRS datasets to develop a new analysis approach specifically designed to identify functional non-coding variants at the cohort scale. Our approach is built on Pacific Biosciences' (PacBio) current suite of LRS whole genome analysis pipelines and extends its functionalities to parse out candidate functional variants for analysis and visualization. The goal of this approach is to improve functional non-coding variant detection by identifying those which alter local patterns of DNA methylation (5-methylcytosine (5mC)). This is achieved by building upon PacBio's MethBat program by aggregating or "stacking" alleles representing samples from an entire cohort and generating consensus patterns of DNA methylation across multiple samples for a given cell type based on the results of MethBat's segmentation algorithm. We utilize these consensus regions to analyze sequence variation within these loci with distinctive DNA methylation in the local haplotype. By extracting the LRS reads from samples which possess these variants, we present an extension of PacBio's LRS analysis tools specifically designed to effectively parse LRS data to identify candidate functional variants for visualization and downstream validation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4012F: An accurate genetic colocalization method for the HLA locus

Authors:

G. Butler-Laporte[1], T. Lu[2], S. Morris[3], W. Zhang[4], F. Hamilton[5], K. Lin[3], R. Nanjala[3], B. Richards[6], M-H. Lee[7], L. Yang[3], P. Yao[3], Y. Luo[3], I. Milwood[3], R. Walters[3], A. Mentzer[3]; [1]McGill Univ., Montreal, QC, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Univ. of Oxford, Oxford, United Kingdom, [4]Montreal Heart Inst., Montreal, QC, Canada, [5]Univ. of Bristol, Bristol, United Kingdom, [6]McGill Univ. | 5 Prime Sci., Montreal, QC, Canada, [7]Natl. Yang-Ming Univ., Taipei, China

Abstract:

When an HLA signal is found in genome-wide association studies, it is common to follow-up with dedicated HLA allele association studies. However, contrary to the rest of the genome, the HLA still lacks a causal inference method allowing for the translation of genetic results to actionable therapeutic targets. This is because the HLA's long range linkage disequilibrium, high number of polymorphisms, and complex pleiotropy prevents the use of methods like genetic colocalization and Mendelian randomization. Hence, when two traits show a genetic association at the HLA, it remains difficult to establish if the genetic signal underlying the two traits is the same, and at which HLA gene it occurs. A method allowing for the identification of these signals is still an unmet need in genetic epidemiology.
Here we present a method that uses HLA alleles to assess genetic colocalization of two traits at HLA genes. The method works by using Bayesian variable selection to assign a signature of allele causality probabilities at each HLA gene, and then by performing genetic colocalization on these probabilities at each gene. If two phenotypes colocalize at a particular HLA gene, then they should share a similar causal probability signature at these genes.
We show through simulation that the method correctly identifies truly colocalizing genes, while limiting the rate of false positives. We then test the method in two clinically relevant real-world scenarios. First, we show genetic colocalization of Epstein-Barr virus and multiple sclerosis in the UK Biobank at *HLA-DRB1* and *HLA-DQB1*, and second, we show colocalization of hepatitis B and liver disease in the China Kadoorie Biobank and the Taiwan Biobank at *HLA-DPB1*. Finally, using the UK Biobank serology data, we show that the method performs well in multiple negative control scenarios, and that it identifies clinically relevant shared causal HLA genes between human diseases and pathogens. To our knowledge, this is the first genetic colocalization method that can provide biologically and clinically meaningful causal insights into shared genetic determinants of diseases and traits at the HLA locus.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4013F: An Angle-based Bayesian Transfer Learning Approach for Enhanced Estimation and Inference

Authors:

D. Lai[1], T. Gu[2]; [1]Dept. of Statistics and Actuarial Sci., The Univ. of Hong Kong, Hong Kong, China, [2]Dept. of Biostatistics, Columbia Mailman Sch. of Publ. Hlth., New York, NY

Abstract:

Background: We consider a high dimensional sparse regression problem. We have a target study of interest where the sample size may be limited, along with a set of related source

studies that can potentially be helpful to the target model learning. Transfer learning techniques aim to enhance the target model performance while leveraging shared knowledge from related source models, which is especially suitable in our case. While many frequentist transfer learning methods have been developed, they mainly focus on improving prediction performance, with less emphasis on statistical inference. Bayesian approaches are appealing because they naturally provide uncertainty quantifications for parameter estimates. There have been few efforts to enhance estimation efficiency for improved inference through Bayesian transfer learning approaches, while the existing ones are either restricted to single-source information transfer or require individual-level data that might not be available in practice. **Methods:** We propose a novel Bayesian angle-based transfer learning approach that can incorporate multiple source studies and only requires summary-level model estimates from source data. The proposed method employs the horseshoe prior, shrinking the target model parameters toward a weighted source estimate, with weights determined by the angle-based similarity between the target and the source parameter vectors. The source estimates are also used to inform the number of non-zero coefficients in the prior. Compared to the standard horseshoe prior used in the target-only study, which shrinks the parameters towards zero, the source estimates can help identify non-zero target parameters and their effect sizes more accurately, especially when the target sample size is small. **Results:** Numerical studies demonstrate significant improvements in estimation and inference accuracy over global-local shrinkage priors that only use the target data. Furthermore, we show the feasibility and superiority of the proposed method to improve the estimation performance of low-density lipoprotein cholesterol using data from the electronic MEdical Records and GEnomics (eMERGE) Network. **Conclusion:** We propose a Bayesian angle-based transfer learning method that uses a source-guided horseshoe prior for enhanced parameter estimation compared to using target data alone and providing uncertainty quantification for inference. It has great potential in applications such as multi-ancestry fine-mapping to enhance accuracy and resolution by leveraging association information across various ancestries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4014F: An efficient network Mendelian randomization method for network structure discovery and effect estimation

Authors:

S. Eng, J. Morrison; Univ. of Michigan, Ann Arbor, MI

Abstract:

Mendelian randomization (MR) is a causal inference method in which genetic variants are used as instrumental variables to estimate causal effects. Univariable and multivariable MR methods are useful when there is a single outcome trait of interest. However, in many cases, we may be more interested in understanding the complete causal network connecting multiple traits. For example, identifying causal networks connecting metabolites or protein-protein interactions may help elucidate biological pathways. It is standard practice in biology to produce undirected correlation networks for these phenotypes. By contrast, use of MR-based techniques can allow us to produce directed, causally interpretable networks. Network inference can also be used to directly answer mediation questions, which are typically addressed through stepwise application of multivariable MR. We propose an extension of the MR framework for estimation of linear acyclic causal networks of multiple traits using GWAS summary statistics. The network MR problem consists of two subproblems: (1) estimation of direct effects in a given network structure and (2) network structure discovery. Our proposed method, network empirical shrinkage Mendelian randomization (NESMR), solves both of these subproblems. First, we develop a method to efficiently estimate causal effects for a specified network structure by maximizing a hierarchical likelihood via variational inference. We show that this method provides consistent estimation of direct effects when the specified network structure contains all true edges. Second, we develop an efficient algorithm to identify the highest likelihood candidate network structures. This algorithm uses repeated application of multivariable MR to build a discovery graph, which asymptotically contains all non-zero edges. We then identify the highest likelihood acyclic sub-graph via a modified weighted feedback arc set algorithm of the discovery graph and provide estimated effects and confidence intervals for each edge. NESMR explores alternative candidate structures with similar likelihoods, allowing us to quantify uncertainty about the estimated network structure. This allows for discovery of graphs without a priori knowledge of the relationship between the phenotypes. We apply NESMR to the METSIM (The Metabolic Syndrome in Men study) metabolomics summary statistics to infer candidate causal networks between 792 metabolites. This novel approach identifies potential causal networks of metabolites, phenotypes and the relationship between the phenotypes and metabolites.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4015F: An omnibus test for multivariable Mendelian randomization with applications to FunGen-xQTL data

Authors:

Y. Cho[1], A. Wang[2], G. Wang[1], Z. Liu[1]; [1]Columbia Univ., New York, NY, [2]Univ. of Hong Kong, Hong Kong SAR, China

Abstract:

Recently, the use of two-sample multivariable Mendelian randomization (MVMR) has been increasing, combined with the public accessibility of summarized data from genome-wide association studies (GWAS) for human diseases and omics quantitative trait loci (QTL) analyses for molecular phenotypes such as mRNA and protein abundances. However, researchers often encounter challenges due to insufficient genetic variants, particularly when employing cis-QTL as instrumental variables (IVs) or following fine-mapping implementation to refine GWAS loci for causal variants in linkage disequilibrium (LD). Such challenges can obstruct model identification, thereby hindering the direct application of current MVMR methods. As an alternative, we propose an omnibus test capable of robustly detecting the shared common effect (fixed) and heterogeneous effects (random) within a linear mixed model framework. This testing involves an estimation procedure under the null hypothesis, excluding the causal effects, thus reducing the number of parameters to be estimated. Extensive simulation studies show that our method has good power to detect moderate effects in practical settings while controlling type I error rate well. We apply our method to the FunGen-xQTL dataset, which contains bulk RNA-seq data from three brain regions, pseudo bulk RNA-seq eQTL data from six brain cell types, RNA-seq pQTL data from one brain region, and bulk RNA-seq eQTL data from blood. Among the 80 genes selected for their strong signals through the mvSuSiE fine-mapping method, our analysis identified 17 genes (*PICALM*, *CASS4, RIN3, KCNN4, ERCC2, PLEKHA1, TSPAN14, PTK2B, BIN1, RASGEF1C, EGFR, USP6NL, SLC39A13, HTRA1, KCNAB1, CR1, HS3ST5*) that are significantly associated with Alzheimer's disease across 11 conditions, a finding not attainable using existing MVMR methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4017F: Ancestry-aware genome-wide association study of African American multiple sclerosis patients in All of Us ⭐

Authors:

S. Brugger[1], J. Denny[2], M. Davis[1]; [1]Brigham Young Univ., Provo, UT, [2]NIH, Bethesda, MD

Abstract:

Multiple sclerosis (MS) is a debilitating neurodegenerative disorder characterized by immune-mediated myelin destruction within the central nervous system. Both genetic and

environmental factors are implicated in an individual's risk of developing the disease. While MS has historically been associated with individuals of European descent, recent evidence suggests that prevalence is high among African American individuals as well. Further, African American MS patients experience more rapid disease progression and greater disease severity compared to European patients. HLA-DRB1*1501, an allele contributing the majority of genetic risk for Europeans, has been found to impart no MS risk if inherited from an African ancestor as opposed to a European ancestor. Additionally, over 200 single nucleotide polymorphisms (SNPs) associated with MS risk replicate poorly in African American patients. This suggests the genetic burden of disease for African-descent individuals is ancestry-dependent. We aimed to replicate these findings, as well as identify novel MS risk alleles specific to African genetic ancestry. We identified 2,692 MS cases in All of Us. We then focused on 271 participants of predominant African American ancestry and 1,000 matched controls within the *All of Us* research platform. Cases were defined as any individual possessing two or more MS billing codes in their electronic health record. Controls were matched for age, sex at birth, and self-reported race and could have no MS billing codes. A genome-wide association study (GWAS) was performed with PLINK v2.0 to serve as a baseline against which ancestry-aware analyses may be compared. Statistical phasing was performed with SHAPEIT2, after which local ancestry was estimated with RFMix. Local ancestry information was used to perform an ancestry-aware GWAS with Tractor to generate ancestry-specific summary statistics. Despite having adequate sample size for the strongest known MS associations, most known non-HLA MS risk variants replicated poorly in the basic GWAS. Overall, peaks of nominal significance were observed primarily in chromosomes 1, 9, and 12. rs140183230, an intergenic variant in chromosome 12, nearly achieved genome-wide significance (p = 5.62e-08, OR = 2.8); lead SNPs for each peak were unassociated with any disease in literature. Ancestry-aware analysis will provide critical context for the associations identified in the basic GWAS. These results are supportive of the current understanding of MS genetics in African American patients and may serve to highlight key differences from patients of a European genetic background.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4018F: Approximation of the polygenic risk score effects in a target population from summary statistics

Authors:

C. Chatzinakos[1], N. Daskalakis[2], M. Singh[3], J. Meyers[1], T. Bigdeli[4], R. Peterson[5]; [1]SUNY Downstate, Brooklyn, NY, [2]Harvard Med. Sch./ McLean, Belmont, NY, [3]Virginia Commonwealth

Univ., Richmond, VA, [4]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [5]SUNY Downstate Hlth.Sci. Univ.State Univ. of New York Downstate Hlth.Sci. Univ, Brooklyn, NY

Abstract:

Background: Polygenic Risk Scores (**PRS**) are essential for quantifying genetic predisposition to complex traits and diseases, including psychiatric disorders and associate with multitude phenotypes in target population. Traditionally, PRS calculations require individual-level genotype and phenotype data of the target population, posing challenges related to data privacy and accessibility. This study aims to develop and validate a method for approximating PRS effects using summary-level genome-wide association studies (**GWAS**) statistics derived in target population. This approach enhances the feasibility and applicability of PRS in large-scale genetic studies. **Methods:** We employed a comprehensive approach to approximate PRS effects at the summary statistic level, integrating data from GWAS and advanced computational models. **PRS Calculation from GWAS**: We derived the Z-score of PRS by aggregating the effects of multiple SNPs, each weighted by their association with the trait of interest. This process involved additional weighting using the Elastic-net method and correction based on their corresponding linkage disequilibrium (LD), calculated using large reference panels comprising 50,000 individuals. **Results. Accuracy:** The approximation of PRS effects from summary statistics demonstrated high concordance with traditional PRS association effects based on individual-level genotype phenotype data. The correlation coefficients (rho) between the approximated PRS and actual PRS effects were consistently high across different traits (i.e., major depressive disorder (MDD)-UKB, and suicide) and biomarkers (rho [0.81-0.99]), indicating robust performance. Notably, the difference between the "**GWAS-PRS**" based Z-scores and the actual PRS based Z-scores was minimal, especially when considering only nominally significant PRS effects. Additionally, there was good concordance (rho [0.89-0.98]) between the "***GWAS-PRS***" based Z-scores and LDSC **rg**-Z-scores, applied across traits and biomarkers. **Conclusion:** Our study presents a robust method for approximating the effect of PRS using summary-level GWAS statistics. This method enhances the feasibility of conducting large amount of PRS-based calculations in large-scale genetic studies and facilitates collaboration by addressing data restrictions and enabling inclusive research across countries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4019F: Artificial Intelligence for Unsupervised Population Phenotyping to Enhance Disease Genotype to Phenotype Association Discovery

Authors:

A. Masino, R. Baminiwatte; Clemson Univ., Clemson, SC

Abstract:

Molecular diagnosis for individuals with rare disease can improve disease management, facilitate access to public health services, and eliminate unnecessary diagnostic tests. Yet, despite increased genetic testing, diagnostic rates remain unsatisfactorily low (<30% depending on clinical domain). In large part, this is due to insufficient knowledge of genotype to phenotype association. Discovery of such associations through traditional methods that start with a disease phenotype and study a small number of individuals to identify causative variants is hampered by the presence of thousands of variants of uncertain significance in each person. Alternatively, researchers have begun to consider a gene-to-patient approach in which one first identifies a high-confidence in-silico predicted pathogenic variant and then identifies a cohort of individuals with that variant from a large database. This cohort is then assessed for the presence of a shared phenotype. However, determination of a shared phenotype within a cohort is difficult. Even when phenotypic data is restricted to ontology terms, it is unlikely that two very phenotypically similar individuals will be annotated with the exact same terms given ontology complexity. For example, the Human Phenotype Ontology (HPO) contains approximately 20,000 terms. Researchers adopting a gene-to-patient approach often rely on identifying the cohort phenotype based on the presence of exact term matches. However, this is likely to either underrepresent the entirety of the overall cohort phenotype or fail to identify one at all. Here, we present our initial efforts to develop artificial intelligence models to identify a shared phenotype within a cohort based on ontology annotations. In this method, we first develop computable representations of the ontology terms using deep leaning methods that utilize term text descriptions and ontology structure. We then apply a genetic search algorithm over the ontology to identify a representative phenotype that maximizes average similarity over the cohort. To validate the approach, we simulated over 45,000 individuals across 5,076 genetic variants using known gene-phenotype relationships from the HPO with varying degrees of imprecision (annotation by related term) and noise (annotations with unrelated terms). We found that the approach recovers the known phenotype with a high degree of accuracy across cohorts with highly variable term counts in the known phenotype, imprecision and noise suggesting it may benefit research using a gene-to-patient approach to discover new genotype to phenotype associations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4021F: Association of orofacial cleft risk loci with elevated risk of prostate cancer

Authors:

N. Mukhopadhyay[1], A. Cenname[2], **J. Shaffer**[3], M. Marazita[4]; [1]Univ Pittsburgh/Sch Dental Med, Pittsburgh, PA, [2]Univ Pittsburgh/Sch. of Med., Pittsburgh, PA, [3]Univ. of Pittsburgh, Pittsburgh, PA, [4]Ctr. for Craniofacial and Dental Genetics, Unive, Pittsburgh, PA

Abstract:

Several well-replicated genetic risk loci/variants for orofacial cleft congenital anomalies (OFCs) are also known risk loci for a variety of cancers: e.g., *CDH1* associated with gastric cancer and breast cancer; *AXIN2* with colorectal cancer; and *8q24* with prostate cancer, colorectal cancer, bladder cancer, and breast cancer. In this study, we formally investigate whether specific OFC risk loci are also associated with prostate cancer. We first We conducted a case-control association study of prostate cancer in All of Us subjects, focusing on genomic regions previously known to influence OFC risk. Unrelated males with a positive or negative diagnosis of "malignant tumor of the prostate" in the All of Us Controlled Tier Dataset v7 (Apr 20, 2023) were partitioned into two groups, European and African-derived (including African American) based on their predicted genetic ancestry. The European ancestry sample consisted of 3,094 cases and 45,197 controls, while the African-derived ancestry sample consisted of 536 cases and 19,165 controls. Genetic association was tested within each group utilizing logistic regression and genotype data for approximately 13,000 variants from the Illumina Global Diversity Array, using HAIL. We found strong associations to prostate cancer within the 8q24.21 OFC risk locus in both samples, as well as moderate associations in other OFC loci not known to be associated with prostate cancer risk. We observed a novel association to prostate cancer within the *WDR1* locus in the African-derived sample (rs75778012, p value 5.8e-07, beta coefficient 0.46). Following the candidate region association, we then tested for genetic correlation between OFC and prostate cancer using GWAS summary statistics from previously published OFC and prostate cancer studies. The OFC GWAS summary statistics were obtained from the Pittsburgh Orofacial Cleft Study, and the prostate cancer summary statistics from a multi-ancestry GWAS meta-analysis of European, African, East Asian and Hispanic male subjects. Our preliminary results from the case-control association of prostate cancer in OFC risk loci support a potential link between OFCs and prostate cancer. The OFC and prostate cancer multi-ancestry GWASs did not show a significant genome-wide correlation. We are currently exploring the shared genetics of OFCs and prostate cancer within specific ancestry-groups using LD score

regression and other techniques, such as conditional GWAS, that are designed to detect shared genetics across co-morbidities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4022F: Augmenting Therapeutic Target Discovery with Polygenic Risk Scores

Authors:

E. Bartell, M. Luth, G. McInnes, B. Cajes, D. Jakubosky, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Abstract:

Large-scale biobanks comprising genetic data linked with rich phenotypic information enable the evaluation of associations between sequence variants and thousands of phenotypes. These analyses can improve our understanding of the biological mechanisms underlying human health and disease, and can facilitate the identification of potential therapeutic targets. Rare nonsynonymous variants, particularly loss-of-function variants, are among the most useful naturally-occurring tools for understanding the phenotypic effects of pharmacological modulation of a target. Such variants typically have large effect sizes and may better approximate the effects of a drug intervention on a disease. However, the discovery of rare variants that are associated with disease is hampered by their reduced statistical power compared with common variants. Recent publications have proposed several approaches for employing common variant-derived polygenic risk scores (PRS) to enhance rare variant discovery, including population sub-sampling and covariate adjustment. However, which of these approaches is best suited for rare variant discovery in support of therapeutic target identification is currently unknown. We used out-of-sample and trained in-sample PRS to predict traits in the UK Biobank, focusing on circulating levels of apolipoprotein B, low-density lipoprotein, and triglycerides, and on coronary artery disease. We then performed genome-wide association studies (GWAS), by (i) adjusting for common-variant PRS, (ii) sub-sampling to individuals in defined PRS strata, and (iii) combining the two approaches. We evaluated the differences in signal identified through these approaches, focusing on associations with common and rare variants in known therapeutic targets, and associations that are likely induced by collider bias. Our findings reveal similarities and some notable differences between these approaches. In general, regardless of the approach used, incorporation of PRS tends to decrease signal for true positive common variant associations and increase signal for true positive rare variant associations. We additionally find that use of in-sample PRS in particular can result in

likely spurious rare variant signal that is proximal to true positive common variant associations, indicative of collider bias.Overall, this study demonstrates that incorporation of PRS can augment rare variant discovery for the purpose of therapeutic target identification, and highlights the potential benefits and risks of the main approaches. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4023F: Bayesian fine-mapping in two ancestry populations to identify putatively causal variants for cross-ancestry prediction of CYP2A6 activity

Authors:

A. Langlois[1,2], C. Lerman[3], J. G. Pouget[1], M. J. Chenoweth[1,2], R. F. Tyndale[1,2]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [3]Univ. of Southern California, Los Angeles, CA

Abstract:

CYP2A6 is the primary metabolic inactivator of nicotine. Its activity is variable, where higher activity is associated with increased smoking and risk of tobacco related diseases. CYP2A6 activity can be measured in vivo in regular smokers using the nicotine metabolite ratio (NMR: 3'-hydroxy-cotinine/cotinine); however, the NMR is not reliable in intermittent and non-smokers. A genetic predictor could overcome these limitations, and allow for accurate estimation of CYP2A6 activity in any datasets with genetic data. *CYP2A6* is polymorphic, with common alleles varying in frequency between populations. Genetic scores for CYP2A6 activity (i.e. NMR) have been derived, but are ancestry-specific with poor portability across populations. We used Bayesian fine-mapping to identify putatively causal *CYP2A6* variants separately in European- (EUR) and African-ancestry (AFR) populations, then cross-ancestry fine-mapping to further investigate these variants common to both populations.
Genotypes for n=933 EUR and n=954 AFR individuals were obtained by merging data from Illumina SNP array genotyping (and imputation), and Taqman and PCR assays of common *CYP2A6* star alleles. Fine-mapping used log-transformed NMR as the outcome; ancestry-specific analyses used the R package SuSiE, while cross-ancestry analyses used the Python tool SuShiE. Analyses were carried out in the overall samples, and in subsets with no known functionally impactful *CYP2A6* star alleles.
In EUR, SuSiE identified five credible sets of putatively causal variants: rs56113850, *CYP2A6*2*, *CYP2A6*9* (and two other correlated SNPs), *CYP2A6*12*,

and *CYP2A6*53*; among those with no known alleles only rs56113850 was identified. In AFR, four credible sets were identified: *CYP2A6*4* and three sets with SNPs of unknown significance; among those with no known alleles no causal variants were identified. SuShiE cross-ancestry analyses identified nine credible sets: *CYP2A6*9*, *CYP2A6*4*, *CYP2A6*53*, and six sets with SNPs of unknown significance. Among those with no known alleles, rs3875149 was the sole putatively causal variant.
Examinations of linkage disequilibrium patterns among CYP2A6 star alleles and SNPs identified in AFR and cross-ancestry fine-mapping suggest that, at least in part, linkage with star alleles drives their association with activity. Given its discovery among those with no known alleles, it is likely that rs3875149 represents a novel causal signal correlated with rs56113850 in EUR. Further investigation is needed to determine whether rs3875149 is causal, or tags a causal variant. This variant will be integrated into a future cross-ancestry genetic predictor of CYP2A6 activity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4024F: Benchmarking of single-cell eQTL methods with experimental data from DICE database

Authors:

C. Gonzalez-Colin[1], B. Schmiedel[2], T. Amariuta[1], P. Vijayanand[2]; [1]Univ. of California San Diego, San Diego, CA, [2]La Jolla Inst. for Immunology, La Jolla, CA

Abstract:

Improvements in single-cell technologies have enabled identification of cell types and cellular populations associated with diseases, developmental stages and multiple biological processes. By profiling millions of cells from multiple donors coupled with genotype data, single cell RNA-sequencing has proven to be a powerful tool to uncover new genes and cellular process linked to disease risk variants. While bulk RNA sequencing has been useful in detecting association with gene expression on common cell types, single cell RNA-sequencing provides a better resolution to resolve those, particularly for rare cell populations, which are challenging to isolate with conventional methods (e.g. using fluorescence-activated cell sorting - FACS). Several methods have been developed to directly map eQTL without aggregating the read counts across single cells of the same donor into 'pseudobulk' counts. However, the performance of these different methods has not been properly benchmarked on consistent reference datasets, nor has the difference in eGene (e.g., genes whose expression is associated with genetic variation) detection been compared between bulk, pseudobulk, and single cell

gene expression methods.
We are uniquely enabled to perform benchmarking across these different resolutions of gene expression measurement, as we have generated bulk, pseudobulk, and single cell gene expression profiles in the same set of 89 individuals from the DICE cohort. Here, we benchmark the following eQTL association tools (GASPACHO, CellRegMap, and SAIGE-QTL) with activated CD4+ T cell gene expression at bulk, pseudobulk, and single cell levels (&gt 60,000 cells). We use eGenes identified by bulk gene expression profiles as the gold standard to evaluate single cell eQTL methods according to scalability, usability and their ability to identify eGenes missed by the pseudobulk approach. By comparing the performance of these methods against bulk RNA-seq-derived eGenes, we aim to assess their ability to capture cell type-specific regulatory signals in homogeneous cell populations and uncover novel eQTL-gene associations that may be missed by previous 'pseudobulk' approaches. With larger single cell data sets slated to arrive in the future (such as the TenK10K resource by the authors of the OneK1K resource), this benchmarking study will provide valuable insights into the strengths and limitations of current and emerging single-cell eQTL mapping methods, highlighting the requirements necessary for a robust and scalable application to linking genetic variation mediated through gene expression with impacts on disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4025F: Beyond Bias: Advancing PCA for Genetic Research

Authors:

C. Trevino[1], A. Talbot[1], A. Kotlar[2], T. Wingo[1]; [1]Emory Univ., Atlanta, GA, [2]Emory Univ, Atlanta, GA

Abstract:

The large p, small n phenomenon, where the number of covariates exceeds the number of samples, is almost ubiquitous in genomics. Because of this, Principal Component Analysis (PCA) is commonly performed to reduce the dimensionality. Unfortunately, when the covariates are strongly influenced by high-variance confounders such as sex or ancestry, these variables dominate the latent representation. This can be problematic due to both fairness and scientific considerations. To address this, researchers commonly exclude the highest-variance principal components due to their strong correlation with undesirable dynamics. Fair machine learning has developed an alternative approach, where the latent representation is constrained to be orthogonal to the confounding variables. In PCA, this results in a modified eigendecomposition of the covariance matrix. However, it is known that the empirical covariance matrix performs poorly in the large p, small n regime. We combine recent developments in covariance matrix

estimation with a modified eigendecomposition to obtain fair representations requiring dramatically reduced sample sizes. We demonstrate the efficacy of our method on both synthetic data, and genetic data from the 1000 genomes project removing the effect of ancestry. Our results highlight the potential of this integrated approach to enhance the fairness and scientific validity of genomic data analyses by effectively disentangling relevant variance from confounding influences, thus offering a robust solution for high-dimensional, small sample size genetic studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4026F: BinBVS: Bayesian variable selection for classification of binary features, applied to multi-genic rare variant data from Phenome-wide association studies in UK Biobank

Authors:

D. Vitsios[1], L. Middleton[1], E. Oerton[1], M. Garg[1], Q. Wang[2], S. Petrovski[1]; [1]AstraZeneca, Cambridge, United Kingdom, [2]AstraZeneca, Chapel Hill, NC

Abstract:

Understanding the mulitgenic causal mechanisms underlying human complex disease is a long-standing problem in statistical genetics. Here we address this problem from the perspective of feature selection, leveraging gene-level collapsed rare-variant matrices generated from ~400,000 UK Biobank (UKB) whole exome sequences. Specifically, we quantify the importance of different genes when training a model to predict cases versus controls for a given phenotype. We developed a model (BinBVS) applicable to binary-valued data that can capture the combined effects of rare gene variants within a Bayesian framework and applied it to 2,569 binary traits from UKB. BinBVS can capture nonlinear dependencies between the binary genotype matrix and the class label for a given case-control cohort. Within the Bayesian framework, performed with either Markov chain Monte Carlo (Gibbs) sampling or Variational Bayes, we learn groups of genes that best explain the phenotype through a process known as Bayesian variable selection. Such an approach allows us to go beyond traditional univariate statistical tests that account for only a single gene at a time. We fit the model across thousands of phenotypes available in UKB, while additionally performing Fisher's exact test (FET) as a benchmark for univariate methods. Encouragingly, after calibrating both tests on an appropriate null distribution, we see a high degree of concordance between the two methods: 90.6% of all genome-wide significant associations ($p<10^{-8}$) are also captured by BinBVS at an equivalent Bayesian probability threshold of 0.95. Furthermore, we are able to uncover 29 putative novel associations in BinBVS

across multiple collapsing models, previously undetected by FET. Next, we leverage the inference scores learnt by BinBVS to construct multigenic signatures for each phenotype. Interestingly, we see that non-insulin dependent diabetes mellitus (ICD10: E11.9) ranks as the most multi-genic phenotype across all 2,569 phenotypes tested. Moreover, when collapsing multigenic signatures to a single feature (i.e. union) we see a significant increase in association levels compared to testing components of the groups individually ($p<10^{-20}$), thereby suggesting that BinBVS is capturing the combined effects of disease-relevant genes. We also validate the novel univariate and inferred multigenic signatures using an independent cohort from FinnGen and protein-protein interaction networks. Finally, BinBVS is scalable to biobank cohorts of scale even larger than UKB and can uncover putative novel associations that would otherwise be missed by traditional univariate phenome-wide association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4028F: Characterisation of diverse global ancestries within participants of the UK Biobank

Authors:

F. Pantring[1,2,3], G. Cavalleri[1,2,3], E. Gilbert[1,2]; [1]Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [2]The FutureNeuro Res. Ctr., Dublin, Ireland, [3]The SFI Ctr. for Res. Training in Genomics Data Sci., Galway, Ireland

Abstract:

The UK Biobank (UKB) is a large dataset containing in-depth phenotype and genotype data of 500,000 UK-based participants. To control for cryptic population genetic confounders, studies leveraging the UKB are typically restricted to a subset of the participants with homogenous European ancestry, according to self-identification and genotype-based principal component analysis. By analysing the 78,573 UKB participants with non-UK ancestries using population genetic approaches, there is an opportunity to better understand the extent to which UKB represents global genetic diversity and enable their inclusion in disease association studies. Here we characterise these diverse ancestries in the UKB by identifying the primary continental-level ancestry clusters and then fine-scale ancestry groups. To determine continental ancestries, an individuals' continental ancestry proportions were estimated using the ADMIXTURE algorithm by projecting the UKB samples onto ADMIXTURE reference populations from the 1000 Genomes and Human Genome Diversity Projects. The machine learning algorithm XGBoost was trained using ADMIXTURE data from reference populations and applied on the ADMIXTURE UKB results to assign each individual to one of eight continental ancestry clusters.

These continental clusters were further divided by applying community detection to a network of Identity-By-Descent sharing. We further characterised these sub-continental communities by demographic history, such as population size and isolation.
We find that the UKB is a repository of diverse ancestries primarily from Europe, Africa, and South Asia. Whilst capturing worldwide diversity, the 102 fine-scale ancestry groups appear to reflect the immigration history of Great Britain and its Commonwealth in the 20th century and are therefore not necessarily represented in other large genomic datasets. Such communities include South Asian ancestry communities born in Southeast Africa, but resident today in Britain, and one of the largest samples of Maltese genotypes available to date.
Our ancestry groups serve as a control for population history and facilitate the detection of rare functional variation in diverse ancestry groups - an example of which is the novel detection of one of the highest worldwide frequencies of the *MUC5B* promoter variant (rs35705950), a strong Idiopathic Pulmonary Fibrosis risk factor, in individuals of Maltese ancestry. Thus, the work provides a framework for future studies of health-related genetic variation specific to otherwise understudied genetic communities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4029F: Characterizing disease-critical contrastive gene programs in multi-context omics assays ⭐

Authors:

B. Turhan[1], D. Nandi[2], A. Ghosh[2], R. Mazumder[3], K. Dey[1]; [1]Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Indian Statistical Inst., Kolkata, India, [3]Massachusetts Inst. of Technology, Cambridge, MA

Abstract:

Genes exhibiting altered expression profiles between healthy and disease contexts have been found to be informative for disease heritability; therefore, it is important to delineate programs of genes that highlight these differences between two or more biological contexts in case-control single-cell RNA-seq assays. To this end, we propose ContraPC, a contrastive global embedding approach that identifies principal components that maximize the ratio of variation in the target (disease) context against that of the background (healthy) context.
We systematically compare ContraPC against the standard cPCA (Abid et al 2018 Nat Commun) which uses a difference-based approach in both simulation experiments and real data analysis. In simulations, across varying signal-to-noise ratios of the target distribution against background, ContraPC directions were on average 19.8-45.5% more cosine-similar to the

original contrastive direction compared to cPCA. We applied ContraPC to two publicly available case-control single-cell RNA-seq datasets, corresponding to Alzheimer's Disease (AD) (N=70,634, Donors=48) and Ulcerative Colitis (UC) (N=366,650, Donors=30). For both ContraPC and cPCA, we considered top genes driving top 5 contrastive principal components in each annotated cell type at 3 different levels of matched contrastive tuning. Out of 120 ContraPC gene programs generated from Alzheimer's Disease scRNA-seq data, 16 of them were found to be significantly associated with MAGMA AD disease association scores (p-value < 0.05), which was 14.3% higher than cPCA. Across 6 cell types in AD data, ContraPC programs were on average, 4.5x enriched in differentially expressed genes and 4.8x enriched in differentially variable genes between the disease and healthy contexts. Similarly, across 6 cell types, ContraPC gene programs showed relatively more unique contrastive profiles compared to cPCA (38.3% versus 37.5% in terms of correlation < 0.8 between the two methods). ContraPC gene programs in microglia were 6x enriched in differentially expressed genes in iPSC-derived microglia pre and post IFNβ treatment. Overall, ContraPC provides a new scalable approach to determine contrastive projections of cells in case-control omics data and assess gene programs driving these directions for matched complex disease heritability.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4030F: Co-expression-wide association studies implicate protein-protein interactions in complex disease risk ⭐

Authors:

M. M. Malakhov, W. Pan; Univ. of Minnesota, Minneapolis, MN

Abstract:

Background: Transcriptome-wide association studies (TWAS) have proven highly successful in prioritizing genes and proteins whose genetically regulated expression modulates disease risk. In standard TWAS approaches, each gene or protein is considered independently of the rest, which provides for a simple implementation but ignores possible synergistic effects among exposures. Although recent work has extended TWAS to a multi-exposure setting, no currently available methods explicitly model the genetic regulation of gene or protein co-expression. **Methods:** Here we introduce the co-expression-wide association study (COWAS) method to identify pairs of co-expressed genes or proteins that are associated with complex traits. COWAS first trains models to predict the expression of each functional unit in the pair, and then trains a third model to predict their genetically regulated co-expression, which we estimate as conditional correlation. Each model is evaluated on a held-out test set to ensure sufficient

predictive accuracy. The model weights are subsequently used to impute expression and co-expression levels into genome-wide association study (GWAS) summary-level data for any trait of interest, enabling COWAS to test for association between co-expression and the trait while also accounting for direct expression-trait effects.
Results: We applied our method to plasma proteomic concentrations from the UK Biobank Pharma Proteomics Project (N = 36k), focusing on 26,433 pairs with known protein-protein interactions comprised of 2,603 proteins coded by autosomal genes. We tested for association with low-density lipoprotein cholesterol, Alzheimer's disease, and Parkinson's disease, demonstrating that COWAS can successfully identify protein pairs whose co-expression impacts complex traits. Notably, our results demonstrate that co-expression between proteins may affect disease risk even if neither protein influences the disease when considered on its own. For example, COWAS discovered a significant effect of co-expression between PLAU and its receptor PLAUR on Alzheimer's disease, while standard TWAS performed on the same dataset failed to identify either protein.
Conclusion: Our contribution provides a novel framework for studying the effects of genetically regulated co-expression on complex traits, facilitating interrogation of the phenotypic consequences of gene-gene and protein-protein interactions using GWAS summary statistics.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4031F: Combining common and rare variants for disease risk prediction ⭐

Authors:

J. Zhang, S. Vrieze, E. Jorgenson, A. Ghosh, A. Marcketta, GHS-RGC DiscovEHR Collaboration, MAYO-RGC Project Generation, Colorado Center for Personalized Medicine - RGC Collaboration, UCLA-RGC ATLAS Collaboration, Penn Medicine BioBank (PMBB), Regeneron Genetics Center, G. Abecasis; Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

Over the past years, great progress has been made on the construction of polygenic risk scores (PRS) using common variants to predict disease risk. With the rapid development of sequencing-based assays and their application to large human genetic studies, rare variants are more accessible, presenting new opportunities for genetic risk prediction.
We investigate whether rare variants can improve common variant PRS and quantify this potential improvement. We use data from six cohorts, GHS, Mayo-clinic, COLORADO, UCLA, UPENN-PMBB, and UTHSC-Rooney, that include diverse ancestries (N = ~443K in total) and

electronic health records. We define case-control status across a range of diseases and health conditions using ICD-10 diagnoses codes. We then train common-variant PRS using genome-wide association studies (GWAS) results for each disease and also search for significant rare variant association signals using exome-wide association studies (ExWAS) results for the same samples.
We then proceed to evaluate and refine the PRS and rare variant signals using UK Biobank samples. First, to select rare variants that contribute to disease risk prediction independently of common variant PRS, we utilize backward selection in a logistic regression model on half of the UK Biobank samples (N=~230K). Then, we test a joint model (PRS + selected rare variants) on the remaining UK Biobank samples (N=~230K). We show that for 72 of 695 diseases and health conditions, there is a significant improvement in prediction when incorporating selected rare variants (P < 0.00005); however, the absolute improvement varies greatly by disease (1.3-fold increase on average and 4.4-fold maximum). For instance, for thalassemia, an inherited blood disorder, we find that the fraction of cases in the top 1% of individuals defined using common variant PRS only was 0.023 (and 25-fold increase relative to the entire sample) whereas this increases to 0.036 (a 1.6-fold further increase) when rare variants are included in the analysis. Similarly, for breast cancer, we find that the fraction of cases in the top 1% of individuals defined using common variant PRS only is 0.027 (and 2.2-fold increase relative to the entire sample) whereas this increases to 0.035 (a 1.3-fold further increase) when rare variants are included in the analysis. On the other hand, for many other traits the benefits of including rare variants are much smaller. Our analyses demonstrate that combining common variant PRS and rare variants can improve disease prediction and help identify high-risk individuals for a subset of diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4032F: Common Variant Gene Score Accurately Aggregates Genetic Effects for Alzheimer's Disease

Authors:

K. Cardone, R. Venkatesh, R. Kumar, Y. Bradford, A. Moore, M. A. Hall, M. Ritchie; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Alzheimer's disease (AD) is a burdensome neurodegenerative disease impacting 55 million individuals worldwide. There are no proven methods of preventative or curative treatment, so it is important to identify new therapeutic targets with innovative methods. One method is knowledge-driven multi-omics integration, integrating genomic data with transcriptome,

methylome, and proteome data using a machine learning (ML) model. One challenge is datatype differences, as genotype data is categorical whereas other omics data are quantitative. So, we developed common variant gene scores for AD, aggregating an individual's genetic risk into quantitative scores at the gene level. We constructed scores with summary statistics from the International Genomics of Alzheimer's Project (IGAP) GWAS meta-analysis of AD (n=54162). For each gene, we multiplied the IGAP betas by the additive genotype encoding, and then divided by the number of variants included. We generated these gene scores in two independent datasets: the Alzheimer's Disease Sequencing Project (ADSP) (n=28517) and the UK Biobank (UKBB) (n=7912). Next, we tested for association between each gene and AD with logistic regressions controlling for age, sex, and 4 principal components, defining significant (p &lt Bonferroni significance threshold) and suggestive genes (p &lt 1 divided by the number of tests). In ADSP, 7 gene scores were significantly associated with AD (p &lt 2.82e-6) and 8 were suggestive (p &lt 5.64e-5). 12 were previously associated with AD and 3 were novel (*CANX, HMGB3P22, TMEFF1*). In UKBB, 10 genes were significantly associated with AD (p &lt 4.03e-6) and 6 genes were suggestive (p &lt 8.05e-5), all of which were previously associated. 5 known genes replicated in both datasets. As this method replicated many known AD associations, we view these positive controls as evidence that gene scores can be used for multiomics integration. Gene scores also have utility in ML models because they reduce the number of features. So, to assess entire model performance, LASSO regression was conducted. In both datasets, the combination of all gene scores predicted AD status modestly (ADSP R^2 = 0.216, UKBB R^2 = 0.115) and prioritized known genes as important. ADSP gene scores also prioritized novel genes as important. The R^2 values are similar to that of polygenic scores and the model prioritizes known genes, demonstrating its reliability and utility. In summary, we aggregated genomic information at the gene level and accurately portrayed genetic risk in a quantitative manner. These gene scores will facilitate integration with other omics data, which we hypothesize will identify novel therapeutic targets for AD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4033F: Computationally efficient approaches for whole-genome quantile inference with related samples

Authors:

F. Wang; Columbia Univ., New York, NY

Abstract:

Quantile regression is an alternative to linear regression for quantitative phenotypes that enables the detection of heterogeneous genotype-phenotype associations. However, the computation of existing quantile regression methods is prohibitively slow for GWAS when modeling complex correlation structures and handling biobank-scale data, as the integration of random effects to calculate likelihood demands ultra high-dimensional integrations. We propose here novel and computationally efficient approaches to perform genome-wide quantile inference with related samples.
We introduce two novel tests: the REGENIE-based quantile rank score test (REGENIE.QRS) and the factored spectrally transformed weighted quantile rank score test (FaST-WQRS). The REGENIE.QRS test first estimates the polygenic effect using the existing REGENIE method within the context of whole-genome regression. It then performs quantile regression for estimation and the quantile rank score test (QRS) for testing after subtracting the polygenic effect. We compare the performance of REGENIE.QRS with the approach using the best linear unbiased predictions for estimating the polygenic effect. For FaST-WQRS, we consider the factored spectrally transformed quantile mixed model and develop the test based on the weighted quantile rank score test by extending it to accommodate more general heteroscedasticity.
Our algorithms can analyze data for over 400,000 individuals in just a few hours, whereas current linear quantile mixed model algorithms fail with data for even 10,000 individuals. Through comprehensive simulations, we validate our approach's robust control of type-1 error rates and demonstrate its power advantage under models with heterogeneous effects, non-additive effects, gene-environment interactions, and heavy-tailed distributions. Moreover, we demonstrate its utility through applications to datasets from the UK Biobank and the ProgeNIA/Sardinia project, underscoring its ability to uncover complex genetic associations in diverse populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4034F: Construction of skin aging score using machine learning model for discovery of skin aging genes

Authors:

M. Yuk[1], S. Yang[1], K. Oh[1], J. Youn[1], H-Y. Yoo[2], S-H. Park[2], D-J. Son[1], N. Song[1]; [1]Chungbuk Natl. Univ., Cheongju, Korea, Republic of, [2]Kolmar Korea Co., Ltd., Seoul, Korea, Republic of

Abstract:

BACKGROUND: Skin aging is considered as a consequence of biological aging. However, skin changes with age varies from person to person even within the same age bracket implying

potential inter-individual variability. In this study, we developed a prediction model for skin aging score based on the five skin phenotypes and conducted genome-wide association study (GWAS) to identify skin aging susceptibility genes. **METHODS:** A total of 749 Korean women aged 30 to 50 years were included in this study. Five aging-related skin phenotypes (i.e., melanin, gloss, hydration, wrinkle, and elasticity) were used to construct skin aging assessment score adjusted for individuals' survey-based lifestyle factors. Machine learning approaches (i.e., elastic net, artificial neural network, extreme gradient boosting, support vector regression, and k-nearest neighbor regression) were considered for candidate model construction of skin aging score. Subsequent GWAS for skin aging score and post-GWAS functional analysis were conducted using a multivariate linear regression. **RESULTS:** We developed a skin aging prediction model using five skin phenotypes (melanin, gloss, hydration, wrinkle, and elasticity) and six lifestyle factors significantly related to skin aging (alcohol consumption, sleep habits, ultraviolet exposure, regular defecation, and the use of sunscreen and basic skincare products). According to the comparison of candidate prediction model performances, elastic net-based skin aging prediction model was chosen to construct skin aging score (RMSE=5.19, MAE=4.33, R^2=0.13). As a result of subsequent GWAS, the strongest association with skin aging score was observed in variants within genetic loci 4q22.1 (Beta (SE)=1.22 (0.17), $P=1.22\times10^{-12}$). Single-tissue quantitative trait loci analysis showed the strong association in two skin tissues regardless of exposure to the Sun ($P<5\times10^{-20}$). **CONCLUSION:** This study developed a novel skin aging prediction model with generating skin aging score and subsequent GWAS was conducted to discover skin-aging susceptibility genes. Developed skin aging score could be helpful to evaluate anti-aging effects of cosmetic or dermatologic intervention in research and clinical practice.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4035F: Convergent functional impact of common polygenic risk for neuropsychiatric disorders (NPDs) in genes associated with NPDs through rare, disruptive variation

Authors:

S. Kuo[1,2]; [1]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA

Abstract:

Exome sequencing studies have identified hundreds of genes associated with neuropsychiatric disorders (NPD) including intellectual disability (ID), autism (ASD), and schizophrenia (SCZ), almost all of which are intolerant of disruptive variation. Observed ultra-rare disruptive variation

in NPD-associated genes is, to this point, mostly expression decreasing. Far less is known about the expression consequences of NPD-associated common polygenic variation within these genes. We aimed to characterize the expression effects of common polygenic risk associated with ID, ASD, and SCZ within NPD genes. We assessed paired genotype and brain-derived gene expression data from 544 adult postmortem tissue donors collected by the Harvard Brain Tissue Resource Center and CommonMind Consortium. 400 autosomal genes associated with ID, ASD, and/or SCZ were compiled from recent and ongoing exome sequencing studies and analyzed together as a set of NPD genes. For each NPD gene in each donor, we computed a gene polygenic score (gPS) based on common polygenic risk for lower-IQ, ASD, and SCZ within the gene body, and regressed normalized gene expression on gPS. We observed that each of the three gPS scores was associated with non-directional inflation of gene expression variation relative to a permuted null. Thus, while observed ultra-rare variants in NPD genes often decrease expression, common risk variants in the same genes may subtly increase or decrease expression. Per-gene expression effects of IQ-increasing common variants in NPD genes were negatively correlated with per-gene expression effects of SCZ common variants ($r = -0.34$, $p = 8.04e-7$). This means that, on average, in genes in which expression increases are associated with higher IQ, expression decreases are associated with SCZ. In genes in which expression decreases are associated with higher IQ, expression increases are on average associated with SCZ. Across genes identified with NPDs through rare variant associations, common polygenic variation creates risk through gene-dependent expression changes and bidirectional gene expression dysregulation. Naturally occuring polygenic variation can clarify the link between NPD genes and neuropsychiatric disease and highlight areas of functional convergence.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4036F: DAESC-GPU: A GPU-powered scalable software for single-cell allele-specific expression analysis

Authors:

G. Qi, T. Cui; Univ. of Washington, Seattle, WA

Abstract:

Background. Single-cell allele-specific expression (ASE) is a powerful signal to study cis-regulatory effects at the resolution of individual cells. Differential ASE across cell types, continuous cell states, and experimental conditions can reveal context-specific regulatory effects. However, only a limited number of statistical tools are available to conduct single-cell differential ASE analysis. We previously developed DAESC, a method based on beta-binomial

regression. DAESC was shown to be robust and powerful in simulation studies and real data analysis. However, the computational efficiency of the DAESC R package is suboptimal, making it challenging to analyze datasets with a large number of cells (e.g., >1 million). This limitation becomes increasingly important with the rapid generation of allele-specific data from new sequencing technologies. **Methods.** We developed DAESC-GPU, an accelerated, scalable version of DAESC powered by Graphics Processing Units (GPUs). GPUs contain thousands of cores which are highly optimized for large-scale parallel operations. DAESC-GPU utilizes CuPy, a Python library that leverages NVIDIA GPUs to provide significantly faster performance compared to traditional CPU-based Python libraries or R packages. Our software can be executed on publicly available GPUs on Google Colab at no cost, or more powerful GPUs for enhanced performance. **Results.** We compared the performance of DAESC-GPU with DAESC (the original R package). Tests for DAESC-GPU were conducted on an NVIDIA T4 or A100 GPU from Google Colab, while tests for DAESC was conducted on a 10-core CPU of an Apple M3 chip. We analyzed single-cell ASE data collected from a three-day endoderm differentiation experiment by Cuomo et al. The dataset consisted of 4,173 genes and 30,470 cells after pre-processing. DAESC was estimated to need ~135 minutes to complete a differential ASE analysis along pseudotime for all 4,173 genes. In contrast, DAESC-GPU completed the same analysis in 570s (9.5min) on a free T4 GPU, or 137s (2.3min) on an A100 GPU. The two software packages led to largely identical results, with ignorable differences likely caused by computational precision. **Conclusions**. DAESC-GPU is a high-performance software which is >10 times faster than the original R package. In the near future, we will use this software to analyze large-scale ASE data of over a million cells, and ASE from long-read sequencing data. DAESC-GPU will be a powerful tool to facilitate the discovery of biological insights from the rapidly growing allele-specific datasets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4037F: Data-driven phenotyping to optimize human facial traits in GWAS and rare variant analysis

Authors:

M. Yuan[1], S. Goovaerts[1], M. Lee[2], H. Hoskens[3,1], S. Richmond[4], S. Walsh[5], M. Shriver[6], J. Shaffer[2], M. Marazita[2], S. Weinberg[2], H. Peeters[1], P. Claes[1,7]; [1]KU Leuven, Leuven, Belgium, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Univ. of Calgary, Calgary, AB, Canada, [4]Cardiff Univ., Cardiff, United Kingdom, [5]Indiana Univ. Indianapolis, Indianapolis, IN, [6]Penn State Univ, University Park, PA, [7]Murdoch Children's Res. Inst., Melbourne, Australia

Abstract:

Genotype-phenotype association analyses for morphological traits, such as human facial phenotypes, typically utilize predetermined and simplified anthropometric measures that are based on biological or clinical knowledge, or use features derived from unsupervised dimension reduction techniques such as principal component analysis (PCA). However, these approaches may not fully capture the complexity or multidimensionality of facial morphology, nor do they necessarily reflect the most relevant aspects of variation for genetic investigations. Therefore, we introduce a data-driven framework that can optimize phenotype extraction for genome-wide association studies (GWAS, for common variants), as well as rare variant analyses. This phenotyping framework comprises two steps: 1) constructing a feature space that captures shape variations through PCA; and 2) within this feature space, identifying genetic-focused directions via an optimization algorithm. In GWAS, we search for phenotypes exhibiting high heritability, estimated from family data or genomic relationship matrices in unrelated individuals. In rare variant analysis, we optimize for the skewness of phenotype distributions, as a commingled distribution provides preliminary evidence for a single (or few) genetic locus with a major effect. We compared our proposed heritability-enriched phenotypes and commingling-based skewed phenotypes against PCA in a GWAS involving 8,426 individuals of European ancestry and in gene-based tests of rare variants (MAF<1%) across the whole exome of 1,906 individuals. Heritability-enriched phenotypes exhibit the highest SNP-heritability as estimated by LD score regression from GWAS, followed by PCs, while commingling-based skewed traits display the lowest SNP-heritability. Regarding the GWAS discovery rate, heritability-enriched phenotypes identify the same number of independent genetic loci with a smaller number of effective traits compared to PCs. For the analysis of rare variants, commingling-based skewed traits result in more genes passing the exome-wide significance threshold than PCA, while heritability-enriched phenotypes lead to very few associations. Overall, our results demonstrate that data-driven phenotyping allows for the extraction of genetically relevant phenotypes that can specifically enhance discovery efforts of common and/or rare variants, as evidenced by their increased power in GWAS and rare variant analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4038F: Deep learning predicts TF binding variation, illuminating breast and ovarian cancer biology

Authors:

S. Salazar[1,2], T. Adeluwa[3], S. Sumner[3], S. Gona[4], F. Nyasimi[3], R. Madduri[5], J. Schildkraut[6], H. Im[4]; [1]Univ. Natl. Autonóma de México, Querétaro, Mexico, [2]Escuela Natl. de Estudios Superiores

Unidad Juriquilla, UNAM, Querétaro, Mexico, [3]The Univ. of Chicago, Chicago, IL, [4]Univ. of Chicago, Chicago, IL, [5]Argonne Natl. Lab., Naperville, IL, [6]Emory Univ., Atlanta, GA

Abstract:

The study of breast and ovarian cancer (BC, OC) genetic susceptibility through genome wide association studies (GWAS) has revealed a large number of associated variants located in regulatory regions. These variants could plausibly disrupt transcription factor (TF) binding motifs, potentially impacting gene expression levels and overall disease risk. Yet, establishing a link between TF binding and disease risk variants is a challenging task as it requires tissue-relevant, population-level TF-binding specificity data, which is scarce and costly to generate.
In this study, we addressed this problem by implementing *TFXcan*, a deep learning-based method that allows us to make associations between predicted TFr binding and GWAS loci. We made binding prediction models for a total of 76 transcription factors in different tissues: 29 for cervix epithelium, 45 for breast epithelium and mammary gland, one for ovary, and one for uterus. We applied these models to the Breast Cancer Association Consortium's 2020 GWAS and the Ovarian Cancer Association Consortium's 2022 GWAS data.
We inspected 212 BC and OC loci where we could identify SNPs with high probability of being causal using the susieR approach. Our preliminary results indicate that for most of the loci neither motif disruption nor chromatin accessibility were the drivers of the association. However, we did find that for 30 loci, variation in TF binding across individuals was strongly associated with BC and OC risk. In these loci, multiple TF were found to be associated, suggesting a complex collaborative regulation of TF binding.
Taken together, our results show that *TFXcan* and related, deep learning tools that predict epigenetic features and TF binding open exciting avenues for further research into the complex regulatory interactions that impact disease susceptibility.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4039F: Detecting meiotic non-crossovers recombination in small pedigrees

Authors:

G. Palsson, M. T. Hardarson, B. V. Halldorsson; deCODE genetics, Reykjavik, Iceland

Abstract:

Meiotic recombination mixes up the genetic material that parents pass on to their offspring. Recombination generates both crossover and non-crossover throughout the genome, and while

crossovers have been well studied in the past, non-crossovers are underreported due to the difficulty in detecting them. We have previously explored non-crossovers in pedigrees using three-generation sequenced families. The requirement of having three generations of sequenced family members restricts substantially the number of probands available and here we show that we can greatly increase the size of our datasets by using whole genome sequenced families with two or more children, omitting the requirement for sequenced grandchildren. We describe an algorithm for detecting non-crossover events and present methods for validating results using both details from the sequence data as well as family based long-range phasing that takes advantage of the genealogy of the Icelandic cohort. We show that our method is consistent with earlier methodology and is able to detect a larger number of events.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4040F: Deviations from additivity in *APOE4*-associated Alzheimer's disease ⭐

Authors:

R. Durodoye[1], T. Ciesielski[1], P. Benchek[1], J. Bartlett[1], X. Zhu[1], M. Pericak-Vance[2], G. Tosto[3], ADGC, J. Haines[1], F. Rajabli[4], W. Bush[1], S. Williams[1]; [1]Case Western Reserve Univ, Cleveland, OH, [2]Univ. of Miami Miller Sch. of Med, Miami, FL, [3]Columbia Univ., New York, NY, [4]Univ. of Miami, Miami, FL

Abstract:

Background: *APOE*'s ε4 haplotype (*APOE4*) is the strongest genetic predictor for late onset Alzheimer's disease (LOAD). *APOE4*'s conferred risk varies by race and ethnicity (R/E), but the mechanisms for this variation remain poorly described. Virtually all recent publications model *APOE4* risk additively, while very few consider deviations from additivity (DA), let alone R/E-specific DA. We sought to describe how these deviations impact the association between *APOE4* and LOAD, and to what extent they affect risk heterogeneity among R/E groups. **Method:** We analyzed 44,366 participants from the Alzheimer's Disease Genetic Consortium: 3,263 East Asian (EAS), 1,922 genetically admixed Hispanic/Latino (HIS), 7,663 non-Hispanic Black (NHB), and 31,518 non-Hispanic White (NHW). We constructed two logistic regression approaches: I) separate, R/E-stratified models and II) a joint, multi-R/E model. All models adjusted for *APOE4* dosage, age, sex, study site, genetic ancestry, and a novel adjustment quantifying DA. The multi-R/E model also adjusted for R/E identity and generated estimates for R/E**APOE4* and R/E*DA interactions.

Result: Consistent with prior studies, we found that additive *APOE4* risk estimates varied across R/E groups. Following DA adjustment, we observed a significant decrease ($p<0.05$) in *APOE4*-associated risk among NHW participants (OR_{APOE4}: 3.34), and significant increases ($p<0.05$) among EAS (OR_{APOE4}: 6.18), HIS (OR_{APOE4}: 3.54), and NHB (OR_{APOE4}: 3.08) participants compared to their respective additive modeling estimates. We observed no difference in *APOE4* risk between NHB and NHW participants after adjusting for DA. Interestingly, *APOE4*-associated risk among NHB individuals carrying a single copy of *APOE4* was closer to estimates for those carrying no *APOE4* copies while risk for NHW *APOE4* heterozygotes trended toward homozygous *APOE4* estimates. DA were also significant in these populations and differed between them. In the multi-R/E context, only R/E**APOE4* interactions involving EAS participants reached significance. Finally, we detected significant R/E*DA interaction in NHB and NHW participants.
Conclusion: Our results indicate that additive genetic models incorrectly estimate *APOE4*'s contribution to LOAD risk, especially in non-White groups. Adding considerations for DA improved our modeling of the *APOE4*-LOAD association. This adjustment may, in turn contribute to more accurate genome-wide association study predictions and contribute to a better understanding of *APOE4*-associated LOAD risk across all R/E populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4041F: Discovery of rare coding variants with loss of function effect using a sliding window technique on clinicogenomic data

Authors:

E. Cirulli[1], A. Bolze[1], K. M. Schiabor Barrett[1], N. Telis[1], M. Levy[1], D. Stoller[2], C. N. Chapman[3], A. A. Chahal[4], D. P. Judge[5], O. A. Douglas[6], J. J. Grzymski[7], W. Lee[1], N. L. Washington[1]; [1]Helix, San Mateo, CA, [2]Univ. of Nebraska Med. Ctr., Omaha, NE, [3]St. Luke's Univ. Hlth.Network, Bethlehem, PA, [4]Wellspan Hlth., Ephrata, PA, [5]Med. Univ. of South Carolina, Charleston, SC, [6]Hlth.Partners, Minneapolis, MN, [7]Univ. of Nevada, Reno Sch. of Med., Reno, NV

Abstract:

Introduction: When analyzing the effects of rare genetic variants on phenotypes, similar variants are often grouped together to improve statistical power. However, this may combine variants with different effects, canceling out the signal. The established 3D Power Window (3DPW) method utilizes a sliding window to group variants based on their position within the 3D structure of the protein to reveal regions where rare variants have a similar effect.
Objective: To identify areas within the 3D structures of proteins in which rare missense variants

cause a loss of function (LoF) effect.
Methods: We first perform gene-based collapsing analysis of rare LoF variants and, separately, missense (coding) variants in 470k UKB exomes for 2,772 phenotypes. We identify 574 associations (160 genes, 251 phenotypes) where the LoF model has a significant association but the coding model does not. We take these gene/phenotype combinations as candidates for 3DPW to identify regions of the protein in which missense variants have a LoF effect. We then train 3DPW models using predicted structures from AlphaFold and 350k UKB exomes. We test our models in 117k unrelated UKB samples and 105k samples from US-based cohorts who underwent clinical-grade sequencing.
Results: Of the 160 genes with significant LoF models but not significant whole-gene coding models, 19% produced significant 3DPW coding models in the test sets, implicating amino acid changes in specific regions of the gene. These models included between 0.6% and 5.0% of the gene, with effect sizes that were a median of 116% of the LoF effect sizes in magnitude. For example, LoF variants in *EPB41* are known to cause the rare condition of hereditary elliptocytosis and are significantly associated with reticulocyte percentage in UKB (p=5.5e-18, normalized beta 0.6). We show that individuals with missense variants in the model regions, which are 2.1% of all rare missense carriers in this gene, also have significantly increased reticulocyte percentage in the test set, with a similar effect size (beta=0.47) despite a lack of overall signal for coding variants in this gene (beta=-0.03).
Conclusion: 3DPW can extend known LoF associations to missense variants by identifying the regions of the protein in which amino acid changes have a LoF effect. We previously demonstrated this for genes with significant whole-gene signals and here show that it can be successfully applied for discovery in genes that do not show significant associations with missense variants at the gene level. This important development moves the field forward in its ability to interpret the effects of missense variants without established LoF properties.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4042F: Disentangling chronic pain genetic heterogeneity from UK Biobank self-reports using graph-embedded topic model

Authors:

H. Tsao[1,2], Y. Wang[3], G. K. Farhood[2,4], Y. Li[3], A. V. Grant[2,4,5]; [1]Dept. of Human Genetics, Faculty of Med. and Hlth.Sci., McGill Univ., Montreal, QC, Canada, [2]Alan Edwards Ctr. for Res. on Pain, McGill Univ., Montreal, QC, Canada, [3]Sch. of Computer Sci., McGill Univ., Montreal, QC, Canada, [4]Dept. of Anesthesia, Faculty of Med. and Hlth.Sci., McGill Univ., Montreal, QC, Canada, [5]Faculty of Dental Med. and Oral Hlth.Sci., McGill Univ., Montreal, QC, Canada

Abstract:

Chronic pain, defined as pain experience persisting for three or more months, is a complex health problem displaying diverse phenotypic expressions including musculoskeletal localization. Individuals with chronic pain often report comorbidities, most commonly from among psychiatric disorders, cardiovascular and autoimmune diseases. In investigating the genetic basis for chronic pain, however, standard genome-wide association (GWA) studies focus on a single phenotype. Topic modeling is a class of machine-learning algorithms originally conceived for document corpuses and adapted to biobank study populations with extensive clinical data, able to detect patterns of comorbidities. The method has previously identified novel genetic loci beyond those using single-condition labels. We recently developed Graph-Embedded Topic Modeling (GETM), which integrates biomedical ontology hierarchies into pre-trained graph embeddings (Wang et al., iScience 2022). We focused specifically on self-reported medical conditions and medications as recorded in the UK Biobank, comprising 259,128 individuals of European descent with at least one condition or medication. Topic modeling assumes each document (or study subject) is a mixture of topics, and each topic is a mixture of words (or medical conditions and medications). Beyond medical conditions and medications, we drew on hospital ICD-10 codes and psychosocial variables along with their respective graph ontologies. We derived topic mixtures for the UK Biobank study population to obtain individuals by topics (θ). For our investigation of genetic heterogeneity of chronic musculoskeletal pain, we selected the most predictive topics for chronic musculoskeletal pain (CMSKP) and used these as quantitative phenotypes. Moreover, these were combined with CMSKP status: we contrasted the top 50% of individuals based on θ values among CMSKP cases to the bottom 50% of individuals based on θ values among CMSKP controls in a series of GWA scans for risk topics, and the inverse for protective topics. We used REGENIE to run GWA scans adjusting for age and sex, and Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) for post-GWA scan analysis, including functional annotations. Our findings provide insights into the heterogeneous components of CMSKP etiology based on differential distributions of comorbidities and related variables.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4043F: Efficient large scale PheWAS analysis with PheTK

Authors:

T. Tran[1], D. Schlueter[2], H. Taylor[1,3], C. Zeng[1], H. Mo[1], R. Carroll[4], J. Denny[1,5]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Univ. of Cambridge,

Cambridge, United Kingdom, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]All of Us Res. Program, Bethesda, MD

Abstract:

Phenome-wide association study (PheWAS) is an analysis method to study gene-phenotype associations utilizing genetic and longitudinal electronic medical record data. Several previous packages have been developed for PheWAS analysis; however, they do not always scale well to the size of modern cloud-based biobanks and may not easily support the most recent phecode mappings. We describe PheTK as a python package designed to support both Phecode 1.2 and PhecodeX 1.0 while being capable of handling biobank-scale data efficiently.
PheTK supports multithreading and multiprocessing, is optimized for working with large datasets, and supports logistic and Cox regression. It supports the end-to-end standard PheWAS workflow, from creating a cohort to regression analysis and result visualization with Manhattan plot. It simplifies cohort generation by genomic variant through support for Hail matrix table and covariate generation from OMOP data are supported for All of Us researcher workbench (All of Us). PheTK also supports Cox proportional hazard regression to analyze time to phecode event data.
To demonstrate the package, we performed PheWAS for CFTR F508del variant in All of Us and the UK Biobank. Our evaluation includes a benchmarking test comparing PheTK with the R PheWAS package, which also supports multithreading. We assessed PheTK's performance across various cloud and local environment, utilizing both Phecode 1.2 and PhecodeX. To demonstrate PheTK scalability in large studies, we ran >10,000 PheWAS analyses on All of Us to find protein - clinical phenotype associations utilizing OmicsPred scores.
We used the All of Us CFTR cohort of 181,776 participants as input data in PheWAS with phecode 1.2 for both the genetic approach demonstration and the benchmark test. In the benchmark test, using the same virtual machine (VM) settings, PheTK took 64% less time than the R PheWAS package (9.9 minutes compared to 27.3 minutes). With the same approach, PheWAS with phecode 1.2 for a UKB CFTR cohort of 469,007 participants completed in 11.9 minutes. PheWAS with phecodeX for the UKB 469,007 CFTR cohort also completed successfully in the local computing resource test using a MacBook Pro M2 with 12 CPUs 32GB RAM. In the large-scale study, 10,732 PheWAS runs were completed in approximately 12 hours for a total cost of less than $250 (~$0.022 per PheWAS run) utilizing 4CPU 8GB RAM cloud VMs vs. initial estimates ~$16,000 using original test methods.
We intend to continue to develop PheTK as a biobank-scale tool to simplify EHR phenotyping. PheTK is available on GitHub (https://github.com/nhgritctran/PheTK) and via an All of Us tutorial workspace.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4044F: Efficient Penalized Regression Approach for Admixture-Aware Polygenic Risk Scores

Authors:

F. Ockerman[1], B. Chen[1], Q. Sun[1], E. Kharitonova[1], L. Zhou[2], L. Raffield[1], R. Tao[3], Y. Li[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Indiana Univ. Sch. of Med., Indianapolis, IN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Polygenic risk scores (PRS) have promising clinical applications for risk stratification, disease screening, and personalized medicine. However, most PRS are trained on predominantly European ancestry cohorts and have limited transportability to external populations. While cross-population PRS methods have demonstrated greater generalizability than single-ancestry PRS, they fail to properly account for individuals with recent admixture between continental ancestry groups. To help overcome this gap, our group recently proposed GAUDI, a PRS method which leverages local ancestry inference to allow and estimate ancestry-specific effects while penalizing ancestry-differential effects. However, the fused lasso approach used by GAUDI is computationally expensive and does not easily accommodate more than two-way admixture scenarios. To address these limitations, we introduce HAUDI, an efficient lasso framework for admixed PRS. HAUDI re-parameterizes the GAUDI model as a standard lasso problem, allowing for extension to multi-way admixture settings and far superior computational speed than GAUDI.

In simulated data, HAUDI performs better than standard lasso (ancestry-agnostic) PRS under a variety of genetic architectures (median relative increase in R^2 of 25.4%). Compared to GAUDI, HAUDI has favorable performance when causal variants are only common in one population (median relative increase in R^2 of 6.3%) and only slightly reduced performance when causal variants are common in both populations (median relative decrease in R^2 of 1.2%). Using real data from African American (n=17,301) and Hispanic/Latino (n=11,777) individuals, we conducted local ancestry inference using FLARE and fit HAUDI, GAUDI, and ancestry-agnostic models for a variety of clinical phenotypes. Among phenotypes with $R^2 > 0.01$ for at least one model, HAUDI out-performs GAUDI (median relative increase in R^2 of 4.3%) in African American participants. HAUDI also has non-inferior performance compared to the ancestry-agnostic approach among African American and Hispanic/Latino participants (median relative increases in R^2 of 0% and 5.0%). These results demonstrate that HAUDI is a highly scalable PRS method which effectively models local ancestry for individuals with recent admixture.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4045F: EIR-auto-GP: A toolkit for deep learning based genomic prediction and analysis

Authors:

A. Sigurdsson[1], J. Gräf[2], Z. Yang[3], K. Ravn[4], J. Meisner[5], J. Meisner[6], R. Thielemann[7], H. Webel[1], R. Smit[8], B. Vilhjalmsson[9], B. Neale[10], A. Ganna[11], T. Hansen[12], R. Loos[7], S. Rasmussen[13]; [1]CBMR, Copenhagen Univ., Copenhagen, Denmark, [2]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Copenhagen, Denmark, [3]Univ. of Helsinki, Helsinki, Finland, [4]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, D, Copenhagen, Denmark, [56]Copenhagen Univ. Hosp., Roedovre, Denmark, [7]Univ. of Copenhagen, Copenhagen, Denmark, [8]Icahn Sch. of Med. at Mount Sinai, New York, NY, [9]Aarhus Univ., Aarhus, Denmark, [10]Massachusetts Gen. Hosp., Boston, MA, [11]Inst. for Molecular Med. Finland, Helsinki, Finland, [12]Univ. of Copenhagen, København N, Copenhagen, Denmark, [13]CBMR, Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

With larger and more precisely measured cohorts, the ability to identify non-linear genetic effects becomes increasingly likely. Deep learning offers a powerful approach to uncover these complex relationships. Here, we present EIR-auto-GP, a toolkit designed to automate genomic prediction with deep learning (DL), supporting both quantitative and binary target phenotypes. The toolkit streamlines the process of processing raw genotype data and associated covariates, performing deep learning modeling and model evaluation, all with a single command. Additionally, the framework includes various built-in variant pre-filtering approaches, allowing for training on large-scale genetic data directly on CPUs. For example, training DL models across 10 cross-validation runs for 400K SNPs and 35K individuals completed in around 3 hours on a 16-core computer. Beyond DL model training, the toolkit includes functionality to analyze the presence and drivers of non-linear effects in the data. This includes fitting both linear and non-linear models on different transformations of the input data to examine whether a non-linear performance gain is due to complex effects in covariates, genetic dominance effects, or potential epistasis between variants. The toolkit also includes fitting ordinary-least-squares (OLS) models on GxG and GxE terms to gain better insights into the factors driving the performance of the DL models. Here, we demonstrate the use of the toolkit in genetic research by identifying complex effects for 171 out of 2,922 measured blood plasma protein levels among 57,200 individuals in the UK Biobank. We show that it identifies known and potentially novel epistatic effects between variants, as well as dominance and complex covariate effects regulating the measured protein levels. We believe that the computational efficiency and ease of use offered by EIR-auto-GP will lower the barrier of entry for other researchers wanting to

perform DL-based genomic prediction. The software is available at https://github.com/arnor-sigurdsson/EIR-auto-GP.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4046F: Empowering Researchers: A Cost-Effective and Versatile PRS Calculator for the All of Us Program

Authors:

A. Khattab, S-F. Chen, N. Wineinger, A. Torkamani; Scripps Res. Translational Inst., La Jolla, CA

Abstract:

Introduction The utilization of Whole Genome Sequencing (WGS) data has revolutionized genomic research. The All of Us Research Program by the NIH aims to accelerate health research and medical breakthroughs by providing a comprehensive resource for researchers. In line with this objective, we have developed a Polygenic Risk Score (PRS) calculator tool that is both cost-effective and versatile. This tool is designed to facilitate researchers, regardless of their experience level, in leveraging WGS data for significant discoveries. **Methods** Our tool queries the complete callset in the Variant Annotation Table (VAT) to determine which variants in the score are present and then uses these variants to calculate the PRS from the sparse Hail format VDS table. This format is less dense, faster, and more cost-effective compared to other tools like PLINK, which require retrieving data from the cloud, a process that is both time-consuming and expensive. **Results** Our step-by-step approach enables researchers, especially those new to the field, to calculate any number of PRS for all individuals in the All of Us study at a fraction of the cost of other tools. For a score comprising approximately one million SNPs, the cost is about $33 to calculate for ~200,000 individuals using traditional methods. In contrast, using our approach, the same score calculation costs only $2, representing a 94% cost reduction. This significant cost difference underscores the efficiency and affordability of our tool. Additionally, our PRS calculator has demonstrated robust performance in generating accurate risk scores across various complex traits and diseases. The tool's results show a high correlation between the generated PRS and established clinical outcomes, validating its utility in genomic research. Furthermore, the reduced computational resources and time required for analysis enhance its practicality and accessibility for a wide range of research applications. **Conclusion** The PRS calculator tool represents a significant advancement in the field of genomic research. By providing a cost-effective, versatile, and user-friendly solution, it empowers researchers to harness the full potential of the All of Us WGS data. This tool not only democratizes access to

sophisticated genomic analysis but also paves the way for novel discoveries that can ultimately contribute to personalized medicine and improved public health outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4047F: Enhancing gene burden analyses in the UK Biobank with proteomic data

Authors:

D. Jakubosky, E. Bartell, R. Avila, Z. Romer, J. Chen, B. Cajes, G. McInnes, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Abstract:

Large-scale biobanks comprising genetic data linked to rich phenotypic information, such as the UK Biobank, provide an opportunity to leverage human genetics for the discovery of therapeutic targets. Rare nonsynonymous variants are potentially powerful tools for predicting the phenotypic effects of pharmacological inhibition; however, these variants are often too rare to be tested individually. Gene burden tests have therefore become a commonly utilized method for rare variant analysis, involving aggregating carriers of predicted-deleterious nonsynonymous variants in a given gene. While gene burden tests have proven to be a useful approach, they are nonetheless limited by our ability to accurately predict which variants have a deleterious effect on protein function. Generally, determining which variants should be included in a gene burden relies primarily on variant annotations based on sequence context (e.g. loss of function/missense), and is supplemented by various computational approaches to predict the deleteriousness of missense variants; based on factors such as conservation (eg. SIFT, PolyPhen) or more recently, learned representations of higher-order features such as structural context (e.g. AlphaMissense, ESM-1b). Although these novel methods have shown increased accuracy for predicting the deleteriousness of variants, gene burden definitions would ideally be based on empirical data derived from experimentation. With the availability of proteomics data in the UK Biobank, there is now an opportunity to leverage directly-observed effects of nonsynonymous variants on protein levels, to aid in gene burden definitions. In this study, we focus on evaluating the contribution of proteomic data to the utility of gene burden tests in the UK Biobank. We first perform a nonsynonymous variant cis-pQTL discovery analysis, and define three types of gene burdens: 1) traditional gene burdens based on variant annotations and deleteriousness predictions, 2) gene burdens based on nonsynonymous variants that decrease protein levels, 3) a hybrid approach filtering traditional gene burdens to those variants that decrease protein levels. We then perform gene burden tests using these definitions across

thousands of traits, comparing and contrasting associations. We find that different gene burden definitions capture unique associations, supporting the use of an ensemble of gene burdens when conducting analyses. This work has important implications for rare variant analysis, and highlights the utility of large scale proteomics data in biobank cohorts. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4048F: Enhancing polygenic scores prediction across multiple ancestries with S4-Multi model

Authors:

P-H. Lai[1], J. Baierl[1], J. Tyrer[2], P-C. Peng[3], P. Pharoah[4]; [1]Cedars Sinai Med. Ctr., West Hollywood, CA, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Cedars-Sinai Med. Ctr., Los Angeles, CA, [4]Cedars-Sinai Med. Ctr., West Hollywood, CA

Abstract:

Polygenic scores (PGS) are important in predicting genetic risk across different phenotypes. However, most models are Eurocentric, limiting their generalizability and clinical utility in non-European populations. In this study, we extended our previously developed select and shrink with summary statistics (S4) PGS method to multi-ancestry cohorts, introducing the new S4-Multi approach. We assessed its cross-ancestry predictive performance using a simulated dataset and data from various biobanks, including All-of-Us, UK Biobank, FinnGen, and Biobank Japan. We compared S4-Multi to three commonly used methods: LDpred2, LDpred2Reg, and PRScsx, across three phenotypes: breast cancer, colorectal cancer, and type 2 diabetes. To ensure consistency, we implemented five scenarios of variant ID matching and harmonized data across biobanks. Our results demonstrate that multi-ancestry methods outperform single-ancestry models. Specifically, the predictive accuracy (R-squared value) of the S4-Multi model exceeded that of the European-only S4 model by 169% in the simulated dataset. Consistent with these findings, incorporating multi-ancestry cohorts into S4-Multi led to increased AUC and log Odds Ratio values across biobanks compared to employing the European-only S4 model. Among the non-European cohorts, we observed improvement across all three phenotypes, with the most significant enhancements noted in type 2 diabetes, exhibiting a 137% increase in log Odds Ratio in the All-of-Us Asian population (from 0.199 to 0.469) and an 81.63% increase in log Odds Ratio in the Biobank Japan East Asian population (from 0.514 to 0.934). Moreover, integrating multi-ancestry cohorts into the model has also shown to be advantageous for the European population. The S4-Multi exhibited superior performance compared to the European-

only S4 model across all three phenotypes when tested on the European population. In addition, in contrast to LDpred2, LDpred2Reg, and PRScsx, S4-multi required only 23% to 66% of the SNPs for PGS calculation. In conclusion, the S4-Multi method significantly improved genetic risk prediction across diverse ancestries and phenotypes, reducing the number of required SNPs, and thereby enhancing the feasibility of clinical application.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4050F: Evaluating effectiveness of polygenic risk scores in African ancestry breast cancer patients from Penn Medicine BioBank

Authors:

L. Caruth[1], R. Sore[2], S. Ehsan[2], A-M. McCarthy[2], S. Setia-Verma[1]; [1]Perelman Sch. of Med. Inst. of BioMed. Informatics, Philadelphia, PA, [2]Perelman Sch. of Med. Dept. of Biostatistics Epidemiology & Informatics, Philadelphia, PA

Abstract:

Breast cancer is among the most common form of cancer in women in the United States, aside from skin cancer. Early detection and treatment significantly reduce mortality rates, highlighting the importance of effective risk prediction methods. Polygenic risk scores (PRS) are promising tools for representing an individual's genetic predisposition to breast cancer. However, the performance of PRS can vary significantly across different ancestral populations. This study aims to evaluate the effectiveness of three published PRS in predicting breast cancer risk among African ancestry patients. One of the primary motivations of Penn Medicine Biobank (PMBB) is to address the need for diversity in genetic studies, making it an ideal candidate on which to validate the effectiveness of Polygenic Risk Scores developed to better predict risk in patients across diverse ancestries, specifically African-American. Three different PRS, as published in the PGS catalog; Gao et al., Mavaddat et al, Zhang et al., were calculated for a subset of PMBB cases and controls with both mammogram and mammogram-subsequent cancer diagnosis information. These scores, along with age, ancestry, Body Mass Index (BMI) principal components, and breast density, were used to fit a Logistic Regression Model for binary classification of patients as either breast cancer cases or controls. All three PRS assessed for the prediction of overall breast cancer similarly across the entire dataset with AUCs in range of 0.69-0.71. The AUC for the Gao PRS in AFR samples (AUC=0.75) was significantly higher than that observed in EUR samples (AUC=0.68), indicating a better predictive power in the African ancestry population. These findings underscore the importance of evaluating and selecting appropriate PRS for different ancestral populations to aid in bench-to-bedside translation of

breast cancer PRS as a clinical risk prediction tool. The superior performance of the Gao PRS in African ancestry patients highlights the potential for improving breast cancer risk prediction in this underrepresented group. Our study emphasizes the need for continued research and development of PRS tailored to diverse populations, which is crucial for equitable healthcare and effective early intervention strategies. We plan to continue this benchmarking investigation by considering PRS that integrate other clinical features. More specifically we are interested in the Tshiaba et al PRS which incorporates an individual patient's Tyer-Cuzick score, as well as the Hughes et al PRS which utilizes ancestry specific weighting.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4051F: Evaluating long-read phasing methods in clinically-relevant genes using data from the 1000 Genomes ONT Sequencing Consortium

Authors:

N. Damaraju, J. Gustafson, S. Gibson, M. Galey, K. Hoekzema, S. H. Storz, J. Goffena, B. Shirts, E. Eichler, D. E. Miller; Univ. of Washington, Seattle, WA

Abstract:

Determining the specific combinations of variants found on each copy of an individual's chromosomes, known as haplotype phasing, is crucial for the comprehensive study of genetic diseases. Previous methods have relied on population frequency data from microarray or short-read sequencing to computationally predict which variants are likely to be inherited together on the same haplotype. However, this statistical approach has limitations as it requires data from large population samples to accurately phase common variants, and may be biased against underrepresented populations that lack extensive genomic data.Long-read sequencing (LRS) enables direct read-based phasing of haplotypes without relying on population data or trios. While several computational tools have been developed to phase LRS data, their effectiveness for phasing clinically relevant genes has not been thoroughly evaluated. The 1000 Genomes ONT Sequencing Consortium is an international effort to sequence at least 800 of the 1000 Genomes Project Samples on the Oxford Nanopore (ONT) Platform. Here, we present preliminary results of phasing of OMIM genes using the first 200 sequenced samples sequenced to an average depth of coverage of 30x and read N50 of 50 kbp using both alignment and assembly-based pipelines. Our analysis evaluates key parameters such as error rate, genotyping errors, length of phase blocks (groups of variants assigned the same phase), read quality, read length and read depth. We explore the impact of down sampling using both coverage and read

length on phasing. Through feature selection methods, we delineate the relative importance of these key parameters on overall phasing quality offering insights into optimizing LRS-based haplotyping methods. These results will be of broad utility to the human genetics community, particularly those leveraging data from the 1000 Genomes ONT Sequencing Consortium, as well as clinicians, researchers, and developers of haplotyping software that aim to optimize LRS-based haplotyping methods. We anticipate that improvements to phasing will increase the identification of disease-causing variants in individuals with suspected Mendelian conditions sequenced on LRS platforms.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4052F: Evaluating Methods for Genome-Wide Association Studies in Diverse Ancestral Populations

Authors:

J-A. Dias[1], P. Kraft[2], H. Zhang[3], T. Chen[4], R. Madduri[5], A. Rodriguez[6]; [1]Harvard Univ., Boston, MA, [2]Natl. Cancer Inst., Rockville, MD, [3]Natl. Cancer Inst., Cambridge, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [5]Argonne Natl. Lab., Naperville, IL, [6]Argonne Natl. Lab., Lincolnwood, IL

Abstract:

Background: Ancestrally diverse biobanks provide opportunities to study complex traits across various ancestries by leveraging differences in linkage disequilibrium (LD) and allele frequencies. However, the effectiveness of different genome-wide association study (GWAS) methods in multi-ancestry groups is debated. Pooled methods bypass the need to classify individuals into ancestry groups, beneficial for admixed populations. Conversely, meta-analysis methods may better account for finer population stratification but often involve arbitrary population binning and may not be feasible with small number of subjects from particular ancestries. **Methods:** We compared pooled and two meta-analysis methods (fixed-effects meta-analysis and MR-MEGA) under various genetic architecture assumptions. We simulated traits using UK Biobank (UKB) and All of Us (AoU) data with and without population structure on both ancestry-group-specific and global scales. We evaluated these methods' ability to control type I error rates under the null hypothesis. To assess power, we simulated continuous and binary traits, using a large-scale simulated multi-ancestry genotype dataset with realistic LD, similar to the 1000 Genomes Project dataset with 600K subjects in five ancestral groups. We also evaluated power for admixed populations using the Admix-kit pipeline. **Results:** The methods showed comparable performance in controlling type I error rates. However, pooled analysis may not adequately

correct for recent population structure in traits with strong ancestry-specific clines. In power assessments, pooled analysis outperformed both meta-analysis methods. Within meta-analysis methodologies, MR-MEGA demonstrated better power over fixed-effects meta-analysis, particularly in situations with balanced population representation and large sample sizes. We corroborate our simulation findings using multi-ancestry GWAS data from UKBB (N ≈ 340K) with six continuous traits and are conducting additional analyses using AoU data. **Conclusion:** Pooled analysis increases power where allele frequencies between ancestries diverge and there is similar representation across ancestries in the cohort. While pooled analysis may have limitations in correcting for recent population structure in some settings, it offers significant advantages in multi-ancestry cohorts. MR-MEGA outperforms fixed-effects meta-analysis under certain conditions, highlighting the importance of method selection based on cohort characteristics and study goals. These findings provide insights into the performance of GWAS methods in ancestrally diverse biobanks.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4053F: Evaluating multi- and cross-ancestry approaches to Alzheimer's disease polygenic risk scores in diverse cohorts

Authors:

M. Okorie[1], C. Jonson[1], P. Castruita[1], A. Oddi[1], C. Udeh-Momoh[1], J. S. Yokoyama[2], S. J. Andrews[3]; [1]UCSF, San Francisco, CA, [2]UC San Francisco, San Francisco, CA, [3]Univ. of California San Francisco, San Francisco, CA

Abstract:

Background: Alzheimer’s disease (AD) is a progressive neurodegenerative disorder characterized by cognitive and functional decline. There are significant health disparities in AD, with African American and Hispanic populations having a higher incidence of AD. The accuracy of polygenic risk scores (PRS) has been observed to decay with increasing genetic distance from the training population due to differences in genetic architecture across populations. The predictive accuracy of PRS can improved by leveraging diverse genetic datasets. Therefore, the goal of this study is to evaluate the predictive accuracy of AD-PRS models in a diverse cohort using single-, multi-, and cross-ancestry approaches.
Methods: AD-PRS was constructed in the Alzheimer’s Disease Sequencing Project (ADSP; EUR = 10,786; AFR = 4,424; AMR = 3,150; SAS = 1,756). We calculated AD-PRS excluding the *APOE* region using: 1) Single-ancestry AD-PRS calculated from EUR AD GWAS summary statistics using clumping (R2 &gt 0.8; 250kb window; 1000 Genomes EUR reference panel) and

thresholding (PT = 5e-8, 0.1) (C+T); 2) Multi-ancestry AD-PRS calculated from multi-ancestry meta-analysis (MAMA) using C+T method; and 3) Cross-ancestry AD-PRS calculated with PRS-CSx using ancestry-specific EUR, AFR, EAS, and AMR AD GWAS summary statistics. Logistic regression models adjusting for age, sex, *APOE* genotype, cohort, and 10 PCs were employed to evaluate the association of each AD-PRS with a clinical diagnosis of AD. Model performance was assessed by estimating odds ratios, Nagelkerke R2, and AUC.
Results: Among participants of EUR ancestry, AD-PRS were consistently associated with higher AD risk across all models. In contrast, AD-PRS did not show an association with AD risk among participants of SAS ancestry. While the EUR AD-PRS was not associated with AD risk in participants of AFR ancestry, with both MAMA and PR-CSx AD-PRS associated with higher AD risk (OR [95%CI] = 1.41 [1.28, 1.54], p = 1.09e-12 & 1.64 [1.50, 1.78], p = 3.61e-30). Similar trends were observed in AMR participants, with MAMA and PR-CSx AD-PRS associated with higher AD risk (OR [95%CI] = 1.22 [1.14, 1.30], p = 2.24e-09 & 1.35 [1.28, 1.43], p = 4.30e-25).
Conclusion: Using multi- and cross-ancestry PRS models improves model performance of AD-PRS in non-EUR populations, indicating that using diverse genetic datasets improves AD-PRS performance in underrepresented populations. AD genetic risk models can further be improved by increasing the sample size of non-EUR AD GWAS, particularly for SAS populations as the current datasets did not improve predictive accuracy and evaluating PRS and environmental interactions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4054F: Evaluating the impact of race, ethnicity, and ancestry (REA) as population labels in genomics research ⭐

Authors:

B. Maldonado, PAGE Race, Ethnicity, & Ancestry Working Group; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract:

The Population Architecture using Genetics and Epidemiology (PAGE) Study is one of the leading consortia working to better understand the genetic architecture of complex traits in underrepresented populations. The PAGE Study has access to genomic data across multiple cohorts and biobanks, consisting of ~120,000 participants that were categorized using race and/or ethnicity labels according to their original recruitment. Race and ethnicity are socially constructed groupings that sometimes are paired with inferred genetic ancestry for analyses; however, their misuse can be problematic in understanding the role of human genetics in

disease. Our goal is to provide quantitative support for recommendations, starting with participant recruitment, to categorization of individuals for association analyses and follow-up analyses such as polygenic risk scoring (PRS) and fine-mapping. Through extensive literature review and applied research examples within PAGE, we evaluated how operationalization of population descriptors impact genomic insights in meta- and mega-analysis studies. In one example, we assessed the association between a previously developed multi-ancestry type 2 diabetes (T2D) PRS and T2D risk across multiple cohorts. Notably, we observed heterogeneity in per-SD odds ratios for PRS in Mexican-American individuals: 3.41-fold (95% CI [2.71, 4.28]) in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL, N=1,009) and 2.23-fold (95% CI [1.89, 2.64]) in the Women's Health Initiative (WHI, N=1,636), illustrating effect heterogeneity across study populations with the same granular population descriptor, which remained when stratifying studies by sex and age categories. As a second example, we use results from exact and approximate conditional analyses in glycemic traits to evaluate how external vs. internal LD reference panels impact results. In the associated *GCK* locus, we observed discordance in the identified top SNPs when using European and African individuals from the U.K. Biobank as the LD reference rather than the same coarse groupings of PAGE individuals. The complexity of population descriptors, combined with differences in genetic architecture, can impact the use of LD matrices, not only in conditional analyses, but any analyses relying on ancestrally-matched reference populations. Additionally, the mismatch between social constructs and genetic ancestry can create challenges, especially when working with different cohorts and contexts. These findings demonstrate the importance of carefully considering the use of study, race, ethnicity, and ancestry labels and how these choices can impact genomic analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4055F: Evaluation of rare variant association methods when incorporating data from multiple sources

Authors:

J. Murphy, S. Sigler, A. Hendricks; Univ. of Colorado Anschutz, Aurora, CO

Abstract:

Rare variant associations are crucial to understanding the genetics of complex traits and diseases. Yet, the large sample sizes required to observe rare variants can be costly and time consuming to ascertain. Combining data, such as through meta-analysis of case-control studies or incorporating public summary data as external controls, can boost power. Combining

multiple study types (e.g., case-control, case-only, and control-only) can further boost power by substantially increasing the sample size and by including samples in the analysis that would otherwise be discarded. However, using data from multiple sources can introduce bias due to differences in sample ascertainment and processing. Here, we compare the performance of rare variant association methods that can incorporate external controls (iECAT-O and ProxECAT) with our new method (LogProx) that can leverage data from multiple sources and study designs. Through simulations, we demonstrate that the type I error is well-controlled (i.e., 0.05) for ProxECAT and LogProx across most scenarios when analyzing internal cases vs only external controls. However, the type I error is inflated (e.g., ~0.1) when there is substantial confounding between internal cases and external controls (e.g., <80% of rare alleles are detected in controls) and for large sample sizes (>20,000). When analyzing internal cases with both internal and external controls, the type I error is well-controlled (i.e., 0.05) across all simulation scenarios for both iECAT-O and LogProx. In scenarios where there are external control sets for each internal case-control dataset, iECAT-O has optimal power compared to LogProx for large sample sizes (>20,000). However, LogProx has higher power when there are additional case-only and control-only datasets that cannot be used with iECAT-O.
Notably, LogProx is the only method that can incorporate different study designs (e.g., internal cases vs only external controls or internal cases vs internal and external controls) and multiple data types (e.g., case-only or control-only) in one analysis. This is likely to be particularly useful for uncommon diseases and underrepresented populations, which may have limited data available. More broadly, we provide guidance in identifying the optimal method (i.e., achieves maximum power while controlling type I error) for various study designs and available data types. Our comprehensive simulation study along with best practice guidelines for incorporating external control data and multiple data sources will aid in the discovery of new genetic associations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4056F: Exome-wide copy number variation in 106,744 individuals from autism spectrum families in the Simons SPARK cohort

Authors:

H. Kim[1,2], G. Shaw[1], J. Ng[3], T. Mosbruger[1], R. Rajagopalan[1,2], T. N. Turner[3,4], T. Hayeck[1,2]; [1]Div. of Genomic Diagnostics, Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Dept. of Genetics, Washington Univ. Sch. of Med., St. Louis, MO, [4]Intellectual and Dev.al Disabilities Res. Ctr., Washington Univ. Sch. of Med., St. Louis, MO

Abstract:

Rare and *de novo* structural variants such as copy number variations (CNVs) are important in understanding genetic contribution to diseases and traits. They have been shown to play a critical role in the onset of neurodevelopmental disorders, like autism spectrum disorder (ASD). However, calling CNVs from short read whole-exome sequencing platforms presents several challenges that limit its use for clinical diagnostics in studying ASD. The inherent biases in targeted approaches such as exome sequencing produce noisy data and the existing callers produce inconsistent and often false calls. It is a common practice to utilize more than one algorithm to create a union call set or a consensus calling approach (combining consistent calls across multiple tools) posing the risk of elevated false negative or false positive rates. For the aforementioned reasons, CNV calling from short read sequence data can requires labor intensive manual curation. The aim of this project is to provide both a tool and a resource. We developed a CNV calling and scoring tool that efficiently converts observed read piles into signals, denoises them, and detects breakpoints using fast signal processing with kernel changepoint detection. Our tool integrates and scores CNV calls from other tools (i.e., XHMM, CoNIFERm cn.mops), providing a comprehensive map of CNVs, from whole-exome sequencing data, for 106,744 individuals from 39,881 ASD families in the SFARI SPARK cohort across multiple callers. To predict the accuracy of CNVs, we use a machine learning model informed by a series of signal features. These features include annotations of sequence context (GC content, segmental duplications, etc.), quality metrics (mappability, alignment, batch effects, etc.), ancestry and population demographics, and regional detection of noise versus common CNVs (normality tests, modality inference etc. across all samples leveraging efficient kernel density using Fast Fourier Transformation). The model is trained using both orthogonally validated CNV calls in conjunction with expert curated CNVs. The tool includes automatically generated graphical representations to pinpoint key features that enhance the prediction accuracy of detecting CNVs. We provide a tool along with a comprehensive and robust map of CNVs across over a hundred thousand whole exome ASD samples in the SFARI SPARK cohort. Since our tool also integrates and scores CNV calls from other tools, it allows for clean integration with existing pipelines to improve prediction accuracy of CNV calling in both research and clinical settings.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4057F: Explore the use of graph neural networks to study the connection between the gut microbiome and bone health

Authors:

L. Jiang[1], Z. Wang[2], L. Zhang[2], B. Tian[3], X. Lin[4], X. Zhang[5], A. Liu[5], Y. Gong[5], C. Qiu[5], K. Su[5], Z. Luo[5], Q. Tian[5], L. Wu[5], H. Shen[5], S. Mungasavalli Gnanesh[1], H-W. Deng[5]; [1]Tulane Univ. Sch. of Med., New Orleans, LA, [2]Beijing Jiaotong Univ., Beijing, China, [3]Central South Univ., Changsha, China, [4]Southern Med. Univ., Foshan, China, [5]Tulane Univ., New Orleans, LA

Abstract:

The increasing recognition of the human gut microbiome's role in health and disease highlights its importance. Particularly, the relationship between gut microbiota composition and bone health, and the methodologies used to explore this link, have drawn significant attention. Our research focuses on evaluating a graph neural network (GNN) model's ability to uncover the relationship between the gut microbiome and bone status. This GNN model combines distance metrics (represented by subject-subject graphs) and feature vectors (such as the compositional relative abundance of the gut microbiome) and is expected to effectively manage nonlinear relationships between variables. Additionally, it can accommodate various covariates like age, gender, and ethnicity, leveraging the high flexibility of deep learning techniques. We gathered data on hip total bone mineral density (BMD), demographic details, and gut microbiome composition at different taxonomic levels (e.g., species, genus, phylum) from over 2,000 individuals in Louisiana. The model's predicted BMD scores were compared to observed BMD scores, and mean squared error was calculated to assess prediction accuracy. An ablation study examined the effects of excluding gut microbiome profiles from the model input, shedding light on the significance of microbiome data in BMD prediction. Initial findings indicate that incorporating microbiome-related information extracted via GNN did not enhance prediction accuracy when demographic and clinical variables were already included. However, a detailed comparison of different distance measures, such as Bray-Curtis, UniFrac, and Aitchison distance, is necessary to fully understand their impact on the subject-subject graphs and GNN model performance. Furthermore, we may need to compare the use of raw relative abundance data with centered log-ratio transformed data to see if the microbiome data's compositional nature significantly affects model performance. This study not only underscores the potential of deep learning in deciphering complex interactions between the gut microbiome and bone health but also sets the stage for more targeted and thorough future research in this area.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4059F: Exploring the Effects of Genetic Variants on Gene Regulatory Networks

Authors:

L. Martini[1], M. Gentili[1], B. Hobbs[2], M. Cho[3], K. Glass[1]; [1]Harvard, Boston, MA, [2]Brigham and Women s Hosp., Needham Heights, MA, [3]Brigham and Women s Hosp., Boston, MA

Abstract:

Understanding how individual genotypes shape gene regulatory networks (GRNs) can significantly advance our knowledge of human health and disease, leading to more precise risk assessments and treatments. Traditional approaches to GRN construction often overlook the influence of genetic variants, despite their known regulatory impacts, such as the disruption of transcription factor (TF) binding sites.
To address this issue, we developed RAVEN (Regulatory Alteration based on Variants Edge aNalysis), a novel framework designed to infer population-based GRNs using genotype data. RAVEN first constructs individual-level GRNs using advanced message-passing algorithms (i.e., LIONESS and PANDA) that incorporate gene expression data, TF protein-protein physical interaction data, and TF-motif predictions. These individual GRNs are combined with individual genotype data to discover edgeQTLs, identifying genetic variants that influence the strength and presence of TF-gene regulatory interactions. To account for potential confounding factors, we adjusted for relevant covariates: age, sex, and population stratification.
We used RAVEN to identify edgeQTLs for 177 genome-wide significant associated variants in 507 chronic obstructive pulmonary disease (COPD) and 222 idiopathic pulmonary fibrosis (IPF) whole genome and RNA-sequencing samples in the Lung Tissue Research Consortium (LTRC) dataset. We first confirmed appropriate control of Type 1 error for edgeQTLs. Comparisons between RAVEN and traditional eQTLs demonstrated a strong correlation, with 40% of genes identified by RAVEN also being eGenes, including DSP, DEPTOR, and MUC5B, previously associated with IPF and COPD. In particular, we found the greatest number of edgeQTLs on chromosome 17 (KANSL1-AS1, target of 626 edgeQTLs) and 11 (MUC5B, target of 153 edgeQTLs). Finally, for each gene, we designed a transcription factor (TF) signature based on its associated edgeQTLs and clustered based on the similarity of these signatures. Our analysis identified a cluster that includes MUC5B and KANSL1-AS1 suggesting that their associated genetic variants influence their regulation through a similar set of transcription factors.
These results show RAVEN's potential in identifying genotype-specific regulatory mechanisms, highlighting significant edgeQTLs, and offering insights into the genetic regulation of COPD and IPF.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4060F: Exploring the mediating role of inflammation and immune-related gene expression in the chronic stress-cognitive function pathway in European and African ancestry populations

Authors:

L. Laiton Jimenez[1], I. Shalev[1], M. J. Sliwinski[1], Y. Veturi[2]; [1]Pennsylvania State Univeristy, University Park, PA, [2]Pennsylvania State Univ., University Park, PA

Abstract:

Cognitive decline is a salient feature of dementia. Chronic stress negatively affects cognitive function, especially in women, and increases inflammatory markers, likely due to elevated immune-related gene expression. The goal of this study was to understand putative causal mediations of (1) inflammation (N=7,111) and (2) immune-related gene expression (N=1,215) on the relationship between chronic stress and cognitive decline. In (1), we used a manually curated list of 50 childhood/adulthood stressors (ST), 5 inflammatory biomarkers (IB), gene expression of 51 immune-related genes (GE) associated with conserved transcriptome response to adversity (CTRA), and 6 cognitive function (CF) measures in the Mid Life in United States (MIDUS) Core and Refresher cohorts, which studies how diverse factors impact adult health over 28 years. In (2), we considered only significant IBs-associated STs, GEs, and CFs. We conducted mediation analyses using GLM models for (a) mediator ~ ST, (b) CF ~ mediator, (c) CF ~ ST+ mediator, and additionally IB ~ GE, adjusting for age, BMI, education, and physical activity, separately for European (EUR) and African American (AA) descents. In (1), we identified 34 STs as significantly associated with at least one IB (30 in EUR, 19 in AA; $p < 0.05$), with 15 affecting C-reactive protein (CRP) in EUR. Only the CF Reasoning showed significant associations with at least one IB in EUR and AA ($p < 0.05$), while Processing Speed and Attention Switching were significant only in EUR ($p < 0.05$). We identified 20 GEs (13 in EUR, 13 in AA) significantly associated with at least one IB ($p < 0.05$). In (2), 13 GEs in EUR and 12 in AA were associated with at least one ST ($p < 0.05$). Overall, five GEs were associated with Reasoning or Processing Speed across populations. We observed significant sex effects in specific segments of (1), but not (2). In EUR, we found a putative causal pathway where "physical abuse by a sister" increased *IRF8* GE and CRP, subsequently reducing Reasoning performance, identifying CRP and *IRF8* as potential mediators. Similar pathways involving two other childhood stressors showed comparable changes in the same GE and CF. The IRF8 transcription factor regulates microglia, which release harmful proinflammatory factors worsening cognitive decline when activated. Our findings suggest chronic adulthood/childhood stressors impact inflammation, immune-related GE, and CF in EUR, elucidating the ST-CF pathway. We also identified segments of the ST-CF pathway influencing sex differences. We plan to improve power of gene detection

related to chronic stress, inflammation, and CF by imputing MIDUS transcriptome-wide gene expression into the UK Biobank data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4061F: Fast and efficient toolkit for rare variants association testing

Authors:

X. Li[1], W. Qiang[2,3], F. Hu[4], S. MacArthur[4], S. Wasilewski[4], Q. Wang[2], S. Petrovski[4,5]; [1]AstraZeneca, Gaithersburg, MD, [2]AstraZeneca, Waltham, MA, [3]Middle Tennessee State Univ., Murfreesboro, TN, [4]AstraZeneca, Cambridge, United Kingdom, [5]Univ. of Melbourne, Melboure, Australia

Abstract:

Next-generation sequencing (NGS) has emerged as a powerful tool for identifying both coding and non-coding rare variants associated with disease, particularly in the clinical diagnosis of inherited disorders. However, traditional approaches for conducting phenome-wide association tests on a genome-wide scale have often been time consuming and resource-intensive. To address this challenge, we have developed a comprehensive toolkit consisting of three fast and efficient computing procedures for rare variants association testing. These procedures have been designed to meet the demands of genome-wide association studies (GWAS) and phenome-wide association studies (PheWAS) in biobanks, where billions of association models and hypothesis tests need to be performed. Our approaches leverage shared information among the individual models, naturally accommodate parallel computing with sparse metrices, resulting in substantial savings in compute time and memory usage while maintaining statistical accuracy. Our approaches are most efficient when we do multiple genotypes and phenotypes together. Specifically, our bulk linear regression approach is 70 to 400-fold faster than the classical method while maintaining statistical accuracy in three UK Biobank (UKB) PheWAS associations. When dealing with binary phenotypes, our proposed binarized confounder effect model efficiently incorporates covariates, demonstrating computational efficiency and robustness for unbalanced case-control phenotypes. Its implementation significantly outperforms Firth regression by about 1900-fold faster when applied to UKB whole genome sequence data. Lastly, we propose an ultra-fast implementation of Fisher Exact Test (FET) by utilizing parallel processing and vectorization to handle multiple genotypes and phenotypes simultaneously. When conducting a collapsing PheWAS association, this new computing procedure is100-fold faster with doubled memory usage or 35-fold faster with optimized memory usage, leading to 99% compute cost saving.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4062F: Fast multi-cohort approximate fine-mapping

Authors:

R. Cui[1], M. Kanai[1], B. Neale[2], M. Daly[2], Z. Fan[3]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Yale Univ., New Haven, CT

Abstract:

Fine-mapping refines GWAS signals and aims to pinpoint causal variants for phenotypes. Incorporating multiple cohorts with diverse ancestries can further refine signals. This is because the increased sample size and different LD patterns can boost statistical power. Methods that jointly model multiple cohorts and use in-sample LD (e.g., MESuSiE) have proven superior to meta-analysis-based fine-mapping methods in terms of accuracy and power. However, meta-analysis-based approaches are much easier to adopt due to simpler data sharing and lower computational costs. Here, we propose FastMap, which has the benefits of both the accuracy and power of joint modeling methods and the convenience of easy data sharing and near-trivial computational time. FastMap requires each cohort to generate full SuSiE results separately, then combines them into approximate joint fine-mapping posterior probabilities by using one cohort's Posterior Inclusion Probabilities (PIPs) as priors to update the other cohort's PIPs for each independently fine-mapped signal. Simulations show that our method is well-calibrated and has comparable recall to MESuSiE. In cases where cohort heterogeneity exists and causal variants may be missing from one of the two cohorts, FastMap outperforms MESuSiE in both calibration and recall, as it doesn't require filtering SNPs to a shared set. Real data benchmarking shows that the approximation is of high quality; PIPs generated by our method are close to those generated by MESuSiE (R^2 = 0.88). Notably, when the two methods disagree, our method tends to be more conservative, indicating minimal additional false discoveries. By applying FastMap to existing FinnGen (N=500K) and UK Biobank (N=360K) fine-mapping results for type 2 diabetes and BMI, we found 6 and 10 additional (9% and 12% increases for T2D and BMI respectively) high-confidence (PIP>0.9) variants that were not discovered from either single-cohort fine-mapping, including four missense variants, one splice variant and one 3 prime UTR variant. 15 out of these 16 variants are in a credible set for at least one cohort, showing evidence of increased power of detection. Our method can be generalized to a higher number of cohorts, opening the door to large-scale multi-cohort refinement of causal signals.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4063F: Functional components of heritability in admixed African-ancestry All of Us whole-genome sequencing data

Authors:

J. Rossen[1], A. L. Price[2]; [1]Harvard Sch. of Publ. Hlth., Boston, MA, [2]Harvard Sch Pub Hlth, Boston, MA

Abstract:

Estimating functional components of heritability can improve our understanding of disease etiology, but has largely been focused on European-ancestry populations. Admixed populations pose a technical challenge due to complexities of admixture-LD, recently addressed by cov-LDSC, which uses ancestry-adjusted in-sample LD scores (Luo et al. 2021 Hum Mol Genet).
In the most comprehensive analysis of functional components of heritability in admixed African-ancestry individuals to date, we applied cov-LDSC to All of Us whole genome-sequencing data for 47,401 African-ancestry individuals (All of Us 2024 Nature), spanning 14 approximately independent quantitative traits, 19 million variants for computing LD scores, and 38 binary functional annotations from the baseline-LD model. We compared our results to the results of LDSC applied to an equal number of European-ancestry individuals from All of Us. Interestingly, common variant heritability was significantly lower in the African ancestry cohort for 11 of 14 traits (22% lower on average, p=10-5 for average difference via genomic block-jackknife). However, estimates of heritability enrichment were highly concordant across the two ancestries (slope=0.92, Pearson correlation = 0.74 for 11 approximately independent annotations, meta-analyzed across traits). This result is consistent with previous studies of European and East Asian ancestries (Kichaev and Pasaniuc 2015 Am J Hum Genet, Kanai et al. 2018 Nat Genet), but discordant with differences in the functional enrichment of fine-mapped variants in the same All of Us cohorts (Rossen et al. 2024 medRxiv).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4064F: Gene Dosage Architecture in Complex Traits and Common Diseases with Mediation of the Brain and Body ⭐

Authors:

S. Kazem[1], K. Kumar[1], T. Renne[1], G. Huguet[2], M. Jean-Louis[2], Z. Saci[2], L. Almasy[3], D. Glahn[4], G. Dumas[1], **S. Jacquemont**[1]; [1]Univ. of Montreal, Montreal, QC, Canada, [2]CHU Sainte-Justine, Montreal, QC, Canada, [3]Children`s Hosp. of Philadelphia, Philadelphia, PA, [4]Boston Children's Hosp. & Harvard, Boston, MA

Abstract:

Deciphering the functions of cells and tissues that mediate genetic associations with complex traits (CTs) and conditions is a central question in human genetics. Genome-wide association studies (GWASs) have identified thousands of significant loci with human CTs that are often non-coding and impact gene expression. While functional genomic studies have shed light on the tissues and cell types disrupted by these variants and how they impact gene expression levels, it remains challenging to link common non-coding variants to genes and corresponding functions. In contrast, copy number variants (CNVs) - deletions or duplications of DNA segments- invariably lead to a large decrease and increase of gene expression when a gene is fully encompassed in a deletion or a duplication respectively. Therefore, they provide direct insight into the effects on CTs of large changes in gene expression. Previous studies focused on a small set of recurrent CNVs, leaving the impact of the majority of rare coding CNVs on CTs unknown.This study aims to characterize the exome-wide CNV architecture of CTs and test the functional convergence between common variants (CVs) and coding CNVs across 28 CTs, using UK-Biobank dataset. We analyzed CNVs >50 kilobases encompassing ≥1 gene. Due to low allele frequencies, we aggregated variants disrupting genes with similar functions across 60 whole body tissues(covering brain and non-brain), 81 and 31 body and brain cell types(from Human Protein Atlas and Human Brain Cell Atlas). We then estimated CNV effect sizes on traits using the burden association test and compared the functional convergence of CNVs and CVs.CNVs disrupting brain-related gene sets showed the highest associations with both brain and non-brain traits. Primarily due to the pleiotropy, measured by the number of independent traits associated with CNVs disrupting a gene set, was highly correlated to proportion of genes under constraint in each gene set. The CNV burden correlations were systematically higher than genetic correlations computed using CVs across CTs (slope:1.53, p=2E-35). We found that at least 50% of the CTs have significant enrichments across our functional gene-sets. Moreover, within GWAS significant enrichment, CNV burden association intersected with 44% and 33% of whole body tissue gene-sets, 31% and 21% for whole body cell type gene-sets, and 21% and 16% of whole brain cell type gene-sets, for deletions and duplications respectively.Our systematic comparison reveals that the same tissues and cell types might mediate the effects of common and rare variants on CTs. Our findings indicate that most CTs are associated with CNVs disrupting gene dosage functions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4065F: Gene-environment interactions are critical for future improvement in translational genetics

Authors:

H. Aschard, L. denis, J. lefevre, L. Henches; Inst. Pasteur, Paris, France

Abstract:

Understanding how genetic effects are modified by individual's environment and lifestyle is commonly cited as an important research domain for moving forward ongoing translational genetic efforts in risk prediction, the identification of drug targets, and pharmacogenetics. However, current research on gene-environment interaction (GxE) has produced very limited evidence in this regard, suggesting instead a very modest contribution of GxE to human quantitative phenotypes and disease, with both limited number of variant-by-exposure interaction identified, and very modest contribution to heritability. Here, we argue that the enthusiasm for GxE is justified and that existing approaches to qualify and quantify the contribution of GxE poorly reflect their potential translational value. We conducted series of GxE interaction with a range of environmental factors (alcohol consumption, sleep behavior, physical activity, etc) using individual-level data from 330K participants of European ancestry from the UK Biobank and body mass index (BMI) as a case study. The narrow-sense heritability of BMI in this sample equals 0.235 (se = 0.008. The estimated contribution to heritability of GxE, as measured using state-of-the-art methods, is very modest, ranging in [0.003, 0.017]. However, stratified analyses for those same environments tells a different story. For example, heritability conditional on walking pace equals 0.183 (se=0.011), 0.230 (se=0.011), and 0.284 (se=0.035), for brisk, steady and slow walking pace, respectively (Ptrend = 0.028). The top 100 (independent) associated variants with BMI show a similar trend with average genetic effect increasing from 0.009, to 0.012 to 0.021, respectively. We also observed substantial variability in effects estimate at single variants across strata, and a significant decreasing trend in genetic correlation when using the most extreme strata (Ptrend = 0.012), suggesting variability in the genetic pathway involved conditional on the environment. Altogether, our systematic screening suggests that these differences in genetic effect and heritability across the exposure spectrum can strongly impact the predictive power of PRS, the identification of the biological pathways impacted, and our ability to improve screening and treatment strategies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4066F: Gene-level Analyses of Novel Olfactory-related Signal from Severe SARS-CoV-2 GWAS Reveal Robust Association with Disease Mortality

Authors:

Y. Zhao[1], L. A. Mucci[1], X. Lin[1], D. C. Christiani[2]; [1]Harvard Sch. of Publ. Hlth., Boston, MA, [2]Harvard Med. Sch., Boston, MA

Abstract:

Importance: The coronavirus disease 2019 (COVID-19) is the third leading cause of mortality in the United States for three years in a row. The genetic contributions to disease severity remains unclear and many previously identified single nucleotide polymorphisms (SNPs) have not been replicated nor linked with functional significance.
Objective: To identify SNPs associated with mortality among hospitalized COVID-19 patients supplemented by expression quantitative trait loci (eQTL) evidence to infer plausible functional mechanisms related to COVID-19 severity.
Design: A quality-controlled genome-wide association study (GWAS) supported by robust gene-level omnibus kernel association tests (SKAT-O), functional prediction, and eQTL analyses of the top GWAS signal.
Setting: Massachusetts General Hospital (MGH).
Participants: 370 adult ICU patients with SARS-CoV-2 infection and acute hypoxemic respiratory failure and floor patients with mild hypoxemia managed with supplemental oxygen consecutively admitted to MGH between March and June 2020 (Surge 1), and January and March 2021 (Surge 2) with baseline clinical characteristics and demographics collected.
Exposures: Low-pass genotyped SNPs from whole blood and aggregated SNP-sets of potential disease susceptibility loci with ±500kb flanking regions.
Main Outcomes and Measures: Genome-wide individual SNP associations and SNP-set associations with mortality outcomes from 370 severe COVID-19 cases.
Results: After LD pruning (<0.8) and false discovery rate adjustment (<0.05), we identified rs7420371 G>A of the receptor transporter protein 5 (RTP5) gene as the top independent signal significantly associated with 30- and 60-day mortality among severe COVID-19 patients (OR, 2.32; 95% CI, 1.59-3.39; P = 4.92e-09 and OR, 2.06; 95% CI, 1.43-2.97; P = 5.43e-08, respectively). SKAT-O analyses on the RTP5 SNP-set showed robust associations with both mortality outcomes (P = 5.90e-05 and 6.17e-05, respectively). eQTL analysis showed rs7420371 A allele significantly upregulated the mRNA expression of RTP5 in 241 cerebellum tissues and 215 cerebellar hemisphere tissues while significantly downregulating RTP5 mRNA expression in 238 adrenal tissues.

Conclusions and Relevance: We discovered a novel, independent, and potentially functional SNP RTP5 rs7420371 G>A to be significantly associated with COVID-19 mortality. The A allele significantly influences the mRNA expression of RTP5, an important protein coding gene that modulates olfactory binding and taste perceptions in response to SARS-CoV-2 infection.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4067F: Genetic determinants of sleep architecture inferred from accelerometer-based wearable devices in the UK Biobank ⭐

Authors:

P. Gundannavar Vijay, **M. Cule**, Z. Chen, M. Onsum, A. Raj; Calico Life Sci., South San Francisco, CA

Abstract:

Sleep is a complex phenotype influenced by the interaction among genetics, age, environment, and lifestyle. Polysomnograms (PSG) are the gold standard for measuring sleep architecture - sleep stages, their duration, and associated physiologic parameters. However, collecting PSG data from large human cohorts is challenging. Accelerometer-based wearable devices offer a practical alternative for measuring sleep in large human cohorts. We developed novel methods to extract features predictive of sleep stages from accelerometer measurements, including heart rate, respiration rate, and several motion-related features. We trained a gradient boosting model on the Physionet dataset, containing matched PSG and accelerometer data, to predict wake, rapid eye movement (REM), and non-rapid eye movement (NREM) sleep, and achieved a balanced class accuracy of 70% across all stages. We applied our model to week-long, triaxial accelerometer measurements in over 100,000 participants in the UK Biobank and derived 25 novel phenotypes such as the duration and stability of sleep stages, the time to onset of these stages, and sleep fragmentation. We observed that NREM sleep decreases with age while REM sleep increases with age, with stronger effects in females. The fraction of time spent in REM sleep predicted reduced mortality due to respiratory and gastrointestinal diseases, while NREM sleep predicted reduced mortality due to cardiovascular and neurological diseases. These 25 sleep phenotypes were genetically correlated to 51 structural and functional phenotypes derived from brain MRI and fMRI measurements, including time to fall asleep linked to reduced volumes of grey matter in the left middle temporal gyrus and right nucleus accumbens. Genetic mapping of these sleep phenotypes identified 42 loci, of which prior studies identified 9 using self-reported sleep duration and 13 using accelerometer-derived sleep. Fine-mapping and colocalization of all 42 loci recovered genes known to be associated with sleep disorders,

including MEIS1, BTBD9, and PAX8, and identified novel genes such as MYT1 and CALB1 associated with deep and REM sleep, respectively. Genetic associations at the MYT1 locus colocalized with MYT1 eQTLs in the nucleus accumbens of basal ganglia, a region of the brain involved in deep sleep. CALB1 expression in the dorsal paragigantocellular nucleus has been linked to REM-associated eye movements and its expression in the substantia nigra might play a protective role against Parkinson's disease. Our results provide novel insights into the genetic underpinnings of sleep stages, and their links to brain morphology and physiology, disease, and mortality.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4068F: Genetic evidence for a dose-dependent response to a *TYK2* inhibitor in psoriasis

Authors:

A. Schaefer[1], M. A. Nalls[2], **S. Esmaeeli**[1]; [1]Neuron23, Inc., South San Francisco, CA, [2]Data Tecnica Intl. LLC, Glen Echo, MD

Abstract:

Psoriasis is an immune-mediated, chronic inflammatory disease leading to significant impairment in quality of life for the around 3% of adults affected in the United States. Intracellular tyrosine kinase 2 (*TYK2*) is involved in psoriasis pathophysiology as a mediator of Interleukin-23 and Interleukin-12 signaling pathways, in turn affecting the production of Interleukin-17 by Th17 cells and TNF-α and IFN-γ release by Th1 cells. These downstream agents play crucial roles in the inflammatory response inherent to psoriasis and other autoimmune diseases, highlighting *TYK2* as a critical disease-modifying link. Recently, Deucravacitinib was approved by the US Food and Drug Administration as an oral treatment for moderate-to-severe plaque psoriasis. In phase 2 and phase 3 trials, the allosteric *TYK2* inhibitor was shown to improve psoriasis symptoms in a dose-dependent manner. During the development of pharmaceutical compounds, the expectation of a dose-dependent response is an important consideration. In this study, we conducted a post-hoc investigation of genetic evidence that could have forecasted the observed dose-dependent response to a *TYK2* inhibitor in psoriasis. We carried out a comparative genome wide association study (GWAS) using genotype data from 487,409 subjects provided by the UK Biobank project. We carefully selected a cohort of psoriasis cases (n=16,011) based on ICD-10 diagnosis codes from in-patient and out-patient health records and controls without skin conditions and autoimmune diagnoses (n=191,267). We identified genetic loci associated with psoriasis risk by conducting GWAS analyses under

dominant, additive, and recessive assumptions of inheritance. By comparing the model fit between the different modes of inheritance, we detected associations with an allele-dose dependent effect on disease risk. Using individual level proteomics data acquired through the UK Biobank Pharma Proteomics Project, we further interrogated the effect of the identified dose-dependent polymorphisms on plasma protein expression levels with the aim to interpret their consequences in the context of the *TYK2* signaling pathway. The study allowed us to identify strongly dose-dependent genetic risk associations for psoriasis within genes directly linked to *TYK2* downstream signaling, conceivably explaining the dose-dependent improvement in symptoms observed in response to an allosteric *TYK2* inhibitor. The results serve as a proof of concept to an approach that could be applied to other indications to predict dose-dependent treatment response from genetic evidence.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4069F: Genetic Variants Associated with Muscle Strength in the United Arab Emirates: Insights from the Emirati Genome Project

Authors:

W. Mohammed[1], S. AlMarzooqi[2], N. Marzouka[3], A. Alsuwaidi[3], A. Al aamri[3], M. Olbrich[3], M. Mousa[3], H. Alnaqbi[3], J. Quilez[4], T. Magalhaes[4], S. Ibrahim[3], M. Alameri[1], A. AlMannaei[1], H. Alsafar[5]; [1]Dept. of Hlth.- Abu Dhabi, Abu Dhabi, United Arab Emirates, [2]Dept. of Hlth., Abu Dhabi, United Arab Emirates, [3]Khalifa Univ., Abu Dhabi, United Arab Emirates, [4]G42 Hlth.care, Abu Dhabi, United Arab Emirates, [5]Khalifa Univ. of Sci., Technology & Res., Abu Dhabi, United Arab Emirates

Abstract:

Muscle strength is one of the most critical factors in athletic performance and physical fitness. It is also independently associated with functional ability in the elderly and may explain up to 25% of the variance in overall functional ability. Research has revealed that genetics influences muscle strength, with several genes found to regulate muscle growth and development. Skeletal muscle is a highly heritable trait, with heritability estimates ranging from 30-85% for muscle strength. Genes associated with muscle strength include the *MSTN, IL6, IGF1*, and *PPARG* genes. *MSTN* codes for the myostatin protein, which is responsible for restraining muscle growth. The *IGF1* gene regulates both anabolic and catabolic pathways in the skeletal muscle, with an important association with muscle mass entity, strength development, and degeneration. The *IL6* gene produces a protein that enhances muscle tissue growth and repair, thus significantly improving athletic performance. The *PPARG* gene primarily involves metabolic

processes and energy utilization while contributing to muscle fiber type and preventing muscle atrophy. Using over 43,000 Emiratis, whole genome sequencing (WGS) data from the Emirati Genome Project were analyzed to identify genetic variants in the *MSTN, IL6, IGF1*, and *PPARG* genes. Variants were annotated using Annovar, and linkage disequilibrium (LD) analysis was performed using Plink. Allele frequencies (AF) of *MSTN, IGF1, IL6*, and *PPARG* variants were compared with global populations. The AF of *IL6*- rs1800795 (G allele) variant was identified in 84% of the UAE cohort, compared to 64% of the global population. The *MSTN*-rs1805086 (C allele) variant was found in 5.5% of the UAE cohort, compared to 0.03% of the global population. The *PPARG*-rs1801282 (G allele) variant was identified in 6% (compared to 0.07%, globally), while the *IGF1*-rs6220 (A allele) and *IGF1*-rs7136446 (T allele) variants were found in 68% (vs. 70% globally) and 65% (vs. 60% globally) of the Emirati cohort, respectively. This large-scale study of the Emirati population is the first to provide new insights into the genetic advantages of muscle strength in the UAE. With an overall higher AF of muscle strength-related variants in the Emirati population, in comparison to global populations, further research into this field is needed to create and tailor lifestyle modifications based on our genomic landscape.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4070F: Genetically regulated enhancer RNA expression predicts enhancer-promoter contact frequency and explains a substantial proportion of GWAS signals not attributable to canonical gene eQTLs

Authors:

M. Betti[1], P. Lin[1], M. C. Aldrich[1], E. R. Gamazon[1,2]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Clare Hall, Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

Introduction: Enhancers are essential mediators of gene expression, regulating spatial and temporal expression patterns through recruitment of DNA-binding proteins and establishment of chromatin conformation. Enhancer mediation of chromatin activity and gene expression has been well-studied. However, the biological functions of enhancer RNAs (eRNAs) transcribed from these regulatory elements and their impact on human disease remain underexplored. We sought to identify whether eRNA eQTLs could partially explain the known discrepancy between colocalization of GWAS signals and eQTLs .
Methods: We trained *in silico* models of genetically regulated expression (GReX) of eRNAs across 49 cell and tissue types from GTEx. We leveraged the estimated GReX for eRNAs and

canonical genes in a large-scale DNA biobank of European ancestry individuals (N > 70,000) and high-resolution Hi-C contact data. We then trained a deep learning-based regression model to predict pairwise 3D chromatin contact frequency for enhancer-enhancer and enhancer-gene pairs in cerebellum and whole blood.
Using eRNA-based TWAS of schizophrenia (SCZ) and Mendelian randomization, we explored putative causal disease associations. Finally, utilizing an integrative set of approaches informed by functional experiments, we identified potential *cis* and *trans*-regulatory mechanisms by which these causal eRNAs influence SCZ risk. Utilizing the same approach for eRNA-based TWAS, we generated a comprehensive resource of tissue-specific enhancer associations with over 4,000 UK Biobank traits. Finally, we mapped eQTLs of eRNAs and canonical genes across 49 tissues and performed colocalization with all significant ($p < 5 \times 10^{-8}$) UK Biobank GWAS associations.
Results: We found that the use of eRNA GReX had substantial power in predicting chromatin contacts assayed using Hi-C ($R^2 > 0.37$), supporting a role for eRNAs in 3D mediating spatial chromatin organization. In the TWAS of SCZ, we identified 104 unique causal eRNAs (64 in the brain) and 396 unique canonical genes (130 in the brain) associated with SCZ, with most acting through independent mechanisms. Across the entire UK Biobank, we identified 467 complex traits with at least one genome-wide significant ($p < 2.60 \times 10^{-10}$) eRNA association. Finally, using colocalization analysis, we demonstrated that the number of significant GWAS associations that colocalizes with an eQTL increases by 63% when eRNA eQTLs are included.
Conclusions: Our results suggest that eRNAs play a key role in mediating 3D chromatin contacts and that a substantial portion of previously unexplained GWAS associations may be attributable to eRNAs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4071F: Genome-wide association study identified novel structural variants associated with Alzheimer's disease pathology in Non-Hispanic Whites ⭐

Authors:

T. Gunasekaran[1], D. Reyes-Dumeyer[2,3], A. Corvelo[4], W. E. Clarke[4,5], E. S. Uday[4], M. Byrska-Bishop[4], A. O. Basile[4], A. S. Runnels[4], R. S. Musunuri[4], G. S. Narzisi[4], K. M. faber[6,7,8], A. M. Goate[9,10], B. F. Boeve[11], C. Cruchaga[12], M. Pericak-Vance[13], J. Haines[14], R. N. Rosenberg[15], D. W. Tsuang[16], D. Mejia[17,18], M. Medrano[19], R. A. Lantigua[2,20], R. A. Sweet[21], D. A. Bennett[22], W. S. Robert[22], T. M. Foroud[8,23], C. L. Dalgard[24], R. Mayeux[1,2,3], M. C. Zody[4], B. N. Vardarajan[1,2,3]; [1]The Gertrude H. Sergievsky Ctr., Coll. of Physicians and Surgeons, Columbia Univ., New York, NY, [2]The

Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Vagelos Coll. of Physicians & Surgeons, Columbia Univ., New York, NY, [3]Dept. of Neurology, Vagelos Coll. of Physicians and Surgeons, Columbia Univ., and the New York Presbyterian Hosp., New York, NY, [4]New York Genome Ctr., New York, NY, [5]Outlier Informatics Inc.,, Saskatoon, SK, Canada, [6]Natl. Centralized Repository for Alzheimer's Disease and Related Dementias (NCRAD), Indianapolis, IN, [7]Indiana Univ. Sch. of Med., Indianapolis, IN, [8]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med., Indianapolis, IN, [9]Ronald M. Loeb Ctr. for Alzheimer's Disease, Friedman Brain Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [11]Dept. of Neurology, Mayo Clinic, Rochester, NY, [12]Dept. of Psychiatry, Washington Univ., Saint Louis, MO, [13]John P. Hussman Inst. for Human Genomics, Miller Sch. of Med., Univ. of Miami, Miami, FL, [14]Case Western Reserve Univ. Sch. of Med., Dept. of Population & Quantitative Hlth.Sci., Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [15]Dept. of Neurology, UT Southwestern Med. Ctr., Dallas, TX, [16]Geriatric Res., Ed., and Clinical Ctr., VA Puget Sound Hlth.Care System, Seattle, WA, [17]CEDIMAT, Santo Domingo, Dominican Republic, [18]Univ. Pedro Henríquez Urena, Santo Domingo, Dominican Republic, [19]Pontificia Univ. Catolica Madre y Maestra, Santiago, Dominican Republic, [20]Dept. of Med., Vagelos Coll. of Physicians & Surgeons, Columbia Univ.,, New York, NY, [21]Dept. of Psychiatry and Neurology, Univ. of Pittsburgh, Pittsburgh, PA, [22]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, [23]Indiana Univ. Sch. of Med./NCRAD, Indianapolis, IN, [24]Dept. of Anatomy, Physiology, and Genetics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Abstract:

Background: Alzheimer's Disease (AD) is characterized by genetic heterogeneity and there is no single model explaining the genetic mode of inheritance. To date, more than 70 genetic loci associated with AD have been identified but they explain only a small proportion of AD heritability. Structural variants (SVs) may explain some of the missing AD heritability, and specifically, their association with AD pathology and intermediate molecular phenotypes such as gene and protein expression has yet to be investigated. **Methods:** SVs were called in 1,162 individuals (687 cognitively normal and 475 AD patients) from the ROSMAP study, an independent familial AD cohort of 197 NHW families (926 subjects, 58.5% with AD), and 214 CH families (1,340 subjects, 59.17% with AD). We used Manta, Absinthe, and MELT to identify large insertions/deletions from whole genome sequencing, augmented with Sniffles2 SV calls from 4 ONT-sequenced genomes and an external SV call set from HGSVC on 32 PacBio-sequenced genomes from the 1000 Genomes Project, genotyped across the cohort with Paragraph. This produced a unified VCF with 45,251 insertions and 76,566 deletions. After excluding SVs with missing data in ≥2% of samples and a minor allele count ≤5, 9,290 insertions and 9,529 deletions were analyzed. Associations of SVs with clinical AD, pathological AD status, Braak

stages, amyloid levels, tangles, NFTs, and plaques were examined, using age, sex, and three principal components as covariates. SVs were validated by segregation among AD patients in NHW and CH families. **Results:** We identified five insertions and ten deletions associated with AD pathological measures after multiple corrections (FDR<0.1). An insertion in the *LRP1B* gene (chr2:141129952:176) was associated with increased amyloid levels (P=5.14×10^{-06}) and segregated with clinical AD in four NHW families. Deletions in the *KIR3DL2* (chr19:54852574:-14949) (P=5.35×10^{-06}), *MYPN* (chr10:68202736:-122) (P=1.70×10^{-05}) and *TTC27* (chr2:32647792:-1642) (P=3.70×10^{-05}), were associated with NFT and segregated with clinical AD in NHW families. The *KIR3DL2* deletion that results in loss of an exon and *TTC27* deletion regulate cis-expression levels in the dorsolateral prefrontal cortex. In addition, seven deletions showed associations with tangles, neuritic and diffused, plaques after FDR corrections. **Conclusion:** Our findings indicate that SVs play a significant role in AD pathology and segregate within AD families. These results provide new potential targets for therapeutic interventions, as SVs, due to their larger size, can have a more substantial impact on gene function or expression compared to SNPs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4072F: Genome-wide association study of protein-altering and regulatory variants with resistance to *Mycobacterium tuberculosis* infection

Authors:

C. Conil[1,2], J. Bohlen[1,2], M. Chaldebas[3], M. Dallmann-Sauer[4,5,6], V. M. Fava[4,5,6], E. E. Kroon[7], M. Möller[7], S. Boisson-Dupuis[3], E. G. Hoal[7], M. A. Jean-Juste[8], D. W. Fitzgerald[9], J-L. Casanova[1,2,3,10], L. Abel[1,2,3], E. Schurr[4,5,6], A. Cobat[1,2,3]; [1]Lab. of Human Genetics of Infectious Diseases, Necker Branch, Inserm U1163, Necker Hosp. for Sick Children, Paris, France, [2]Paris Cité Univ., Imagine Inst., Paris, France, [3]St. Giles Lab. of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller Univ., New York City, NY, [4]Dept. of Biochemistry, Faculty of Med., McGill Univ., Montreal, QC, Canada, [5]Program in Infectious Diseases and Global Hlth., The Res. Inst. of the McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [6]McGill Intl. TB Ctr., Dept. of Med., Faculty of Med., McGill Univ., Montreal, QC, Canada, [7]DSI-NRF Ctr. of Excellence for BioMed. Tuberculosis Res., South African Med. Res. Council Ctr. for Tuberculosis Res., Div. of Molecular Biology and Human Genetics, Faculty of Med. and Hlth.Sci., Stellenbosch Univ., Cape Town, South Africa, [8]Haitian Study Group for Kaposi's Sarcoma and Opportunistic Infections (GHESKIO), Port-Au-Prince, Haiti, [9]Ctr. for Global Hlth., Weill Cornell Med., New York City, NY, [10]Howard Hughes Med. Inst., New York City, NY

Abstract:

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (*Mtb*), is a major cause of morbidity and mortality worldwide. The existence of genetic factors predisposing to the development of clinical forms of the disease is now well established. However, there is another level of interindividual variability, much less studied, at the initial stage of infection by *Mtb*. Despite intense exposure to *Mtb*, some individuals seem naturally resistant to infection and remain negative to tuberculin skin test (TST) and/or interferon-gamma release assays (IGRAs). Few studies have attempted to unravel the genetic factors underlying resistance to *Mtb* infection. We performed a whole genome sequencing study of individuals with extreme phenotypes of resistance who remained TST and IGRA negative (so-called resisters) despite high vulnerability due to HIV infection and high levels of exposure to *Mtb* due to living in area highly endemic for TB. We enrolled 55 resisters and 100 *Mtb* infected people living with HIV (PLWH) from South Africa and 66 resisters and 57 *Mtb* infected PLWH from Haiti. We performed a genome-wide association study (GWAS) of resistance to *Mtb* with candidate common variants (frequency > 5% in the African population of the gnomAD database), defined as protein altering variants or non-coding regulatory variants affecting expression (eQTLs) and splicing (sQTLs) of at least one gene encoding a protein in the lungs or whole blood. We identified a locus on chromosome 12 of 7 SNPs associated with resistance to infection in South Africa and replicated in Haiti (combined OR[95%CI]=0.22[0.13-0.39], *P-value* = 1.03×10^{-7} for a genetic additive model). We further showed that the *Mtb* resistance allele was significantly depleted from our in-house African TB cohort (9.1% of homozygotes in 121 TB *vs.* 16.4% in 586 general population, *P-value* = 0.036) and UK Biobank European individuals with history of TB (6.4% of homozygotes in 581 TB *vs.* 9.3% in 107,350 general population, *P-value* = 0.017). The main candidate causal variant suppresses an upstream open reading frame (uORF) in the 5' untranslated region of the downstream gene. We showed experimentally that suppression of this uORF increases the activity of the downstream gene *in vitro*. This gene is involved in innate lymphoid cells differentiation, a key protective subset of immune cells in the defense against *Mtb*. This project identified a new candidate gene for studying the different mechanisms involved in the initial phase of *Mtb* infection. Ultimately, these results will help identify pathways towards new therapeutics aimed at restoring a partially deficient immune response.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4073F: Genome-wide comparison of allele specific gene expression in Non-Hispanic White, African American, Caribbean Hispanic, and Peruvian populations ⭐

Authors:

E. Vazquez-Hidalgo[1], M. Mews[1], N. Wheeler[2], P. Benchek[1], T. Starks[3], G. Byrd[4], L. Adams[5], M. Cornejo-Olivas[6], J. Haines[7], M. A. Pericak-Vance[5], A. Griswold[5], W. Bush[7]; [1]Case Western Reserve Univ., cleveland, OH, [2]Case Western Reserve Univ., Cleveland, OH, [3]Wake Forest Sch. of Med., Winston-Salem, NC, [4]Wake Forest Univ., Winston-Salem, NC, [5]Univ. of Miami, Miami, FL, [6]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [7]Case Western Reserve Univ, Cleveland, OH

Abstract:

Genetic ancestry is associated with genetic and phenotypic variation contributing to diversity among human populations. Another key contributor to phenotypic diversity is the imbalanced expression of the two inherited copies of each gene, which can play a significant role in disease susceptibility. Allele specific expression (ASE) analyses quantify the relative levels of gene expression in diploid individuals, providing new insights into genomic variation and their potential impact on disease susceptibility. However, much like studies of disease, studies of ASE in diverse populations have been limited, and there is very little understanding of how genetic ancestry influences ASE. To investigate differences in ASE across populations, we analyzed whole-blood RNA-seq from Non-Hispanic White (NHW, n=212), African American (AA, n=201), Caribbean Hispanic (CHIS, n=295), and Peruvian (PER, n=84) participants. We used phased genotype data, ancestry proportions estimated from FLARE, and phASER to quantify ASE. To understand genome-wide diversity in gene expression, each participant was evaluated for the number of genes containing at least one RNA-seq read for heterozygous alleles (median NHW=34165 , AA=16988 , CHIS=15138 , and PER=13909). Each participant was further evaluated for genes exhibiting ASE. Median ASE gene count for NHW=4384, AA=6507, CHIS=6104, and PER=6171. NHW showed significantly lower numbers of genes with ASE (Wilcoxon $p<2.3x10^{-10}$) when compared to AA, CHIS, and PER. There are no significant differences between AA, CHIS, and PER. As there is considerable genetic admixture in these populations, we investigated how the proportion of African and Amerindian ancestry related to the number of ASE genes within each individual. Linear regression models showed a significant positive correlation between African ancestry ($p=3.26x10^{-13}$) and Amerindian ancestry (p=0.037) and ASE. Possibly due to biases in imputation accuracy for rare variants by ancestry, we observed higher numbers of heterozygous variants in NHW compared to other populations. Despite this bias, we observed larger numbers of statistically significant genes with ASE in AA, HISP, and PER populations. We anticipate that increased variability in the expression of alleles

within non-European-descent populations will impact risk prediction modeling and if not accounted for, may impact statistical power to detect genetic variants through association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4074F: Genome-wide investigation of genetic variants that function as both eQTLs and pQTLs independently and via meta analysis

Authors:

J. Landman[1], W. Zhu[2], L. Petty[1], E. Frankel[2], R. Roshani[2], H-H. Chen[3], M. Yaser[4], M. Graff[5], M. Krishnan[6], V. Buchanan[7], M. Lee[8], A. Gutierrez[9], K. Young[5], J. McCormick[10], S. Fisher-Hoch[11], K. North[5], H. Highland[12], J. Below[13]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Univ., Nashville, TN, [3]Academia Sinica, Taipei, Taiwan, [4]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [5]Univ North Carolina, Chapel Hill, NC, [6]Univ. of North Carolina, Chapel Hill, NC, [7]Univ. of North Carolina - Chapel Hill, Westfield, IN, [8]The Univ. of Texas Hlth.Sci. Ctr. at Houston Sch. of Publ. Hlth., Brownsville, TX, [9]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Brownsville, TX, [10]UTH SPH Brownsville, Brownsville, TX, [11]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [12]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [13]Vanderbilt Univ Med Ctr., Nashville, TN

Abstract:

Background: Identification of expression quantitative trait loci (eQTLs) has enabled several recent advances, including improved understanding of gene regulation, functional characterization of GWAS loci, and novel gene discovery methods, such as imputation of genetically regulated gene expression levels. Similar approaches have also begun to be applied to identify and leverage protein QTLs (pQTLs). However, there is little work that examines the impact of genetic variants across the transcriptional-translational pipeline. Using genetic, transcriptomic, and proteomic data from the Cameron County Hispanic Cohort (CCHC), we identified genetic variants that are both eQTLs and pQTLs in blood. We hypothesize that these highly functional variants may have important consequences for precision medicine and uncovering molecular mechanisms of disease.
Methods: We used *fastqtl* to identify cis-QTLs for both gene expression and protein expression in the CCHC cohort. We used two non-overlapping datasets of 1,288 and 261 participants, respectively, to identify eQTLs and pQTLs for 2,252 protein-coding genes. We then meta-analyzed the eQTL and pQTL datasets using *METAL* to identify QTLs with significant effects on both transcription and translation.
Results: In our dataset, we found 217,498 eQTLs (per-gene mean = 96.58, sd = 187.90) and

53,817 pQTLs (per-gene mean = 23.90, sd = 98.33); 1,312 genes had no eQTLs or pQTLs. Of those, 16,487 were independently statistically significant as both eQTLs and pQTLs following false discovery rate (FDR) p-value correction (p < 0.05). Of those e+pQTLs, 12,819 had the same direction of effect across 186 genes, while 3,668 had opposite directions of effect across 74 genes. Following the meta-analysis of both datasets with *METAL*, 229,653 variants were identified as statistically significant following FDR correction ($p < 0.05$); 1,280 genes had no eQTLs or pQTLs.
Conclusion: This approach details a method for identifying variants that function as both eQTLs and pQTLs. It provides a potential avenue forward for future research toward better understanding how those e+pQTLs may be associated with specific health and disease outcomes via genetic studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4075F: Genome-Wide Linkage and Association of Novel Genes and Pathways with Type 2 Diabetes in Italian Families

Authors:

M. Amin[1], C. Gragnoli[2]; [1]Faculty of Med., Univ. of Khartoum, Khartoum, Sudan, [2]Creighton Univ., Omaha, NE

Abstract:

Background: Type 2 diabetes mellitus (T2D) stands as one of the most prevalent chronic diseases globally, posing substantial health and economic burdens on society. Within the spectrum of T2D, familial cases emerge as a distinct entity characterized by a strong familial clustering of the disease. This phenomenon has long suggested that genetics contributes substantially to T2D susceptibility, motivating extensive research into the genetic determinants of familial T2D. **Methods:** We recruited 212 multigenerational Italian families with multiple cases of T2D. The families were genotyped using genomic array (≥ 600k) derived from the UK Biobank Axiom Array platform. Informative markers were tested via Pseudomarker for linkage to and linkage disequilibrium (i.e., linkage joint to association) with T2D across the following models: dominant with complete penetrance (D1), dominant with incomplete penetrance (D2), recessive with complete penetrance (R1), and recessive with incomplete penetrance (R2). **Results:** We identified a total of 566 variants reaching genome-wide significant ($P < 0.00005$) linkage and/or association to/with the risk of T2D in Italian families. Of the 356 genes identified in our study, 343 (96%) are novel and have not been reported with T2D or any of its related phenotypes (i.e., obesity, metabolic syndrome, insulin resistance, polycystic ovary

syndrome, hyperglycemia). **Conclusion:** Our study constitutes the first familial T2D-linkage and association study in the Italian population. However, the functional relevance of the novel variants and genes reported in our study remains to be explored.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4076F: Genome-wide meta-analysis derived polygenic risk score predicts chronic back pain

Authors:

R. Osagie, G. K. Farhood, M. Parisien, C. B. Meloto, A. V. Grant; McGill Univ., Montréal, QC, Canada

Abstract:

Background/Aim: Back pain is the leading cause of disability worldwide. The most debilitating type of back pain is chronic back pain (CBP) defined as back pain>=3months. A previous PRS was found to predict CBP with only a slight improvement over chance. The predictive power of a polygenic risk score (PRS) depends on the sample size of the base dataset. We therefore aimed to construct an informative PRS for CBP derived from genome-wide association study (GWAS) meta-analysis, maximizing on sample size. **Methods**: We performed meta-analysis of GWASs from 3 cohorts (160,251 cases & 568,388 controls). CBP cohorts: 1. UKB & 2. CHARGE and dorsalgia cohort (genetically correlated with CBP in 1&2) 3. FinnGen. We applied the weighted Z-score method based on sample size via METAL which addresses sample overlap between our input GWAS data (1&2). Standard meta-analysis quality control was done resulting in 28,010,544 overlapping markers. Additionally, we removed markers with high heterogeneity ($p<0.05$, & $I2>80$). With the Clump+Threshold method using PRSice2, we then derived a PRS. Next, we derived a PRS using only one CBP cohort (1) as base data. The target dataset used for deriving both PRSs comprised participants of European descent from the CLSA. We defined CBP in the CLSA as those who had back pain ≥ 3 months, and controls were those usually free of pain/discomfort and had no history of back pain lasting >= 1month (3,794 cases & 10,769 controls). We adjusted for the first 20 principal components, province of residences, age, age2, and sex. We assessed discrimination and calibration using the area under the curve (AUC) and loess smoother calibration curve, respectively. **Results**: Individuals from the 99th percentile of the GWAS meta-analysis derived- and UKB derived PRS distributions showed an increased risk for CBP (OR = 2.11, 95% CI: 1.50-2.97; OR = 2.10, 95% CI: 1.49-2.95, respectively). The performance for the full models of PRS derived by GWAS meta-analysis and UKB GWAS respectively was 3.70% & 3.58% (at p-value thresholds of 0.3565 & 0.4775), with AUC (95% CI)

of 0.603(0.593-0.613) and 0.602(0.592-0.612). Visually, the loess calibration curve was qualitatively better for the GWAS meta-analysis derived PRS, demonstrating a higher degree of consistency between the predicted and observed probabilities. **Conclusion**: Overall, GWAS meta-analysis derived PRS showed similar performance compared to that of a single cohort. Meta-analysis derived PRS showed greater potential for accuracy and reliability when applied to other cohorts. By adding biobank resources to the meta-analysis, we expect further improvement. A high-performing PRS for CBP has the potential to impact public health.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4077F: gmmcoda: Graphical model for the mixture of compositional data and absolute abundance data with applications to microbiome studies

Authors:

H. Fang, S. Zhang, T. Hu; Capital Normal Univ., Beijing, China

Abstract:

Probabilistic graphical models provide efficient approaches to exploring the relationship of variables in various applications. Gaussian graphical model (GGM) is popular for constructing the conditional dependence network of interested variables. However, GMM is inappropriate in some applications such as microbiome studies, in which only relative abundances (referred to as compositional data in statistics) can be observed for some variables. Recently, some algorithms have been proposed to deal with the graphical modeling problem for compositional data. Nevertheless, there is a lack of statistical methods for inferring the interaction network for the mixture of compositional data and absolute abundance data. In this study, we propose a probabilistic graphical model for modeling interactions of variables for the mixture of compositional data and absolute abundance data. A novel maximum penalized likelihood estimator, called gmmcoda, is introduced for inferring the network from the mixture data. We develop a majorization-minimization algorithm to solve the optimization problem involved in gmmcoda. The performance of gmmcoda is evaluated and compared with other existing methods by simulation studies. Additionally, we apply gmmcoda to one microbiome data including microbial abundance as compositional data and gene expression data as absolute abundance data. The microbe-gene interactions detected by gmmcoda are further validated using previous studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4078F: GWAS of 59,102 glaucoma cases reveals 230 independent loci and highlights pathways involved in glaucoma

Authors:

X. Zhang[1], N. Parikshak[1], R. Wang[1], GHS-RGC DiscovEHR Collaboration, MAYO-RGC Project Generation, O. Melander[2], Regeneron Genetics Center, Penn Medicine Biobank, UCLA-RGC ATLAS Collaboration, FinnGen, Colorado Center for Personalized Medicine - RGC collaboration, UK Biobank, G. Coppola[1], E. Jorgenson[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Malmo Univ Hosp, Malmo, Sweden

Abstract:

Glaucoma is a leading cause of irreversible blindness worldwide. Elevated intraocular pressure (IOP) is the major modifiable risk factor in most glaucoma cases and has been linked to retinal ganglion cell (RGC) death. Current treatments can reduce IOP and reduce damage to RGCs, but no treatments to date increase RGC resilience or restore RGC function. Human genetic studies have the potential to identify new therapeutic targets that could address this need. In this study, we performed the largest GWAS for glaucoma to date, which consists of 59,102 cases (38,198 European and 20,904 Finnish cases) and 1,190,670 controls. We identified 230 independent common variant loci - nearly doubling the number of loci discovered by the largest previous study (Gharahkhani, P. *et al.* 2021). Out of these independent loci, 138 our loci are novel, and 81 overlap with loci previously associated with GWAS of IOP. For example, previous studies have shown that rare LoF variation in *ANGPTL7* reduces IOP. Our increased sample size and inclusion of FinnGen enabled identification of a common variant *rs147660927* [p.Arg220Cys, MAF EUR = 0.01%, MAF FIN = 4.1%] in *ANGPTL7*, which is associated with 31% lower disease risk (p-value = $2.2x10^{-38}$). To further understand the contribution of diverse molecular pathways and potential mechanisms beyond IOP that might contribute to glaucoma genetics, we used Multi-marker Analysis of GenoMic Annotation (MAGMA) analysis to search for significantly enriched pathways. Our analysis highlighted the following pathways: circulatory system development and cardiovascular system pathways ($P=7.3x10^{-9}$), molecules associated with elastic fibers ($P=8.2x10^{-8}$), response to growth factors, ($P=1.0x10^{-7}$) and enzyme linked receptor protein signaling pathway ($P=6.7x10^{-7}$). Our study expands the spectrum of variation linking glaucoma to IOP, supports known roles of inflammation and extracellular matrix biology, and suggests new protein signaling pathways that might play a role in glaucoma.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4079F: High Prevalence of *ALPK3* Premature Terminating Variants in Korean Hypertrophic Cardiomyopathy Patients

Authors:

S. Ryu[1], W. Jeong[1], G. Hong[2], J. Cho[3], S. Lee[4], H. Kim[5], J. Jang[6], S. Lee[7], D-H. Bae[8], J. Cho[9], J. Kim[10], K-H. Kim[11], J. Son[12], B. Han[1], G. Seo[1], H. Lee[1]; [1]3billion inc, Seoul, Korea, Republic of, [2]Div. of Cardiology, Yonsei Univ. Coll. of Med. Severance Hosp., Seoul, Korea, Republic of, [3]Div. of Cardiology, Daejeon St. Mary's Hosp., Coll. of Med., The Catholic Univ. of Korea, Daejeon, Korea, Republic of, [4]Div. of Cardiology, Dept. of Internal Med. and Res. Inst. for Convergence of BioMed. Sci. and Technology, Pusan Natl. Univ. Yangsan Hosp., Pusan Natl. Univ. Sch. of Med., Yangsan, Korea, Republic of, [5]Div. of Cardiology, Dept. of Internal Med., Keimyung Univ. Dongsan Med. Ctr., Daegu, Korea, Republic of, [6]Div. of Cardiology, Dept. of Internal Med., Gyeongsang Natl. Univ. Sch. of Med., Gyeongsang Natl. Univ. Changwon Hosp., Changwon, Korea, Republic of, [7]Div. of Cardiology, Dept. of Internal Med., Jeonbuk Natl. Univ. Med. Sch., Jeonbuk Natl. Univ. Hosp., Jeonju, Korea, Republic of, [8]Div. of Cardiology, Dept. of Internal Med., Chungbuk Natl. Univ. Hosp., Cheongju, Korea, Republic of, [9]Dept. of Cardiovascular Med., Chonnam Natl. Univ. Hosp., Chonnam Natl. Univ. Med. Sch., Gwangju, Korea, Republic of, [10]Coll. of Med., The Catholic Univ. of Korea, Seoul, Korea, Republic of, [11]Div. of Cardiology, Incheon Sejong Hosp., Incheon, Korea, Republic of, [12]Div. of Cardiology, Dept. of Internal medicine, Yeungnam Univ. Hosp., Daegu, Korea, Republic of

Abstract:

Hypertrophic cardiomyopathy (HCMP) is a disease affecting approximately 1 in 300 people in South Korea. HCMP is characterized by thickening of the myocardial ventricular walls and is both phenotypically and genetically heterogeneous. Although more than 60 different genes associated with HCMP have been identified, molecular diagnosis is established in only 30% of cases, suggesting that additional genetic loci and disease mechanisms remain to be discovered for known genes. The alpha protein kinase 3 (*ALPK3*) gene associated with autosomal recessive familial hypertrophic cardiomyopathy-27 (OMIM: 618052) is an atypical protein kinase thought to play an essential role as a transcription factor during cardiomyocyte differentiation. Patients with autosomal recessive familial hypertrophic cardiomyopathy-27 have early-onset severe dilated cardiomyopathy in infancy that progresses to HCMP over time. Although *ALPK3* is associated with autosomal recessive HCMP, several recent studies have reported that monoallelic premature termination variants (PTVs) in *ALPK3* confer an increased risk of adult-onset autosomal dominant hypertrophic cardiomyopathy (HCMP). However, these studies were conducted in patient cohorts of predominantly European Caucasian background.To determine whether these findings are replicable in a Korean HCMP cohort, we compared 2,366 Korean patients with nonsyndromic HCMP to three independent population databases: the East Asian

population from the gnomAD v4.0 database, the Korean Genome and Epidemiology Study (KoGES) database, the Korean Variant Archive (KOVA2) database, and the in-house Korean exome database. We observed that monoallelic PTV of *ALPK3* was significantly enriched in Korean HCMP patients, with an odds ratio of 15.33, similar to the odds ratio of 16.11 reported in a previous study. Among 2,366 unrelated individuals with HCMP patients studied, heterozygous *ALPK3* PTVs were identified in ~2% of all unexplained Korean HCMP cases. Thus, the HCMP-associated risk for *ALPK3* does not seem to be restricted to specific ethnic groups. In conclusion, our study suggests that *ALPK3* PTV carriers are at high risk for developing adult-onset HCMP. Therefore, diagnostic laboratories should consider reporting monoallelic *ALPK3* PTVs when the patient is suspected of having non-syndromic HCMP.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4080F: HLA region local heritability and genetic correlation reveals novel genetic pattern underlying complex traits across multiple ancestries

Authors:

H. Gui[1], M. Wei[2], **Y. Liu**[3], Z. Meng[3], I. Adrianto[3], A. Levin[4], M. LI[5], Q. Wang[2]; [1]Henry Ford Hlth.-Michigan State Univ., Detroit, MI, [2]Sichuan Univ., Chengdu, China, [3]Henry Ford Hlth., Detroit, MI, [4]Henry Ford Hlth System, Detroit, MI, [5]Sun Yat-Sen Univ., Guangzhou, China

Abstract:

Background: The human leukocyte antigen (HLA) region (~5 Mb) has been associated with multiple complex traits across physical and mental health. However, advancement of genomic knowledge for HLA region lags that for other regions. This study aims to better understand its contribution to complex traits by estimating and comparing local heritability and genetic correlation. **Methods**: Access to summary statistics for genome-wide association studies (GWAS) were approved by Million Veteran Program via dbGap. In total 11 common phenotypes (e.g., body mass index [BMI], lipid traits, and mental conditions) were included, covering >500,00 individuals with 3 different ancestries (European, African, and Hispanic). Whole genome was divided into 550 regions (including HLA) with 5 Mb size, and local heritability was calculated using effective heritability estimator implemented in KGGSEE; global heritability was calculated using LDSC or LDAK with overall GWAS summary statistics. For cross-phenotype, LDSC or LAVA was used to estimate global or local genetic correlations. Linkage disequilibrium data was from ancestry-matched population in 1000 Genome. For each ancestry, HLA region local heritability per phenotype and genetic correlation per phenotype-pair were compared with

other regions by Monte Carlo procedure. Across ancestries, Wilcoxon signed-rank test was used to compare local heritability across different regions. Bonferroni method was used for multiple testing correction. **Results**: Estimated heritability of HLA region ranged from 0.0001 (behavioural ideal health score in Africans) to 0.013 (low-density lipoprotein in Hispanics), while the median of estimated heritability of other regions ranged from 0.0001 to 0.0028 for 11 phenotypes across ancestries. Global heritability ranged from 0.0368 (suicidal ideation in Europeans) to 0.7332 (BMI in Africans), and the percentages for HLA region ranged from 0.11% to 5.47%. When comparing to others, HLA region contributed to significant higher heritability for 8 traits in Europeans and 2 traits in Hispanics. The distributions of estimated local heritability in EA and AA are significantly different across all 11 phenotypes. When looking into local genetic correlations at HLA region, estimates from 11 or 9 phenotype pairs were significant for Europeans or Africans, respectively; among them, only a few of them have also significant global genetic correlations. **Discussion**: The heritability and genetic correlation estimated for HLA varies differently across 11 complex phenotypes in 3 ancestry groups. This new pattern will be helpful to better understand complex traits from HLA-mediated immune pathways.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4081F: Human protein-small molecule interaction networks reveal the cross-talks between genomic components and metabolome across complex metabolic traits

Authors:

V. Akbary Moghaddam[1], S. Acharya[1], S. Liao[1], W. Jung[1], B. Thyagarajan[2], K. Christensen[3], M. R. Brent[1], G. J. Patti[1], P. An[1], M. A. Province[1]; [1]Washington Univ. in St. Louis, St. Louis, MO, [2]Univ. of Minnesota, Twin Cities, MN, [3]Southern Denmark Univ., Odense, Denmark

Abstract:

Small molecules (SMs) modulate biological processes through influencing macromolecules, gene expression, and epigenetic changes. However, their functions remain largely unknown due to their diverse origins and technological limitations. Here, we introduce a three-step framework to: 1. construct statistical protein-SM interaction networks (PSI-Nets), 2. enrich them with biological information, and 3. use them to study complex metabolic traits. We initially used transcriptomics, lipidomics, and metabolomics data from 3839 healthy participants of the Long Life Family Study (LLFS) to construct statistical PSI-Nets with a combined linear mixed model / correlated meta-analysis (CMA) approach following adjusting the omics measurements for technical and biological confounders. Next, we enriched gene-SM connections by known gene-

SM interactions and common gene regulatory and protein structural signatures of SM-associated genes to make biologically informed PSI-Nets. We constructed 529 PSI-Nets containing 5080 gene-SM interactions with 90% enrichment for biological features for 3347 genes and 401 SMs. Network association analysis for insulin sensitivity (HOMA2-S) in non-diabetic LLFS participants revealed nine significant PSI-Nets (*P* value < 9.45×10^{-5}). Among our major findings, using PSI-Nets' architecture and their enriched biological information, a protective effect of mast cell genes (*FCER1A*, *MS4A2*, *HDC*, *CPA3*, *GATA2*) on HOMA2-S was revealed, which can be downregulated by elevated dimethylguanidino valeric acid in higher BMI and triglyceride (TG) conditions, decreasing HOMA2-S and increasing glucose and L-alanine levels. The TG/BMI-mast cell genes-HOMA2-S axis was also replicated in the Framingham Heart Study. Lastly, combining whole genome sequencing association scans of HOMA2-S, genes, and SMs by CMA identified 26 genome-wide significant SNPs across five gene-SM pairs for HOMA2-S (*P* value < 5×10^{-8}). Notably, we found *rs6605000* (CMA *P* value = 4.17×10^{-12}) with protective effect for HOMA2-S and suggestive to significant associations for HOMA2-S-*TLR5*-leucylproline. *TLR5*-leucylproline interaction is represented in one of the significant PSI-Nets for HOMA2-S. Furthermore, *rs6605000* is also annotated to *TLR5* expression across multiple tissues in the GTEx database. Our framework provides a promising paradigm to uncover the biological effects of SMs and their role in complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4082F: Identification of microRNA-related SNPs associated with Lung, Kidney, Gastric, and Thyroid Cancer in Korean Cancer Prevention Study-II Biobank

Authors:

J-H. Lee, S. Jee; Yonsei Univ., Seoul, Korea, Republic of

Abstract:

MicroRNAs (miRNAs) are endogenous, small non-coding RNAs that play crucial roles in regulating gene expression. MiRNAs play crucial roles in cancer development and progression, and their dysregulation has been implicated in various cancer types. In recent years, there has been a surge in research on miRNA and its relationship with cancer. The expression patterns of miRNAs have been found to correlate with cancer type, stage, and other clinical variables, highlighting their potential utility as diagnostic and prognostic tools in cancer management. Furthermore, miRNA expression analyses have shed light on the oncogenic or tumor-suppressive roles of specific miRNAs, implicating them in almost all aspects of cancer biology,

including proliferation, apoptosis, invasion/metastasis, and angiogenesis. However, the expression patterns of miRNAs in different cancer types among Korean patients remains poorly understood. To address this gap, we conducted a comprehensive analysis utilizing genome-wide association study (GWAS) and meta-analysis approaches to investigate the miRNA expression patterns associated with lung cancer, kidney carcinoma, gastric cancer, and thyroid cancer in the Korean population. Our analysis identified distinct significant miRNA-related single nucleotide polymorphisms (SNPs) associated with each cancer type. Additionally, we observed the miRNAs associated with each SNPs, highlighting the genetic heterogeneity underlying cancer predisposition. Notably, hsa-miR-4999 and hsa-mir-612 demonstrated significant correlations with lung cancer, whereas hsa-mir-6730 and hsa-mir-603 exhibited associations with kidney cancer. Additionally, hsa-mir-6128, hsa-mir-6839, and hsa-mir-4779 were found to be associated with gastric cancer, while hsa-miR-564 and hsa-miR-6868 showed associations with thyroid cancer. Our study elucidates cancer type-specific miRNA and SNPs in the Korean population. We have identified miRNAs as potential biomarkers for cancer diagnosis, prognosis, and therapeutic targets among Korean patients. Such insights contribute to the development of effective approaches in cancer diagnosis and treatment. Furthermore, the identification of miRNA-cancer associations provides valuable foundations for the development of targeted therapeutic strategies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4083F: Identification of new variants associated with urinary incontinence and its subtypes: a meta-GWAS analysis

Authors:

E. Moreno[1], L. F. Thomas[1,2,3,4], B. Wolford[1], M. R. Moksnes[1], E. Coward[1], J. Horn[5,6], I. Surakka[7], K. Hveem[1], B. M. Brumpton[1]; [1]HUNT Ctr. for Molecular and Clinical Epidemiology, Dept. of Publ. Hlth.and Nursing, Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [2]Dept. of Clinical and Molecular Med., Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [3]BioCore-Bioinformatics Core Facility, Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [4]Clinic of Lab. Med., St. Olav's Hosp., Trondheim Univ. Hosp., Trondheim, Norway, [5]Dept. of Obstetrics and Gynecology, Levanger Hosp., Nord-Trøndelag Hosp. Trust, Levanger, Norway, [6]Dept. of Publ. Hlth.and Nursing, Faculty of Med., Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [7]Univ. of Michigan, Ann Arbor, MI

Abstract:

Urinary incontinence (UI) - involuntary urine leakage - is highly prevalent among women and strongly associated with poor quality of life. Two main subtypes are described: stress urinary incontinence (SUI) and urgency urinary incontinence (UUI). While the clinical subtypes differ in etiology, genetic predisposition plays a large role with an estimated heritability in twin studies of nearly 40% for both subtypes of UI. To date, four genome-wide association studies (GWAS) have been conducted, including up to 13,066 participants (SUI cases= 3,393; UUI cases= 1,614). However, the results are inconsistent, and the underlying biological mechanisms of UI remain unknown. We performed a GWAS for several UI phenotypes including up to 46,026 women (e.g. any type of UI [cases=16,726], SUI [cases=13,239] and UUI [cases=11,234]) using self-reported data from the Trøndelag Health Study (HUNT). We further conducted a fixed-effects inverse-variance weighted meta-analysis with publicly summary statistics from the Nurses' Health study (NHS) thereby including up to 56,957 individuals in the analyses. We used FUMA platform for single nucleotide polymorphism (SNP) annotation, gene prioritization, and eQTL analyses. We identified seven genetic loci (six novel) significantly associated with UI and/or its subtypes ($p<5\times10^{-8}$). For SUI, we replicated a previously identified intergenic locus (rs2270016/rs7607995) near *LBX2*, a gene known for its role in muscle cell differentiation. Also associated with SUI is a potentially novel locus (rs78579538) near *PCOLCE2* (Procollagen C-Endopeptidase Enhancer 2), a gene involved in collagen chain trimerization and in the organization of the extracellular matrix. We also found two rare variants associated with UUI. Located in an intergenic region on the chromosome 12, rs554550612 (MAF <1%) is 240 kb upstream of the transcription factor gene *SOX5* and in perfect linkage-disequilibrium (r^2=1) with an intergenic SNP situated 3kb upstream of *SOX5*. SOX family transcription factors are known to regulate cell fate during development and to potentially play a role in adult tissue regeneration and homeostasis. On chromosome 16, the rare variant rs139186923 (MAF<2%) is in a proximal enhancer-like signature, in a short interspersed nuclear element (from the Alu family). It is upstream (19.5 kb) of the *ZNF263* gene that codes for a transcription factor protein. These genetic loci are potential targets for future functional validation. Our findings contribute to a better understanding of the involvement of genetic factors in the etiology of UI while highlighting the potential implication of myogenesis, tissue regeneration and muscle support.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4084F: Identifying Causal Biomarkers via Knockoffs in the Presence of Unmeasured Confounding

Authors:

J. Gu, Z. He; Stanford Univ., Palo Alto, CA

Abstract:

Identifying causal genes or proteins for complex phenotypes is crucial for comprehending disease mechanisms and developing therapies while existing methods can only infer biomarker-phenotype associations and are vulnerable to unmeasured confounders. In this study, we present a novel framework that mitigates confounding effects when detecting causal biomarkers. Without the need to use genetic variants as instrumental variables, our framework leverages a probabilistic factor model and effectively eliminates confounding effects by simulating a set of synthetic backdoor blockers. We then select causal biomarkers using knockoffs, which seamlessly integrate with any modern machine learning model and possess rigorous error rate control. Theoretical studies establish our method's nonparametric identifiability and false discovery rate control under mild conditions. Experiments on simulated and real-world biological datasets demonstrate the empirical advantages of the proposed framework in discerning biological causality.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4085F: Identifying Disease-Relevant Cell States via Integration of GWAS and Single-Cell Omics Data Using Context-Specific Annotated Polygenic Risk Scores

Authors:

Y. Lin, L. Xu, j. hu, W. Jiang, H. Zhao; Yale Univ., New Haven, CT

Abstract:

Uncovering the disease relevant cell types is crucial for the understanding of disease mechanisms and the development of novel targeted therapies The recent application of single-cell technologies to large-cohort studies has facilitated the investigation of diverse cell types and molecular signatures across different diseases. Together with genome-wide association studies (GWAS) summary statistics, it is now possible to identify cellular subpopulations associated with genetic risk, having the potential to accurately detect risk variants causally linked to disease in a cell-type-specific manner. However, the current approaches, such as sLDSC, heavily rely on the cell type information defined in the single-cell analysis to detect cell types enriched with genetic heritability, limiting their ability to pinpoint disease-relevant cell states within major cell types. Here, we develop scGRS, a computational framework that effectively integrates information from GWAS summary statistics, linkage disequilibrium (LD), context-specific eQTL, and single-cell omics data to identify cellular subpopulations associated with disease risk. The framework

begins with estimating the effect sizes of genetic variants on diseases using Genome-Wide Empirical Bayes (GWEB), a Polygenic Risk Score (PRS) method that incorporates cell-type-specific annotations with GWAS summary statistics and LD information. scGRS then constructs a PRS-weighted score matrix by weighting single-cell gene expression data with the estimated annotation-specific PRS weight matrix that takes cell-type-specific eQTL effects into account. This cell-level PRS-weighted score matrix then allows us to perform graph clustering to identify cell states relevant to disease risk.
Applying our framework to multiple single-cell datasets from two diseases, Alzheimer's Disease (AD) and Inflammatory Bowel Disease (IBD), we demonstrate its ability to uncover cell states uniquely marked by one or two genes (e.g. RORA and IL32 in IBD; DOC2A in AD). Remarkably, some of these cell states exhibit significantly higher abundance in disease samples and are marked by genes that are closely related to the disease in the literature. Notably, these cell states are not cell-type-specific and therefore might be overlooked by conventional single-cell clustering analysis. By performing our analysis across multiple single-cell studies, we consistently identified similar cell states, demonstrating the robustness and reliability of our framework in capturing disease-relevant cellular subpopulations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4086F: Identifying single-cell eQTLs by leveraging cell type specificity and similarity and incorporating linkage disequilibrium

Authors:

C. Lin, W. Li, Y. Lin, H. Zhao; Yale Univ., New Haven, CT

Abstract:

Expression quantitative trait locus (eQTL) studies aim to infer the effects of genetic variants on gene expression levels, providing insights into the genetic regulations of complex traits. Fine-mapping methods that incorporate linkage disequilibrium (LD) information further identify putative causal variants from association results. In recent years, the emergence of population-scale single-cell RNA sequencing (scRNA-seq) datasets with annotated cell types has greatly improved our understanding of cell-type-specific eQTLs. However, existing strategies often neglect the shared genetic effects among similar cell types, potentially leading to a loss of statistical power.
In this study, we propose CASE, a statistical framework that leverages cell type specificity and similarity with an Empirical Bayesian model to identify cell-type-specific and shared eQTLs. Our framework begins by separately analyzing each cell type using MatrixEQTL with a pseudo-bulk

aggregation of single-cell data. The estimated eQTL effect sizes are then modeled in a mixture matrix normal distribution, considering the LD and various cell type similarity structures, where the priors are fitted using a Monte Carlo Expectation-Maximization (MCEM) algorithm. With the fitted priors, we can obtain the posterior probabilities of the variants having nonzero eQTL effect sizes. Compared to mvSuSiE, another multi-trait fine-mapping method, our method better characterizes cell type shared structures by calibrating the LD mix-up.
Through extensive simulation studies, we demonstrated that CASE outperformed other single-trait and multi-trait methods in terms of F-scores (at a significance level of 0.05) and precision-recall curves. Additionally, we observed that the performance of CASE improved as sample sizes and cis-heritability increased. In the application to the OneK1K scRNA-seq data with peripheral blood samples from 981 donors, CASE showed better performance compared to other methods like mvSuSiE by identifying more eGenes (genes associated with at least one significant eQTL) and eQTLs. Notably, CASE identified more weak signals and mitigated the false discovery issue in mvSuSiE, where distinct cell types tended to align with similar significance. These findings underscore the importance of leveraging cell type specificity and similarity and incorporating LD structures when studying the regulatory mechanisms underlying complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4087F: Identity-by-descent mapping using multi-individual IBD with empirical genome-wide multiple testing adjustment

Authors:

R. Cai, S. Browning; Univ. of Washington, Seattle, WA

Abstract:

Despite the many common genetic variants identified through genome-wide association studies (GWAS), the genetic architecture of rare variants underlying complex traits remains largely unknown. When testing for rare variants that are usually untyped or have minimal individual effect sizes, it is important to consider linkage and co-inheritance relationships between genetic markers, which are naturally revealed through identity-by-descent (IBD) segments. By searching for genomic positions where levels of IBD sharing are associated with phenotypic variation within a group of individuals for a trait of interest, IBD mapping offers a promising opportunity to identify the effects of rare causal variants underlying complex traits. Using IBD haplotypes detected among a group of individuals may enhance power for IBD mapping due to more accurate IBD calls and the additional information provided by specific IBD group effects that are ignored in pairwise IBD analysis.

In this work, we develop an IBD mapping test using multi-individual IBD sharing between distantly related individuals in large, outbred populations. Motivated by variance component models in linkage analysis, we characterize genetic similarities between individuals at consecutive locations across the genome by constructing local relatedness matrices from multi-individual IBD sharing inferred with the software ibd-cluster [1]. We then employ a likelihood ratio statistic to test the contribution of local variations in genetic similarities to the total phenotypic variation of a complex trait of interest. Additionally, we develop an approach to adjust for multiple testing based on an analytical formula of the correlation between test statistics across the genome. Through simulation studies, we demonstrate that our test has a well-controlled family-wise type-I error rate, superior power to detect rare and untyped variants in SNP array data compared to standard single-variant tests in GWAS, and enhanced numerical stability compared to a recently published IBD mapping test by Chen et al. [2].

[1] Browning, S. R., & Browning, B. L. (2024). Biobank-scale inference of multi-individual identity by descent and gene conversion. *The American Journal of Human Genetics, 111*(4), 691-700.
[2] Chen, H., Naseri, A., & Zhi, D. (2023). FiMAP: A fast identity-by-descent mapping test for Biobank-scale cohorts. *Plos Genetics, 19*(12), e1011057.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4088F: Improved Genetic Association analysis identifies 29 novel loci influencing preclinical to disease Progression in lupus ⭐

Authors:

L. Wang, H. Markus, D. Liu; Penn State Coll. of Med., Hershey, PA

Abstract:

Genome-wide association study (GWAS) have successfully discovered many loci for autoimmune diseases, which implicates new biology and lead to new drugs. Many autoimmune diseases have a preclinical phase, where individuals have a subset of symptoms but do not meet all the criteria for disease diagnosis. Only a subset of preclinical individuals will progress to disease phase, while others may stay preclinical or get back to normal. Given that autoimmune disease may quickly cause irreversible organ damage, it is important to identify genetic variants that influence preclinical to disease progressions.
Biobanks with genetic data and electronic health records becomes a valuable resource to identify preclinical individuals and study disease progressions. Yet, compared to large case-control GWAS, biobanks often only have a limited number of disease cases and can be underpowered. To address this challenge, we propose to a novel Bayesian model GAP (genetic

association with progression) to decompose the genetic effects influencing control vs case status into the sum of the effects influencing control to preclinical (stage 1) and that influencing preclinical to disease (stage 2). This insight can leverage the large sample sizes of the case-control studies and substantially improve the estimates of genetic effects affecting both stages. As an application, we apply GAP to study systemic lupus erythematosus (SLE). Individuals in preclinical stage were identified through a positive antinuclear antibody (ANA) test (titer ≥ 1:80). Individuals with SLE were identified based on previously evaluated algorithms, which requires presence of both SLE ICD code and ANA positive test result.
We applied GAP to case-control GWAS based on meta-analysis from six European SLE studies (14,355 cases and 505,956 controls) and the BioVU preclinical GWAS (2,784 ANA+ cases and 46,174 controls) and progression GWAS (749 SLE+/ANA+ cases and 2,784 ANA+ controls). GAP identified 8 and 15 novel loci for stage 1 and 2 respectively, with 6 and 7 loci being replicated in All of Us. Transcriptome-wide association study (TWAS) also identifies an additional 4 loci for stage 1 and 5 loci for stage 2, including *BTN3A2*, known to be involved in maintaining self-tolerance, for stage 1 and *IL12RB2*, known to promote inflammation and increased in SLE patients, for stage 2.
Together we showcase a novel strategy to jointly model large case-control with small preclinical and progression studies. Results from our studies will allow identification of individuals at risk of being in preclinical stage and identify novel drug targets to prevent or slow down the transition to disease stage.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4089F: Improved heritability partitioning and enrichment analyses using summary statistics with graphREML ⭐

Authors:

H. Li[1], R. Mazumder[2], X. Lin[3], **L. J. O'Connor**[1]; [1]Harvard Med. Sch., Boston, MA, [2]MIT Sloan Sch. of Management, Cambridge, MA, [3]Harvard T.H. Chan Sch Pub Hlth, Boston, MA

Abstract:

Heritability partitioning is a powerful approach to integrate genetic association data with almost any other genomic assay, by quantifying the heritability enrichment of a derived annotation. For common and low-frequency variants, the most widely used method is stratified LD score regression (S-LDSC). This method is fast, and it can jointly analyze a large number of overlapping annotations. However, S-LDSC can have much lower statistical power, such that many enrichments may go undetected. Moreover, S-LDSC assumes an unrealistic linear relationship

between the heritability of a SNP and its annotations, leading for example to negative per-SNP heritability estimates. Intuitively, this trade-off between statistical efficiency and the convenience of using variant-level summary statistics arises from the difficulty of fully modeling linkage disequilibrium (LD), and in particular incorporating it into likelihood calculations. Recently, Nowbandegani et al. proposed LD graphical models (LDGMs), which represent LD patterns using extremely sparse matrices derived from genome-wide genealogies. LDGMs enable the use of efficient sparse matrix operations, potentially addressing the challenge of likelihood-based heritability partitioning. We propose "graphREML", a new heritability partitioning method that operates on GWAS summary statistics and LDGMs. graphREML improves upon S-LDSC by modeling the full likelihood of the summary statistics, and improves upon existing likelihood-based methods by incorporating overlapping annotations. We validate the method in simulations and in analyses of real traits, comparing its estimates with those of S-LDSC. We found that graphREML improves upon S-LDSC by modeling the full likelihood of the summary statistics. On average across traits, it produces enrichment estimates that are concordant with S-LDSC, indicating that both methods are unbiased; however, graphREML identifies a larger number of significant enrichments (2.5 times more trait-annotation pairs). graphREML is applicable to summary association statistics for almost any trait, and its statistical efficiency will enable the identification of highly specific disease relevant functional features. graphREML is robust to out-of-sample application of the LDGM precision matrices and is effective at accounting for missingness in the summary statistics. We apply graphREML to estimate the heritability enrichment of a wide range of annotations, including for diseases that are underpowered using S-LDSC. We also estimate heritability at a per-SNP level, as opposed to at an aggregate level, highlighting the advantages of our approach.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4090F: Incorporating functional annotations to improve polygenic risk prediction in admixed individuals

Authors:

B. Chen[1], F. Ockerman[1], Q. Sun[1], E. Kharitonova[1], L. Zhou[2], A. Reiner[3], H. Cao[4], R. Tao[5], Y. Li[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Indiana Univ. Sch. of Med., Carmel, IN, [3]Univ. of Washington, Seattle, WA, [4]Florida State Univ., Tallahassee, FL, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Polygenic risk scores (PRS) have been shown to improve clinical risk prediction. However, a majority of PRS studies have been done on participants with European ancestry and have been shown not transferable to other populations, which can lead to health disparities. This problem becomes even more challenging with admixed individuals, who have multiple genetic ancestries. Our previous work, GAUDI, leverages local ancestry to improve PRS in admixed individuals by modeling ancestry-specific effects through a penalized regression framework. To incorporate multiple functional annotations into GAUDI, we use annotation-specific penalties that prioritize more annotated SNPs, which are more likely to be causal. We first tested our method on admixed European and African individuals simulated using COSI. To simulate annotations, 20%, 10%, and 5% of the causal SNPs were annotated with 1, 2, or 3 random binary annotations, respectively. Additionally, to add noise, we annotated non-causal SNPs with a frequency of 10% of the number of annotated causal SNPs. Within our simulations, using annotations showed an average of 7% (2.5%) increase in testing r^2 in simulated phenotypes with 20% (60%) heritability compared to not including annotation information. Furthermore, our method outperformed two other methods, PANPRS and BFGWAS_QUANT, which incorporate multiple functional annotations but are not optimized for admixed individuals. Our method had an average of 77% (145%) increase in testing r^2 in simulated phenotypes with 20% (60%) heritability over PANPRS. Additionally, our method had an average of 132% (171%) increase in testing r^2 in simulated phenotypes with 20% (60%) heritability over BFGWAS_QUANT. For real data analysis, we predicted monocyte count (adjusted for age, sex, top 10 principal components, and center) for 7,832 self-reported Africans in the UK Biobank. Local ancestry inference was done using FLARE with a 3-way admixture consisting of European, African, and American Indian. Leveraging monocyte eQTL and ATAC-seq peak information, we see an average of 27% increase in testing r^2 in monocyte count compared to not using annotations. Moreover, our method outperformed PANPRS (which did not converge) and BFGWAS_QUANT (an average of 230% increase in testing r^2). In summary, we have developed a PRS method for admixed individuals incorporating function annotations, which show advantages in predictions compared to both our original GAUDI and other annotation-incorporated methods

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4091F: Indiscriminate usage of surrogate variables leads to overfitting in differential gene expression analyses

Authors:

D. Hoang[1], A. N. Lund[2], R. C. Thompson[1], N. D. Beckmann[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Icahn Sch. of Med., New York, NY

Abstract:

Background/Objectives: Surrogate variable analysis (SVA) aims to reduce heterogeneity in high-throughput experiments by identifying and accounting for latent confounders (surrogate variables, SVs). This approach, essential for certain analyses (e.g., eQTL), is becoming increasingly used for differential expression analyses (DEA). This study evaluates the impact of SVA on DEA, using simulated and published datasets.**Methods:** Expensive simulations (20,000 genes for 1,000 samples) with varying signal (known and unknown) to noise (2.9 x 106 parameters) were performed to characterize the impact of SVA on DEA where the true differentially expressed genes (DEGs) are known. Five DEA approaches were employed: oracle (all parameters are known), no SV adjustment, SV adjustments for: known number of SVs and number of SVs estimated by SVA “BE” and “leek” algorithms. Effect of SVA was assessed in real data by evaluating the reproducibility of DEA results with and without SVA in two large Alzheimer's disease (ROSMAP, MSBB) and schizophrenia (MSSM-Penn-Pitt, NIMH-HBCC) gene expression datasets. Gene ontology (GO) analyses were performed to characterize the effects of SVA usage on pathways enriched for DEGs.**Results:** Overall, accounting for SVs in DEA led to overfitting in simulation analyses, with a high false positive rate of DEG identification (mean = 0.22, standard deviation = 0.26). When SVA was applied to real data, the reproducibility of DEG signatures dropped substantially for Alzheimer’s disease between datasets (Spearman’s correlations without SVA rho = 0.46 and with SVA rho = 0.12), within dataset (two balanced half of the ROSMAP data; without SVA rho = 0.43 and with SVA rho = 0.16), and for schizophrenia between datasets (without SVA rho = 0.33 and with SVA rho = 0.13). While significant enrichment of distinct GO terms previously involved in Alzheimer’s disease were found in the MSBB dataset without SVA, (e.g., respiratory electron transport chain, q-value = 0.001), no significant GO terms were found in MSBB with SVA. Interestingly, although no pathways were enriched in ROSMAP without SVA, significant GO terms emerged that were not related to Alzheimer’s disease processes in ROSMAP with SVA (e.g., cell-substrate adhesion, q-value < 0.001). For schizophrenia datasets, the only differential expression signature with significant GO terms was NIMH-HBCC without SVA, with terms not related with schizophrenia (e.g., signaling, q-value <0.001).**Conclusion:** Our results of decreased of reproducibility and less informative GO terms emphasized that the use of SVA in DEA require careful consideration to avoid overfitting, cautioning against its indiscriminate usage.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4092F: Inferring causal direction between two traits using $R^2$with application to Transcriptome-Wide Association Studies (TWAS)

Authors:

H. Liao[1], X. Haoran[2], W. Pan[1]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Dept. of Biostatistics, City Univ. of Hong Kong, Hong Kong, Hong Kong

Abstract:

In the framework of Mendelian randomization (MR), two single SNP-trait Pearson correlation-based methods have been developed to infer the causal direction between an exposure (e.g., a gene) and an outcome (e.g., a trait), including the widely used MR Steiger's method and its recent extension called Causal Direction-Ratio (CD-Ratio). Steiger's method uses a single SNP as an instrumental variable (IV), while CD-Ratio combines the results from each of the multiple SNPs. Here we propose a new method based on R^2, the coefficient of determination, to combine information from multiple (possibly correlated) SNPs to simultaneously infer the presence and direction of a causal relationship. It can be regarded as a generalization of Steiger's method from using a single to multiple SNPs as IVs, making it useful in Transcriptome-Wide Association Studies (TWAS) (or similar applications) with typically small sample sizes for gene expression (or other molecular trait) data and applicable to GWAS summary-level data with a reference panel. We further discuss the influence of invalid IVs on our R^2-based method and introduce an approach called R2S to incorporate an invalid IV selection step to better control the IVs included in the model and enhance the robustness.
To illustrate the advantages of our proposed method, we compared its performance to TWAS, Steiger's method with majority voting (Steiger-MV) and CD-Ratio in simulations and in identifying causal genes for high/low-density lipoprotein cholesterol (HDL/LDL) using individual GTEx (V8) gene expression data and UK Biobank GWAS data. Our method confirmed some well-known causal genes, such as *LPL*, *LIPC*, and *TTC39B* for HDL and *MPI*, *FADS2*, and *NECTIN2* for LDL. It also identified some novel gene-trait relationships, most of which have been studied to have an association with HDL/LDL or their closely related diseases in the literature, suggesting the power gains of our method from using multiple correlated SNPs as IVs. Additionally, we showed its application with GWAS summary data through inferring causal relationships between HDL/LDL and stroke/coronary artery disease (CAD). Our proposed method correctly identified the causality from LDL to CAD and stroke and no causality from HDL to stroke but falsely identifying HDL as causal to CAD, while Steiger-MV suggested the presence of all 4 causal relationships. Furthermore, we applied the R2S method to the HDL/LDL application, which generally resulted in fewer IVs being included in the model and a smaller R^2 thereafter. The causal genes identified by R2S form a proper subset of those identified by our approach without IV selection.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4093F: Inferring relative gene-level effect sizes across multiple complex traits using GWAS summary statistics

Authors:

S. Zandavi[1], S. Yoshiji[2], E. Llamas[3], N. Burtt[4], J. Flannick[5]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Broad Inst., Cambridge, MA, [3]Broad Inst., Cambridge, MA, [4]Knowledge Portal Network, Cambridge, MA, [5]Boston Children s Hosp., Boston, MA

Abstract:

Genome-wide association studies (GWAS) have identified disease-associated variants but do not indicate the directional relationship between gene perturbation and disease risk. Here, we introduce Multi-trait Association of Gene Perturbation Inferred Effects (MAGPIE), which infers the relative effect sizes of gene perturbation across multiple traits by combining GWAS summary statistics (SNP standard error and effect size), linkage disequilibrium (LD), and SNP-to-Gene (S2G) linkage strategies. MAGPIE uses S2G to “phase” relative effect sizes, using Expectation-Maximization to maximize the likelihood of observed SNP effect sizes as a function of a) unobserved “gene perturbation effect sizes” across traits; b) unobserved trait-independent SNP effect parameters on gene expression; and c) known trait-independent S2G linkage probabilities. Simulations indicate that this method is accurate and robust. Using MAGPIE, we identified genes that matched the type 2 diabetes (T2D) “lipodystrophy” subtype, a severe insulin resistance form with detrimental cardiovascular outcomes. This subtype shows increased T2D risk, Triglycerides (TG), 2-hour Glucose (2hrG), Waist-Hip Ratio (WHR), Corrected Insulin Response (CIR), Fasting Glucose (FG), and Homeostatic Model Assessment (HOMA), and decreased Adiponectin (Adipo), Insulin Sensitivity Index (ISI), Disposition Index (DI), and High-Density Lipoprotein (HDL). By inferring relative gene-level effect sizes for these 11 traits, we recapitulated patterns for control genes like PPARG and LPL. Clustering genes according to their effect sizes across these 11 traits produced 20 clusters, including one with 193 genes matching the lipodystrophy pattern. This cluster showed enrichment for AMPK signaling (p=2.59e-4), cholesterol metabolism (p=1.25e-3), adipogenesis (p=0.018), and PPAR signaling (p=2.7e-3), processes that have been previously implicated in the T2D lipodystrophy subtype. Genes in this cluster included CMIP, IRS1, COBLL1, LIPC, LPL, CCDC92, GCK, IGF2BP2, and PEPD, previously linked to insulin resistance. Four novel genes (ZFAT, SYN2, FNDC3B, and CEBPA) have been shown to be involved in adipogenesis and adipocyte differentiation. In summary, MAGPIE, the first statistical method for inferring gene effect sizes from GWAS data, helps identify genes that

match trait signatures for disease endophenotypes. Additional applications include informing drug target validation through prediction of beneficial or detrimental secondary effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4094F: Integrating large-scale GWAS and single-cell disease atlas data to identify disease-critical cell populations corroborated by genetic associations

Authors:

H. Shi[1], O. Whitley[2], C. Simopoulos[2], D. Chang[1], R. Zhang[1], A. Stockwell[1], W. Gong[1], K. Fletez-Brant[1], L. Zou[1], T. Lucas[1], E. Kenigsberg[1], A. Missarova[1], D. Garfield[1], B. Yaspan[1], M. McCarthy[1], A. Mahajan[1], J. Kaminker[1]; [1]Genentech, South San Francisco, CA, [2]Roche, Mississauga, ON, Canada

Abstract:

Recent computational methods, CNA (Reshef 2021 *Nat Biotech*) and Milo (Dann 2021 *Nat Biotech*), detect *disease-associated* cell populations using scRNA-seq data of healthy and diseased donors. We developed a new method, single-cell Expression exPlainability Statistics (scEPS), to detect putative *disease-causal* cell populations, by identifying neighborhoods of cells where the expression of genetically implicated genes explains more disease variability than randomly selected matched control genes. At each cell neighborhood, scEPS estimates a *d statistic*, the difference between the variance in disease risk attributable to each genetically implicated and randomly selected matched control gene, using methods of moments. We confirmed via extensive simulations that scEPS provided approximately unbiased estimates of the d statistic and calibrated test statistics.

We applied scEPS to identify disease-critical cell populations for Alzheimer's Disease (AD). We used 570 (out of 18K) MAGMA (de Leeuw 2015 *Plos Compbio*) nominated genes (FDR<0.05) for AD based on summary statistics of a large-scale (N_{eff}=382K) publicly available AD GWAS (Bellenguez 2022 *Nat Genet*) to obtain scEPS d statistics for 1.3M cells (each indexing a cell neighborhood) of 24 brain cell types from 81 donors in the publicly available snRNA-seq data from the SEA-AD consortium (Gabitto 2023 *bioRxiv*). scEPS attained a significant average d statistic of 3.0×10^{-4} ($p<10^{-30}$) across all cell neighborhoods (i.e. the expression of each GWAS gene explained an additional 0.03% of variance in disease risk than a randomly selected matched control gene), recapitulating AD brain pathobiology. Across all cell types, oligodendrocyte progenitor cells (OPC) had the highest average d statistic ($d=1.3\times10^{-3}$, $p<10^{-30}$), implicating the causal role of OPC. The average d statistic was also significant for microglia

($d=6\times10^{-4}$, $p<10^{-30}$), a key cell type involved in AD risk. Despite a small overall average d statistic ($d=3\times10^{-5}$, $p=10^{-3}$), astrocytes harbored a subpopulation with strong and significant average d statistic ($d=0.0098$, $p<10^{-30}$). Pathway analysis of this subpopulation's marker genes revealed neurogenesis, neuron migration, and neuron development.
Finally, across cell types, we observed only limited Pearson's correlations between the average scEPS d statistics and the average CNA neighborhood correlations for quantifying the association between differential cell abundance and AD risk ($R=0.08$, $p=0.34$) , and the average scDRS statistics (Zhang 2022 *Nat Genet*) for detecting specific expression of GWAS implicated genes ($R=0.18$, $p=0.04$). Future work investigating these discrepancies is warranted.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4095F: Integrating Neuroimaging with Genetics to Predict Risk of Hypertension

Authors:

P. Sutaporn, Y. Veturi, H. Song; Pennsylvania State Univ., University Park, PA

Abstract:

Background: Hypertension, a leading cause of cardiovascular diseases, arises from a complex mix of genetic and environmental factors. Recent advancements in neuroimaging and genetics offer new possibilities for understanding the mechanisms underlying hypertension and creating models to predict its risk. Objectives: This study aims to develop statistical models that improve prediction of hypertension onset time by integrating neuroimaging and genetic data. Methods: We conducted regression analyses on time-to-event data to predict onset of hypertension from the first imaging visit. Our study population consisted of 16,734 participants (53% females, 96.99% white European, and 3.01% other ethnic background) in the UK Biobank who had no prior diagnosis of hypertension before their first imaging visit. We calculated the time-to-event as the duration from the first imaging visit to the date of hypertension diagnosis for individuals who developed the disease during the observation period, or up to the censoring point for those who remained free of hypertension until the end of the study period. We considered 849 T1 and T2 structural and diffusion brain MRI Image-Derived Phenotypes (IDPs) and 336 SNPs corresponding to hypertension-associated loci from the GWAS catalog as our predictor variables. We included sex, age, body mass index (BMI), and 40 ancestry-derived principal components as covariates. We fit Cox proportional hazards regression and random forest models using as predictors (i) only base covariate variables, (ii) base covariates plus IDPs, and (iii) base covariates plus IDPs and SNPs. We compared their out-of-sample prediction accuracy

based on the Concordance index (CI). Results: We observed a general improvement in out-of-sample prediction performance from both models. Using only base covariates, the Cox proportional hazards regression and random forest models resulted in a CI of 0.5525 and 0.5609, respectively. When IDPs were added to the base covariates, the performance improved significantly, with a CI of 0.6205 and 0.6086, respectively. Finally, using IDPs and SNPs as predictors yielded a CI of 0.6131 and 0.6075, respectively from both models. Conclusions: Both Cox proportional hazards regression and random forest models underscored the importance of integrating genetic factors and neuroimaging in predicting hypertension risk. We plan to incorporate results from a genome-wide association study for hypertension onset time on 488,377 individuals (independent from our study population) from UK Biobank to further improve the prediction accuracy of our models, enabling more precise hypertension risk predictions in clinical settings.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4096F: Investigating the Phenomic and Genomic Cognitive Architecture in a Multiethnic Asian population

Authors:

M. Lam[1,2], T. Mina[1], D. Low[1], P. Jain[1], N. Sadhu[1], J. Lee[2], M. Loh[1], J. Chambers[1]; [1]Nanyang Technological Univ., Singapore, Singapore, [2]Inst. of Mental Hlth., Singapore, Singapore

Abstract:

Cognition is one of the most widely used measure of brain health. Poor cognitive function accelerates the progression of neurodegenerative disorders and exacerbates cardiometabolic conditions. The genetic underpinnings of cognitive function are also pleiotropic with various aspects of physical health and mortality. There is little research representing diverse Asian populations. To understanding the epidemiological role of cognition and its biological underpinnings, we performed PheWAS on 616 unique research variables collected in the HELIOS Study (Singapore); comprising up to 9,407 Chinese, Indian and Malay adults aged 30-84 years in whom ‘g’ was derived. We used ‘g’ as predictor and adjusted for age, sex, ethnicity, the interaction of age and sex and genetically determined ancestries. We then performed 2-sample Mendelian Randomization and downstream sensitivity analyses using the GWAS summary statistics from the UK Biobank to ascertain causality; ‘g’ was associated with broad range of phenotypes (Bonferroni p&lt8.1x10-5, 146 [23.7%] phenotypes), ranging from greater familiarity with digital devices (b=0.21, p=5.0x10-153), to smaller waist-hip-ratio (b =-0.06, p=1.2x10-17) and lower white blood cell counts (b=-0.08, p=1.4x10-21). Hierarchical clustering

was conducted to reduce redundancies and ensure representativeness across domains. We selected 23 (3.7%) representative phenotypes (range b=-0.23-0.33; p=4x10-13:5.0x10-153) for Inverse Variance Weighted 2-sample MR, demonstrating causal linkage of genetically predicted 'g' with 11 (47.8%) demographic and physiological phenotypes, including less body fat percentage, lower alcohol consumptions, higher PUFA/MUFA consumptions, and increased household income and educations levels, with significant horizontal pleiotropy. Lower visceral fat, overall body fat, FVC, FEV1, household income and educations levels were also causally linked with higher 'g'. From the initial 147 SNPs in the instrument-exposure 'g', we selected 134 SNPs which uniquely predicts 'g' and performed further annotations. The SNPs are linked with genes involved in neural development, neurotransmission, neuroprotection, and chromatin modification and transcriptional regulation. Our study demonstrates broad implications of cognitive function in lifestyle choices and cardiometabolic health indices in multiethnic Asian populations as potential targets for future cognitive maintenance at population levels. The HELIOS Study addresses the diversity gap in the cognitive genetic research and presents an opportunity to improve the cardiometabolic and psychiatric health in this population.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4097F: Investigating the role of neighborhood socio-economic status and germline genetic risk on prostate cancer risk

Authors:

J. Judd, J. P. Spence, J. K. Pritchard, L. Kachuri, J. S. Witte; Stanford Univ., Stanford, CA

Abstract:

Genetic factors play an important role in prostate cancer (PCa) with polygenic risk scores (PRS) predicting the germline component of disease risk across genetic ancestries. While genetics are one of the strongest risk factors for PCa, there are few convincing modifiable factors, and little is known about their potential interaction with genetic risk. In this study, we used data from the United Kingdom Biobank (UKB) cohort to evaluate the role of neighborhood socioeconomic status (nSES)-and how it may interact with PRS-on PCa risk. For nSES, we used indices of deprivation, a set of validated metrics across seven domains that quantify the relative lack of resources within defined geographical areas across the UK. We evaluated a multi-ancestry PCa PRS containing 451 genetic variants that is strongly associated with prostate cancer in the UKB (OR = 2.28; 95% CI = 2.23-2.35) per standard deviation (SD) increase in the PRS. Analyzing incident PCa cases (n=6,155) and controls (n=98,257) across European, African, South Asian, and Mixed genetic ancestries, we found that several nSES indices of deprivation were inversely

associated with PCa: employment (OR = 0.91; 95% CI = 0.86-0.96), education (OR = 0.94; 95% CI = 0.83-0.98), health (OR = 0.91; 95% CI = 0.86-0.96), and income (OR = 0.91; 95% CI = 0.86-0.96) (all ORs for a SD unit increase in European ancestry models). Additionally, we analyzed the relationship between nSES deprivation and mediators that may explain the association with PCa, such as PSA screening (OR = 0.91; 95% CI = 0.89-0.92) and number of clinical visits (Estimate = 0.75; 95% CI = 0.68-0.83) (all ORs for SD unit increase in the Townsend Index in European ancestry models). These results agree with some previous work (Wu et al. 2017) and may reflect factors that reduce PCa screening and thus influence PCa detection. Additionally, the effect of PRS on PCa did not appear to be modified across nSES indices of deprivation. When testing heterogeneity using Cochran's Q-test, PRS ORs little changed across different quartiles of the indices of deprivation (all p-values > 0.05). These findings suggest that socioeconomic and genetic risk factors for PCa may act independently.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4098F: Is the association between eye disease and cognitive function modified by a *PDE7A/MTFR1* genetic variant?An analysis of hospital data with replication in the Canadian Longitudinal Study on Aging

Authors:

E. Tran[1], M. Rakesh[1], G. Li[2], E. Freeman[1,3], M-h. Roy-gagnon[1]; [1]Univ. of Ottawa, Ottawa, ON, Canada, [2]Maisonneuve-Rosemont Hosp., Montreal, QC, Canada, [3]Ottawa Hosp. Res. Inst., Ottawa, ON, Canada

Abstract:

Background: Age-related eye diseases, such as glaucoma and age-related macular degeneration (AMD), are inconsistently associated with cognitive decline, which could be due to effect modification. In this study, we investigated whether rs10808746, a *PDE7A/MTFR1* genetic variant, previously found to be associated with cognitive decline, modified the association between eye disease and cognitive function.
Methods: Data from a Montreal hospital-based cross-sectional study (n=302) was used for our primary analysis. Interactions between eye disease (glaucoma or AMD) and rs10808746 genotypes were tested in linear regression models for each of six oral cognitive tests. We performed a replication analysis using data from the Canadian Longitudinal Study on Aging (CLSA) (n=22,622). We used linear mixed models to test for interactions between eye disease

(glaucoma or AMD) and genetic variants in a genomic region around rs10808746. All models were adjusted for relevant covariates.
Results: In the Montreal data, we observed two statistically significant (p-value<0.05) interactions between rs10808746 and glaucoma on the letter verbal fluency test and on the immediate story recall (ISR). For both cognitive tests, among individuals with the rs10808746 AA genotype, those with glaucoma performed worse than those with normal vision, while among individuals with the GG genotype, those with glaucoma performed better than those with normal vision. We found similar interactions in the CLSA, although not with the same cognitive measure.
Conclusion: Our results suggest that the *PDE7A/MTFR1* gene variants may modify the association between eye disease and cognitive function. These findings may help to better understand the mechanism by which glaucoma and AMD are related to cognitive function.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4099F: LAWS-HiC: A locally adaptive weighting and screening (LAWS) approach to improve detection of long-range chromatin interactions from Hi-C data

Authors:

L. Zhou[1], W. Chen[1], Y. Li[1], M. Hu[2]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Lerner Res. Inst., Cleveland Clinic Fndn., Cleveland, OH

Abstract:

Hi-C technologies have been widely used to measure genome-wide chromosome conformation, and shed important insights of chromatin spatial organization. Among various Hi-C analyses, detection of long-range chromatin interactions (or 3D peak calling) remains a most commonly performed crucial task. However, widely used tools for 3D peak calling usually assume statistical independence among all pairwise loci interactions, which is violated when dealing with high-resolution (e.g. <= 10 kb) Hi-C data, where spatial dependency among neighbor loci is non-negligible, leading to invalid analysis or power loss. In this study, we present LAWS-HiC, a 3D peak caller adopting the Locally Adaptive Weighting and Screening (LAWS) approach, designed to enhance detectability by incorporating local patterns into the post-processing of results from any standard 3D peak caller. We benchmarked LAWS-HiC using two high-depth Hi-C datasets, a GM12878 dataset from a human lymphoblastoid cell line and a mouse embryonic stem cell (mESC) dataset, with 4.9 and 7.2 billion raw reads, respectively. Our results indicate that LAWS-HiC improves overall recall while maintaining nearly identical precision compared to state-of-

the-art alternatives, including FitHiC2. For example, on the 20% (10%) downsampled mESC dataset, HiC-LAWS improves the recall from FitHiC2's 0.369 to 0.453 (0.107 to 0.136), while both methods achieve perfect precision. On the 20% (10%) downsampled GM12878 dataset, LAWS-HiC improves the recall from 0.251 to 0.432 (0.123 to 0.271), while remaining a precision > 0.99.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4100F: Leveraging genetic similarity to investigate understudied genetic variation associated with EHR-derived phenotypes in diverse patient biobanks

Authors:

D. Zhang, S. Damrauer, M. Ritchie, D. Rader; Perelman Sch. of Med., Philadelphia, PA

Abstract:

The lack of representative population diversity in existing genetic association studies not only hinders our understanding of disease pathobiology, but also contributes and exacerbates existing health disparities. We herein propose a tailored genome-first approach that leverages existing patient diversity in large-scale biobanks to investigate commonly understudied genetic variations. The Penn Medicine Biobank includes over 11,000 individuals genetically similar to African reference populations (AFR) with whole-exome sequencing. For each sequenced variant, we extracted its gnomAD allele counts in the African (AFR) and non-Finnish European (EUR) population groups and performed a chi-sq test with stringent p-value correction to identify variants significantly more common in the AFR population. We then filtered for EUR minor allele frequency (MAF) 5%, AFR MAF > 0.05%, and predicted loss-of-function (pLOF) and missense variants. An AFR/EUR MAF ratio 200 cutoff was used to identify ~70,000 final variants that are more common in AFR relative to EUR and likely to have been understudied. Using linked clinical data in the biobank, we performed phenome-wide association studies for each variant against diagnosis codes, laboratory values, and CT-derived imaging traits. We identified well-known associations such as *HBB* p.Glu7Val with sickle cell anemia (odds ratio [OR]=103.54, p=9.80E-181), *APOL1* p.Ser342Gly with chronic kidney disease (OR=2.04, p=2.41E-23), and *G6PD* p.Val68Met with hereditary hemolytic anemia (OR=4.79, p=7.58E-19) that served as strong positive controls for our approach. Additionally, we found novel associations including *APOL3* p.Gln58* with end-stage renal disease (OR=1.39, p=7.36E-09), *SPATA5L1* p.Arg119Pro with hypocalcemia (OR=0.56, p=4.40E-08), *RAET1G* p.Thr84Arg with gastric ulcers (OR=3.90, p=5.61E-08), and *UNC45B* p.Ala881Thr with hearing loss (OR=3.06, p=8.03E-08). Multiple associations were significantly replicated in AFR populations in the Million

Veteran Program (MVP) and the *All of Us* (AoU) Research Program. The variant *SPATA5L1* p.Arg119Pro was associated with chronic renal failure in MVP (OR=1.10, p=5.63E-17), consistent with its observed negative association with hypocalcemia. The variant *UNC45B* p.Ala881Thr was found to be associated with hearing loss in AoU (OR=1.57, p=0.038). Our approach represents a unique methodological strategy for targeting underexplored variants and uncovering novel biology in understanding genetic risk for disease. Future work will involve applying this approach in other non-AFR populations and leveraging large-scale genomic studies to address health disparities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4101F: Leveraging single cell transcriptomics to identify pathogenic cell lineages

Authors:

Z. Du, M. Tsabar, S. Lessard, T. Iyyanki, D. Liu, H. Jeremy, V. Savova, F. Rapaport, E. de Rinaldis, S. Khader, C. Chatelain, G. Gaglia; Sanofi US Services Inc., Cambridge, MA

Abstract:

The abnormal physiology of disease can be conceptualized as driven by a subset of cells either acquiring novel functions or losing canonical traits. While in cancer this is a well-established model for disease pathology, its application to autoimmune and neurological diseases is still under debate. In these diseases, a crucial missing step is identifying the lineages that drive disease, or “pathogenic” cell types. Here, we started from the basic assumption that genetic disease modifiers would have an increased impact on cell types that actively expressed the gene, and in turn that a pathogenic cell would be more likely to express genes that have a causal link to disease. In key autoimmune diseases, we combined single cell transcriptomics and human genetics to rank cell types by their likelihood of pathogenicity. We leveraged scRNA-seq data to identify cell type specific genes in each disease and intersected them with the disease’s casual gene modifiers. We found that the gene expression of genetic modifiers broadly clustered cells by their lineage, with several modifiers exclusively expressed in a single lineage. Moreover, we were able to rank cell types by their pathogenicity in atopic dermatitis, lupus nephritis, type 1 diabetes and inflammatory bowel syndrome, and validated the scoring by comparing it to an extensive literature review effort. The data-driven pathogenicity score was overall able to predict the pathogenic cell types cited in the literature and find new cell types potentially driving disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4102F: Leveraging transfer learning to bridge data gaps: Identifying unique sepsis risk factors in chronic disease patients

Authors:

R. Caldino Bohn[1], J. 2. Carcillo[2], K. Kernan[2], H. Park[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Children's Hosp. of Pittsburgh, Pittsburgh, PA

Abstract:

Sepsis, a life-threatening dysregulated host response to infection, poses a significant global health challenge with high mortality rates and a substantial healthcare burden. Individuals suffering from chronic diseases, such as Chronic Obstructive Pulmonary Disease (COPD) or Type 2 Diabetes (T2D), are at elevated risk of developing sepsis. However, the underlying mechanisms and risk factors driving this increased susceptibility remain poorly understood, hindering the development of targeted prevention and management strategies. The advent of population-scale health databases, such as the UK Biobank (UKBB), has revolutionized our ability to uncover the underlying mechanisms and sepsis risk factors. However, existing risk models often lack statistical power when attempting to identify sepsis risk factors specific to smaller cohorts characterized by particular chronic diseases. To address this limitation, we employed a novel transfer learning framework that first learns a general sepsis risk model from all available sepsis cases, capturing broad risk patterns. Subsequently, this model is fine-tuned using sepsis cases in individuals with COPD or T2D, enabling the identification of unique genetic and exposome risk factors associated with these chronic conditions. Through comprehensive evaluations of UKBB data, we demonstrate the ability of our transfer learning framework to uncover novel risk factors and biomarker signatures specific to COPD and T2D patients with sepsis. Our interpretability analyses provide insights into the underlying biological mechanisms and pathways associated with increased sepsis susceptibility in these chronic disease populations. By bridging the gap between the general sepsis cohort and chronic disease-specific sepsis cohorts, our transfer learning approach represents a significant advancement in sepsis risk stratification and personalized management strategies for individuals with chronic comorbidities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4103F: MACHINE: a robust multi-ancestry fine-mapping method using a continuous global-local shrinkage prior with applications in psychiatric diseases

Authors:

X. Li, Z. Xiong, P. Sham, Y. Zhang; The Univ. of Hong Kong, Hong Kong, China

Abstract:

The purpose of fine mapping is to identify genetic variants that have nonzero effects on the phenotypes. Leveraging data from diverse ancestries can improve the accuracy and resolution of fine mapping due to variations in linkage disequilibrium patterns and increased sample sizes. Existing multi-ancestry fine-mapping methods are all based on discrete priors and most of them assume that all causal variants are shared across ancestries, which may not hold true. A recently developed method, MESuSiE, accounts for both shared and ancestry-specific causal variants across ancestries but requires prespecified prior probabilities for causal variants being shared or specific to an ancestry. Moreover, fine-mapping methods based on discrete mixture priors are prone to sub-optimal solutions and result in false discoveries.To address these limitations, we introduce Multi-AnCestry Heritability INducEd Dirichlet decomposition (MACHINE), a Bayesian fine-mapping method that employs a continuous prior. Simulation studies demonstrate that MACHINE performs robustly across various scenarios and has a better control of false discovery rate compared to existing multi-ancestry fine-mapping methods. In the real data analysis, we applied MACHINE to several psychiatric diseases, identifying previously unknown causal variants and depicting their genetic architectures across ancestries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4104F: Measuring causal effect heterogeneity across the human phenome by linking biobank-scale genomic data and electronic health records

Authors:

J. Wang, S. Zollner, J. Morrison; Univ. of Michigan, Ann Arbor, MI

Abstract:

Revealing causal relationships between modifiable risk factors and diseases is one of the central problems in epidemiological, human genetic, and public health research. Several experimental

designs as well as statistical methods have been proposed to estimate these causal effects. Mendelian randomization (MR) is a causal inference technique that can be used to estimate causal effects in the presence of unmeasured confounding by using genetic variants as instrumental variables. Standard MR methods estimate the average treatment effect, representing the average effect of intervening on the exposure across all individuals in the population. These methods assume either homogeneity of the causal effect or monotonicity in the association between the genetic instruments and the exposure. However, heterogeneity of causal effects may lead to a lack of generalizability of effects measured by MR in non-representative populations. In this study, we aim to assess the extent and magnitude of effect heterogeneity across age and sex for a range of human diseases in the Michigan Genomics Initiative (MGI) and UK Biobank. We use structural-equation-based two-stage regression models to estimate main causal effects and putative effect modification, using externally estimated polygenic risk scores as instruments in order to explore the causal relationships between > 30 quantitative blood and urine biomarkers and > 3,000 binary traits. Our results identify a small number of age-related effect modification signals, including the associations between HDL cholesterol, Apolipoprotein A and type 2 diabetes. We did not identify pervasive sex-related effect modification. Additionally, we use simulations to assess the impact of violation of instrumental variable assumptions (weak instrument, instrument heterogeneity, horizontal pleiotropy), and regression modeling strategies. These results provide a survey of causal effect modification by age and sex over a wide range of human phenomes and provide context for interpreting marginal effects from Mendelian randomization.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4105F: Meta-analysis approach for copy number variation aggregation association tests in multi-ethnic studies

Authors:

Y. Cheng[1], S. Holloway[2], W. Lu[1], Y. Si[1], A. Tucci[3], H. Wang[4], G. Schellenberg[4], L-S. Wang[4], W-P. Lee[4], J-Y. Tzeng[1]; [1]North Carolina State Univ., Raleigh, NC, [2]Duke Univ., Durham, NC, [3]North Carolina State Univ., Wendell, NC, [4]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

With the availability of large-scale sequencing projects such as the Alzheimer's Disease Sequencing Project (ADSP), association studies have expanded to include variants beyond SNPs, such as CNVs, and multi-ethnic samples. CNVs significantly contribute to genetic diversity and impact phenotypes and diseases by altering gene dosage and structure. Including CNVs in

genetic studies provides a more comprehensive understanding of complex traits, and analyzing CNV in multi-ethnic samples ensures the robustness and generalizability of findings. For multi-ethnic association studies, aggregation tests using summary statistics (SS) are particularly useful as they enhance statistical power and interpretability by combining information from multiple genetic variants. SS-based aggregation tests help detect associations across diverse populations, where variant frequencies may differ between ethnic groups. Using SS also protects privacy, improves computational efficiency, enables data sharing, all of which facilitate meta-analysis across studies with different ethnic backgrounds.While many SS-based aggregation tests are available for SNP, few exist for CNVs. Unlike SNPs, CNVs lack a natural definition for CNV loci, and CNV loci in a study are defined based on unique breakpoint patterns observed in that particular study. In this work, we propose a meta-analysis approach for CNV aggregation tests, which is built on the kernel-based aggregation test CONCUR and is referred to as "MetaCONCUR". We base our work on CONCUR because (i) CONCUR considers the CNV events of an individual using a CNV profile curve, bypassing the need to define a CNV locus; and (ii) CONCUR models the effects of CNV length and dosage. Due to the study-specific nature of CNV loci and the lack of standard reporting practices for CNV association results, we adopt a framework similar to MetaSTAAR, and design MetaCONCUR with two main steps: the worker step and the meta step. The worker step requires investigators of each study to compute score statistics at study-specific CNV loci using individual-level data and output relevant summary statistics. The meta step then aligns CNV summary statistics from different studies and combines them to compute meta-association test statistics and p-values. We ensure the feasibility and scalability of MetaCONCUR in numerical computations and storage of summary statistics. We demonstrate the validity and utilities of the proposed framework through simulation studies with various levels of population heterogeneities and by conducting ADSP multi-ethnic association analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4107F: Modernizing crlmm for Robust Genotyping and Efficient Genomic Analysis

Authors:

B. Carvalho[1], R. Irizarry[2], I. Lopes-Cendes[3]; [1]Univ. of Campinas, Campinas, SP, Brazil, [2]Harvard Univ., Cambridge, MA, [3]Univ. of Campinas, Campinas, Brazil

Abstract:

The *crlmm* package, a well-established package within the Bioconductor ecosystem, has been widely utilized for genotyping analysis of SNP array data. However, we recognize the need to enhance the package's utility as genomic datasets grow in scale and complexity. This work presents a series of focused improvements to the *crlmm* package, specifically designed to improve genotype calling accuracy, expand quality control capabilities, and streamline integration with downstream analyses. These enhancements benefit research directly, making work more efficient and accurate.Specifically, we have expanded the reference dataset used for prior knowledge determination, improving the accuracy of rare SNP calls. We have introduced additional QC metrics at the SNP level, including signal intensity, heterozygosity, and call rate, allowing for a more comprehensive data quality assessment. To facilitate downstream analyses, we have added functionality to export results in the widely adopted VCF format and implemented tools for genome-wide association studies (GWAS). Additionally, we have optimized the package to handle larger datasets by improving memory usage and code refactoring computational bottlenecks. Finally, we have incorporated visualization tools to facilitate the exploration and interpretation of genotyping results.These enhancements were not just implemented in isolation, but with the collaborative spirit of the research community in mind. We aimed to make the *crlmm* package an even more powerful and versatile tool for analyzing genotyping data, with researcher needs and feedback in mind. By improving this tool, we are benefiting our research and contributing to the collective advancement of genomics and personalized medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4108F: MR.RGM: An R Package for Fitting Bayesian Multivariate Bidirectional Mendelian Randomization Networks

Authors:

B. Sarkar, Y. Ni; Texas A&M Univ., College Station, TX

Abstract:

Mendelian randomization (MR) infers causal relationships between exposures and outcomes using genetic variants as instrumental variables. Typically, MR considers only a pair of exposure and outcome at a time, limiting its capability of capturing the entire causal network. We overcome this limitation by developing **'MR.RGM'** (Mendelian randomization via reciprocal graphical model), a fast R-package that implements the Bayesian reciprocal graphical model and enables practitioners to construct holistic causal networks with possibly cyclic/reciprocal causation/measured confounders and proper uncertainty quantifications, offering a

comprehensive understanding of complex biological systems and their interconnections. **'MR.RGM'** is an open-source R package that applies bidirectional MR using a network-based strategy, enabling the exploration of causal relationships among multiple variables in complex biological systems. **'MR.RGM'** holds the promise of unveiling intricate interactions and advancing our understanding of genetic networks, disease risks, and phenotypic complexities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4109F: Multi-level gene-environment interaction analysis of fish oil supplementation and 14 circulating polyunsaturated fatty acid-related traits

Authors:

S. Ihejirika, A. H. Chiang, A. Singh, K. Ye; Univ. of Georgia, Athens, GA

Abstract:

Genetic and environmental factors, such as diet, influence circulating polyunsaturated fatty acid (PUFA) levels. Fish oil supplements (FOS) are a rich source of omega-3 PUFAs, an essential component of human diet. FOS are known to alter circulating PUFA levels among individuals heterogeneously. These varied responses may result from unidentified gene-FOS interactions. To identify genetic factors that interact with omega-3 PUFA supplementation to alter the circulating levels of PUFAs, we performed a large multi-level genome-wide interaction study (GWIS) of fish oil supplementation on 14 plasma measurements related to PUFAs and monounsaturated fatty acids (MUFAs) in 200,060 unrelated European-ancestry individuals from the UK Biobank. We conducted single-variant interaction analysis and a gene-level aggregate test. We then performed a transcriptome-wide interaction study (TWIS) to identify genes whose expression interacts with fish oil. Lastly, we estimated the phenotypic variance explained by genome-wide interactions with fish oil supplementation. From our single-variant tests, we identified genome-wide significant interacting SNPs (P < 5e-8) in the FADS1-FADS2 gene cluster on total omega-3 concentration and its proportion in total fatty acids (expressed as a percentage), DHA concentration and its percentage, and the ratio of omega-6 to omega-3. We performed an exposure-stratified genetic association analysis to quantify the effect of the C major allele of the lead SNP, rs35473591 (chr 11:61586328; CT>C, MAF = 0.34), on omega-3%. It had a lower effect size in the fish oil-taking group (beta = 0.35) than in the group that does not use FOS (beta = 0.42). Furthermore, fish oil supplementation significantly altered omega-3% levels with different effect sizes across the three genotype

groups (beta: = 0.45 in CT/CT, 0.50 in C/CT, 0.59 SD units in CC). We also identified significant signals at the FADS1-FADS2 locus in our gene-level aggregate and TWIS tests. The contribution of gene-FOS interactions to phenotypic variance was not statistically significant across all 14 PUFA and MUFA-related traits. Our study is the most extensive gene-FOS GWIS study on the circulating levels of PUFAs to date. Our study replicated the FADS1-FADS2 locus, which is relevant in the biosynthesis of omega-3 and omega-6 PUFAs. This study contributes to our understanding of the genetic architecture underlying FOS response and can help to improve personalized diet recommendations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4110F: On the use of Random Forest for obtaining Genetic Risk Score

Authors:

V. Venkat[1], K. Clark[2], Y. Si[3], H. Wang[4], T-P. Lu[5], T-H. HSIAO[6], C-H. Lin[7], H-J. Tsai[8], T-C. Yao[9], G. Schellenberger[10], L-S. Wang[4], W-P. Lee[4], J-Y. Tzeng[3]; [1]North Carolina State Univ., raleigh, NC, [2]Univ. of Pennsylvania, Baltimore, MD, [3]North Carolina State Univ., Raleigh, NC, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Natl. Taiwan Univ., Taipei, Taiwan, [6]Taichung Veterans Gen. Hosp., Taichung City, Taiwan, [7]Taichung Veterans Gen. Hosp., Taichung, Taiwan, [8]Natl. Hlth.Res. Inst.s, Zhunan Town, Taiwan, [9]Chang Gung Univ. Coll. of Med., Taoyuan, Taiwan, [10]Univ. of Pennsylvania, Pennsylvania, PA

Abstract:

Genetic risk score (GRS) is an essential tool for estimating an individual's genetic liability to various traits, aiding in disease risk assessment, and elucidating the genetic architecture of complex traits. Traditional GRS computation methods primarily assume additive linear effects of risk variants. However, emerging evidence suggests that non-additive and epistatic effects significantly influence complex traits. This study investigates the application of random forest (RF), a model-free ensemble learning method, for constructing GRS to capture these complex genetic interactions. RF's ability to model non-linear and interactive effects without assuming a predefined genetic model offers significant advantages over conventional methods. Despite its demonstrated superiority in various genomic studies, RF has not been fully leveraged for GRS construction, especially in utilizing extensive summary statistics from GWAS base data. We propose novel strategies for RF-based GRS construction, aiming to enhance prediction accuracy and robustness by incorporating C+T and GWAS summary statistics. Our primary strategy integrates C+T into RF, optimizing hyperparameters to maximize predictive accuracy.We evaluate

these strategies through simulation studies using real genotype data from the Taiwan Biobank and real data applications with various levels of similarities of ancestral backgrounds between base data and target data, including data from ADSP, LIGHTS, Taiwan Biobank, and CoLaus/PsyCoLaus studies. Our results demonstrate that incorporating C+T into RF significantly boosts predictive performance and robustness, particularly when etiological homogeneity between base and target data varies. In simulations and real data applications, the RF-based GRS outperformed traditional methods. These findings underscore the potential of RF to construct more accurate and robust GRS by capturing complex genetic architectures and utilizing comprehensive GWAS summary statistics. This work paves the way for improved genetic risk prediction, ultimately contributing to personalized medicine and public health initiatives.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4111F: Optimizing multitrait approach in Brain imaging GWAS ⭐

Authors:

A. Auvergne[1], L. Henches[1], H. Julienne[1], J. Lefèvre-daumal[2], H. Aschard[1]; [1]Inst. Pasteur, Paris, France, [2]INSERM, Paris, France

Abstract:

There is raising interest for multitrait genome-wide association studies (GWAS) of magnetic resonance imaging (MRI)-derived brain phenotypes. MRI traits have high dimension, are highly correlated, and are expected to harbor substantial shared genetic. Multitrait analysis can help reduce the multiple testing burden, and improve statistical power to detect associated genetic variants. One open challenge for such applications is the definition of sets of variants to be analyzed jointly out of millions of possibilities. Intuitively, multivariate analyses are applied on sets of related phenotypes, belonging to the same cortical structure, or the same functional specialization (e.g. Eliott et al, 2018; Wu, 2020 ; van Der Meer, 2020). However, there is increasing evidence that multitrait analyses might lead to larger statistical power when analyzing sets of fairly heterogeneous phenotypes. Here, we investigated the impact of subset-definition strategies on the variant discoverability for multitrait GWAS of 1,010 neuroanatomical phenotypes. We derived the GWAS summary statistics of these phenotypes using the UK Biobank data (32,947 individuals and 9,583,000 variants divided into 1,703 independent regions). We applied an omnibus multitrait test to four biologically-informed homogeneous subsets of phenotypes, mean thickness (Nphe=30), surface area (Nphe=10), mean intensity (Nphe=40), and volume of various brain regions (Nphe=40), and series of heterogeneous

subsets of phenotypes including 10 to 40 phenotypes. The number of regions harboring significant association from the homogeneous subsets varies in [103, 226]. In comparison, the number of significant regions from the heterogeneous sets equals on average 248, corresponding to a 60% increase in power. Altogether, our work confirms that using sets of heterogenous brain MRI phenotypes can lead to a substantial boost in discoverability as compared to the common practice of using related MRI phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4112F: Performance of two-phase designs for the time-to-event outcome of genomic sequencing studies and a case study assessing B-ALL primary mutation associated relapse risk

Authors:

W. Chen[1], T-C. Chang[2], C. Mullighan[2], M. Loh[3], G. Wu[2]; [1]Mayo Clinic, Rochester, MN, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Seattle Children's Hosp., Seattle, WA

Abstract:

To reduce costs in genomic association studies with time-to-event phenotypes (e.g., survival analysis), sequencing a subset of samples in the second phase from the large full cohort using available variables collected for the full cohort in the first phase is becoming increasingly common. Recently an optimal two-phase design was proposed. The corresponding maximum likelihood method has the advantage of dealing with arbitrary sample selection in the second phase. However, there is a lack of direct comparison with other two-phase designs in terms of power and computational cost, which are important factors for planning genome-wide analysis of sequencing data. In this study, we evaluated the performance of important two-phase designs including the optimal designs and the coupled analysis methods, in terms of type I error, power, and effect size estimation. We showed that the optimal design with the corresponding maximum likelihood based method had the highest power and consistent effect size estimation under the Cox regression model. Surprisingly, the simple logistic regression achieved similar or only slightly lower power than the maximum likelihood method for both the optimal design and another extreme case control design, but with much less computational cost. One drawback of logistic regression is that it cannot provide a consistent estimation of the hazard ratio. We then applied both logistic regression and the maximum likelihood method to an extreme case control design employed in the Molecular Profiling to Predict Response to Therapy (MP2PRT) study, where the association between cancer subtypes and cancer relapse after treatment in pediatric B-cell acute lymphoblastic leukemia patients was investigated. We reported the estimated

hazard ratio of each main cancer subtype using the maximum likelihood method, which will be useful for comparison with other full cohort based studies. Logistic regression takes less than 1 second for the analysis of the MP2PRT data, but the maximum likelihood method can take several hours depending on the number of second-phase variables. Finally, we provided our recommendations on the selection of two-phase designs and analysis methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4113F: Pharmacogenomic genome-wide association study of LDL response to statins using high-throughput electronic health records analysis

Authors:

S. Goleva[1], J. Keaton[1], B. Waxse[2], A. Williams[1], A. Awan[1], J. Denny[1,3]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [3]All of Us Res. Program, Bethesda, MD

Abstract:

We developed a high-throughput, generalizable self-controlled case series method to determine drug effectiveness for pharmacogenomic analysis using electronic health record (EHR) data from the All of Us Research Program (All of Us). To demonstrate the utility of this method, we applied it to determine HMG-CoA reductase inhibitor (statin) effectiveness, which we quantified using the difference in LDL laboratory values before and after statin prescription.
We examined whether atorvastatin, simvastatin, rosuvastatin, pravastatin, lovastatin, or fluvastatin monotherapy effectiveness differences exist between Hispanic, White, and Black populations using adjusted linear regressions. We then identified any pharmacogenetic variants associated with statin effectiveness by conducting a trans-ancestry genome-wide association study (GWAS) adjusted for age at statin prescription, sex at birth, ten principal components, and baseline LDL value prior to statin initiation.
The mean reduction in LDL following statin prescription in All of Us was 28.3 mg/dL (95% CI = 27.1 - 29.4; n = 4,695). Consistent with published studies, we observed that statins are less effective in Black (mean reduction in LDL = 23.25; 95% CI = 20.47 - 26.03; n = 760; p = 0.003) and Hispanic (reduction = 19.58, 95% CI = 16.07 - 23.09; n = 503; p = 0.004) compared to White (reduction = 30.62, 95% CI = 29.31 - 31.9; n = 3,432) populations. Our GWAS (n = 4,035) revealed a significant association, rs7412 (β= 7.19, SE = 1.32, p = 5.9e-09), a variant within APOE previously associated with hypercholesterolemia and atorvastatin response. Another significant association, rs73151468 (Beta = -8.35, SE = 1.52, p = 4.2e-08), resides within a lncRNA gene,

ENSG00000234223. Data from the Gene-Tissue Expression (GTEx) project suggests ENSG00000234223 is only expressed in the liver and previous findings suggest its association with hepatocarcinoma, which may be relevant since lipogenesis occurs in the liver.
We used EHR data to determine statin effectiveness and identified differences in race and ethnicity in statin effectiveness as well as genetic variants associated with statin response. Our findings demonstrated the validity of our EHR-based self-controlled case series method to determine drug effectiveness, and highlight the generalizability to determine drug effectiveness using EHR data in other phenotypes. Replication of the pharmacogenomic association in APOE validates the accuracy of this method, and a novel association demonstrate the possibility to discover clinically and mechanistically relevant pharmacogenomic targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4114F: PhenoGPT2: adapting domain-specific Retrieval Augmented Fine-Tuning and In-context learning to enhance recognition of phenotypic abnormality from clinical notes

Authors:

Q. Nguyen[1,2], J. D. Wu[2], K. Wang[2,1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

Recognizing phenotypic abnormalities in genetic disorders often provides important clues to facilitate diagnoses, by expediting referrals to specialized experts or suggesting genetic tests. These clinical phenotypes are usually documented in unstructured clinical notes as free texts. Thus, developing efficient informatics methods for extracting and interpreting phenotypic data from clinical documentation is crucial for enhancing physicians' ability to recognize and manage genetic disorders promptly. Recently, large language models (LLMs) have demonstrated significant potential in various medical tasks; for example, we developed PhenoGPT for extracting clinical phenotype terms from clinical notes. However, LLMs still face challenges, such as hallucinations when mapping phenotype terms to standardized vocabularies like the Human Phenotype Ontology (HPO). In this study, we introduce a novel computational pipeline called PhenoGPT2 designed to improve LLMs' ability to extract phenotypes and accurately associate them with HPO identifiers. Firstly, we fine-tune the Llama 3 (8B) model using the HPO database to let the model learn generic medical knowledge related to HPO terms. More specifically, we supplied LLMs with all the HPO identifiers, official names, synonyms, and parental phenotypes for over 13,000 phenotypes. Then, we continue to fine-tune the model utilizing the Gold

Standardized Corpora (GSC+) Biolark, which includes short paragraphs (containing multiple phenotypes), alongside contextual information sourced from Retrieval Augmented Generation (RAG), a combination of generative and retrieval-based AI models, for HPO recognition task. Additionally, we specified the response in a chain-of-thought (COT) format with crafted-prompts, guiding the model to accurately extract the phenotypic abnormalities and correctly convert them to HPO identifiers. During inference, RAG was used to enhance LLMs' responses with accurate source attributions. We compared PhenoGPT2 to its predecessors (original PhenoGPT and PhenoGPT+BioSent2Vec). Our preliminary results indicate that PhenoGPT2 correctly extracts 81.5% of phenotypes and converts them to HPO identifiers with an average similarity of 72.2% on the ID-68 dataset. In comparison, PhenoGPT achieved 65.6% for phenotype extraction and 27.3% for HPO conversion, while PhenoGPT+Biosent2Vec achieved 69.5% for phenotype extraction and 71.9% for HPO conversion. This result indicates that the fine-tuning with RAG and COT significantly improves the capabilities of detection and standardization of phenotypes in clinical notes, which can be helpful for various downstream medical tasks.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4115F: Polygenic Modeling with Sparse LD Representation for Enhanced Efficiency and Performance

Authors:

M. Witteveen, T. Ge; Massachusetts Gen. Hosp., Boston, MA

Abstract:

Background: Polygenic Risk Scores (PRS) are increasingly used to predict individual genetic predispositions to various diseases. The conventional PRS-CS method, while effective, is restricted to HapMap3 variants due to challenges in computational efficiency. This restriction may reduce its predictive performance, particularly in non-European populations due to the unequal tagging capabilities of HapMap variants across different ancestries. To address these challenges, we developed SPRS-CS, a novel iteration of the PRS-CS method that combines a highly sparse and efficient representation of LD matrices, derived from Linkage Disequilibrium Graphical Models (LDGMs), with a newly designed posterior sampler to substantially improve the computational and memory efficiency of PRS-CS.

Methods: Our SPRS-CS method incorporates LDGMs to infer posterior SNP effect sizes under continuous shrinkage (CS) priors. The model is implemented within a Markov Chain Monte Carlo (MCMC) framework with various enhancements, including a specialized sampler algorithm

optimized for LDGM blocks. We validated our model using datasets from five ancestry groups and compared its performance to traditional methods.
Results: Initial evaluations of SPRS-CS demonstrated significant improvements in runtime and computational efficiency. When applied to the HapMap3 dataset (1.1 million variants), SPRS-CS outperformed the original PRS-CS method in speed while maintaining prediction accuracy. When scaled to 6.2 million variants, which was not feasible with PRS-CS, the performance gains were more pronounced, highlighting the model's scalability and efficiency. Preliminary results indicated that SPRS-CS also enhanced the prediction accuracy in individuals of diverse ancestries.
Conclusion: SPRS-CS represents a substantial advancement in polygenic risk scoring methodologies by leveraging the sparse representation of LD and highly efficient posterior sampling algorithms. This approach improves the coverage of association signals across the genome without relying on the selection of tagging variants, enabling more accurate and efficient genetic predictions in both clinical and research settings.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4116F: Population-based biobank analysis reveals elevated frequencies of G6PD deficiency Variants in Quebec.

Authors:

I. Pender[1,2,3], J. Pelletier[1,3], D. Ashton-Beaucage[1,2,3], D. Taliun[1,2,3], C. Bhérer[1,2,3]; [1]McGill Univ., Montreal, QC, Canada, [2]Dept. of Human Genetics, Faculty of Med. and Hlth.Sci., McGill Univ., Montreal, QC, Canada, [3]Canada Excellence Res. Chair Program in Genomic Med., VP Dahdaleh Inst. of Genomic Med., McGill Univ., Montreal, QC, Canada

Abstract:

The role of X chromosome variation as a determinant of human health is broadly understudied, despite its role in many diseases. Sex chromosomes are frequently excluded from genomic analysis and underrepresented in research despite being known to play major roles in many genetic diseases. We investigated the prevalence of X-linked Mendelian diseases in the CARTaGENE (CaG) population-based cohort of Quebec residents. We used high-depth whole genome sequences from CaG, including participants of French-Canadian (FC) (n=1,873), Haitian (HI) (n=238), and Moroccan (MO) (n=166) genetic ancestries. We used Ensembl Variant Effect Predictor to annotate known ClinVar pathogenic variants. This scan revealed multiple G6PD deficiency variants at elevated frequencies. G6PD deficiency is a highly polymorphic condition with over a hundred previously described pathogenic variants. In Quebec, G6PD deficiency is

typically imagined as a condition that primarily affects individuals from the Middle East and Africa. There are no published Canada-wide or Quebec-wide prevalence estimates, and it is not part of neonatal screening program in Quebec. In 2,157 individuals sequenced at high depth, we discovered the presence of five G6PD variants with some prior evidence of pathogenicity. c.934G&gtC, c.770G&gtA, c.466A&gtG, c.961G&gtA, and c.292G&gtA (GRCh38). Two of these variants are present in the CaG FC sample at statistically elevated rates compared to non-Finnish European from gnomAD, c.770G&gtA, (Fisher exact test odds ratio (OR) = 11.1, p-value = 0.00082), and c.466A&gtG, (OR= 5.9, p-value =3.11E-05). In the CaG-FC sample we estimated a cumulative carrier frequency of 1/292 for G6PD variants classified as Class III by the World Health Organization; i.e. 10-60% enzymatic activity, in addition to carrier frequencies of 1/13 in CaG HI and 1/37 in CaG MO. These phenotypes are characterized by a degraded form of the G6PD enzyme, leading to increased hemolysis (breakdown of red blood cells). G6PD deficiency can lead to serious reactions to certain foods and is a contraindication for several common antibiotics. This work highlights the value of population-based biobanks for analyzing rare disease-causingvariants, including on the traditionally understudied X chromosome. These results will inform policy regarding G6PD deficiency population screening.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4117F: Population-enriched germline variants in Finns pinpoint shared mechanisms shaping hematopoietic mosaic chromosomal alterations and diverse solid tumors ⭐

Authors:

A. Liu[1,2,3,4,5], G. Genovese[2,3,6], M. Kanai[1,2,7], N. B. Rockweiler[2,3,6], A. Sekar[8], Z. Zheng[1,2,3,4], T. P. Sipilä[5], A. Ghazal[5], M. Aavikko[5], FinnGen, A. Ganna[1,2,3,5], A. Palotie[1,2,3,5], S. McCarroll[2,3,6], M. Daly[1,2,3,4,5]; [1]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [2]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [5]Inst. for Molecular Med. Finland (FIMM), HiLIFE, Univ. of Helsinki, Helsinki, Finland, [6]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [7]Program in Bioinformatics and Integrative Genomics, Harvard Med. Sch., Boston, MA, [8]Dept. of Med. Oncology, Dana-Farber Cancer Inst., Boston, MA

Abstract:

With age, humans can acquire genetic changes in somatic cells; such changes are hallmarks of biological age and are linked to aging-associated diseases including cancers. In biobank-scale

data from blood-derived DNA, diverse types of somatic mutations can be identified, providing a unique opportunity to study clonal hematopoiesis in detail and giving insights into the origins of other somatic mutations. Thus Far, research in the field has centered on commonly recurring hematopoietic clones such as clonal hematopoiesis of indeterminate potential and sex chromosome loss, with autosomal mosaic chromosomal alterations (mCAs) less addressed. Autosomal mCAs are individually less common and comprise a heterogeneous group of events. Here, we screened Affymetrix microarray data of 444K FinnGen participants (43% males, median age 65) for autosomal mCAs and identified 32765 events in 19978 individuals. We performed GWAS by pooling all autosomal mCAs together and identified several germline variants with broad effects in acquiring any clones; many of these variants are also associated with myeloproliferative diseases in Finns (e.g., JAK2, TERT, CHEK2, IFT80). Additionally, the wide age range of FinnGen samples allowed us, for the first time, to identify genome-wide significant trans signals for 30 mCA subtypes. For example, we found that a missense variant in MADL1 (rs1801368), previously reported for Y chromosome loss, was also a germline risk factor for somatically-acquired trisomy 15. Knowing that autosomal mCAs, as a group, are associated with both myeloid and lymphoid leukemias, we then divided autosomal mCAs into those that specifically increase the risk of myeloid versus lymphoid malignancies. We identified HLA alleles and a Finnish-enriched missense variant in RPA1 (rs202068855, AF=1% in Finns, 50-fold enriched than in non-Finnish Europeans) as novel signals for lymphoid mCAs, and a Finnish-enriched frameshift in CHEK2 (rs555607708, AF=1% in Finns, 4-fold enriched) as shared for both myeloid and lymphoid mCAs. Interestingly, several of these Finnish-enriched variants are also cancer driver mutations. We therefore hypothesized that some autosomal mCA drivers might have broad roles in shaping diverse types of somatic mutations predisposing to solid tumor development, and found supportive evidence also from genes ATM, TP53, and SPDL1, all involved in cell-cycle regulation. Importantly, these results support a partially shared mechanism for the origins of hematopoietic mCAs and diverse solid tumors. Given that blood DNA is relatively easy to obtain, these findings suggest a potential utility of hematopoietic mCAs as a biomarker of solid tumor risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4118F: Powerful estimation of the genetic component of gene expression with a Poisson mixed effect model of single cell RNA-sequencing reads

Authors:

Z. Xu[1], W. Zhou[2], A. Massarat[3], L. Rumker[4], S. Raychaudhuri[5], M. Gymrek[6], T. Amariuta[7]; [1]Univ. of California, San Diego, La Jolla, CA, [2]Massachusetts Gen. Hosp., Chestnut Hill, MA, [3]Univ. of California San Diego, La Jolla, CA, [4]Harvard Med. Sch., Boston, MA, [5]Brigham and Women s Hosp., Brookline, MA, [6]Univ California San Diego, La Jolla, CA, [7]Univ. of California San Diego, San Diego, CA

Abstract:

Unlike marginal eQTL mapping identifies redundant signals due to linkage disequilibrium, gene expression heritability is the proportion of variance in expression due to causal genetic effects jointly. While genetic variation in regulating gene expression is highly cell-type specific, single cell RNA-sequencing (scRNA-seq) enabled comparative analysis at unprecedented resolution. Pseudo-bulking (PB), combining reads from the same donor across cells, is a common practice in many analyses. But we found it leads to over-estimated variances due to non-independence between cells of the same donor. We propose a robust statistical method to accurately estimate gene expression heritability from scRNA-seq reads and quantify cell-type-specificity in genetic regulation. Our method scGeneHE, **S**ingle **C**ell **Gene** expression **H**eritability **E**stimation, an extension of the recently developed SAIGE-QTL mixed model by Zhou et al., exploits (1) a Poisson regression model to effectively address the sparsity of scRNA-seq read counts and (2) multiple random effects to partition gene expression variance into inter-subject (genetic) and intra-subject (repeat measurements) variances. It transforms Poisson regression model variances to linear space, comparable to conventional GCTA estimates. scGeneHE is computationally efficient for millions of scRNA-seq reads and generates empirical standard errors via bootstrapping to identify significantly *cis*-heritable genes. We simulated *cis*-heritable gene expression with 80% sparsity for 500 individuals and variable number of cells (2~50) per individual. Our simulations demonstrated that scGeneHE is well-powered across multiple settings. For instance, setting gene expression heritability to 1% with 50 cells per individual (representative of the OneK1K cohort), we detect *cis*-heritable genes with 65% power and under estimation of heritability proportional to sparsity. At all settings, PB over-estimates heritability by 10-fold. As expected, scGeneHE produced standard errors up to 84% smaller than PB analysis. We applied scGeneHE to OneK1K cohort of 982 donors and 1.27 million blood cells. Across 419 genes near genome-wide significant variants associated with rheumatoid arthritis, scGeneHE identified 147% more cis-heritable genes across cell types, with 237 newly detected genes compared to bulk (GTEx whole blood) or PB analyses ($p<0.01$). 52 of them are under cell-type-specific regulation ($p<0.01$ in only one cell type) with notable heritability estimates, e.g., 15.5% for ARID3A - CD8+ T central memory cells specific, 38.6% for SDF2L1 - CD14+ Monocyte cells specific, and 16.3% for HHAT specific to CD4+ cytotoxic T cells.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4119F: Predict Disease Risks and Trait Values by Screening Thousands of Polygenic Scores

Authors:

W-M. Chen[1], S. Onengut-Gumuscu[1], A. Manichaikul[2], B. Worrall[2], S. Rich[1]; [1]Univ. of Virginia, Charlottesville, VA, [2]Univ Virginia, Charlottesville, VA

Abstract:

Introduction: A polygenic score (PGS) summarizes a person's genetic information in a single number for a trait, and its utility in genomics and medicine has been well-recognized in recent years. Although PGSs have been widely generated for many traits, traits that are not widely measured, such as treatment/drug responses could greatly benefit from trait prediction, but PGSs are still not broadly available due to the limited sample sizes in available studies. To allow for more versatile trait prediction, we present a prediction method that can be used in any study with genome-wide SNP data. **Method:** We first generate 1,000s of PGSs for each study subject using their genome-wide SNP data and public resources, such as the Michigan Imputation Server and the PGS Catalog. We then compute the Area Under the Curve (AUC) of prediction for a binary trait, or the R-squared of association for a quantitative trait, for each PGS. The screening process selects the best predicting PGSs. We implement this algorithm in a scalable and user-friendly software tool, SECRET (Screen and Evaluate Catalogued Risk scores to Enhance Trait predictions). This tool takes PGSs and binary/quantitative traits as input, evaluates and identifies the most predictive PGSs, and provides advanced inferences to potentially further improve the prediction power. We applied the SECRET method to a type 1 diabetes (T1D) dataset and a dataset consisting of 2,100 samples each with 13 blood test-related traits measured, with the top principal components of ancestry being properly adjusted. **Results:** For 9 traits with existing PGSs available in the PGS catalog, which included T1D, creatinine, total cholesterol, HDL, triglycerides, vitamin B12, CRP, DBP, and SBP, all traits were predicted correctly, with the same-trait-related PGSs being identified as the top predictors. For the 5 traits without existing PGSs, the top predictors all were closely related to the trait, such as the SBP PGS for Factor XII (a plasma protein in coagulation), and the phosphate PGS for B6. We discovered that some traits could be better predicted when stratified by sex, such as in T1D, where the prediction power was consistently larger in males than in females across dozens of T1D PGSs. New risk scores were automatically developed in SECRET as linear combinations of best PGSs; *e.g.*, the newly developed PGS improved the Factor XII prediction from R^2=0.007 to 0.039. **Conclusion:** The method and tool developed here will be a valuable resource for researchers, such as making full use of their genomic data in studies of limited sample sizes,

enabling systematic PGS evaluation, and generating insights and hypotheses regarding their genetic traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4120F: Prediction at scale of lumping and splitting decisions to define monogenic gene-disease entities

Authors:

S. Khanam[1], A. J. Marcello[1], J. X. Chong[1,2]; [1]Univ. of Washington, Seattle, WA, [2]Brotman-Baty Inst., Seattle, WA

Abstract:

Over 4,500 genes have been reported to underlie more than 6,300 Mendelian conditions (MC), and ~25% of these genes have been asserted to underlie two more putatively distinct MCs. However, assertion of distinct MCs due to variants in the same gene has largely been based on subjective grouping of individuals perceived to share arbitrarily selected clinical features, while the underlying distribution of phenotypic effects from variants in the gene remains unknown. In 2019, the ClinGen Consortium, OMIM, and the Monarch Initiative established criteria for so-called lumping and splitting (L&S) of same-gene disease entities to facilitate assessment of the evidence strength and validity of asserted gene-disease relationships. To date, ClinGen Expert Panels have used these criteria for L&S of 616 disease entities during gene-disease validity curation. This leaves 2/3 of the >1800 MCs asserted to be distinct from other same-gene MCs unassessed, as well as the hundreds of new gene-disease relationships reported each year. We sought to develop a machine learning model to predict L&S for any pair of same-gene disease entities.

If distinct MCs exist for a single gene, it is often due to differences in mode of inheritance, the functional consequences of the causal variants (e.g. loss vs. gain of function), or which protein functions are perturbed. Because there are few structured databases linking variants, their functional consequences, and corresponding disease entities, our model included both molecular mechanism curations from Gene2Phenotype and multiple gene-level computational predictors of functional consequences. We also incorporated multiple phenotype similarity metrics and a measure for pleiotropy within each disease entity, both calculated from structured phenotype data encoded in the Human Phenotype Ontology. We developed an ensemble model trained on 378 disease entity pairs for which a L&S decision was made by a ClinGen Expert Panel and evaluated performance on the remaining 162 pairs. To mitigate the small sample size, we used repeated random subsampling validation. Phenotype similarity and

pleiotropy, mode of inheritance, and variant functional consequence/molecular mechanism alone correctly classified 81% (136/162), 60% (91/162), and 78% (126/162) disease entity pairs, respectively. Combining all three categories of features achieves 84% (136/162) accuracy. Examination of incorrect predictions identified disease pairs that appear to be outliers versus other L&S decisions, highlighting our model's potential utility in prioritizing prior L&S decisions for reevaluation. Our model also provides L&S predictions for all same-gene MCs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4121F: Proteome-wide adjusted pQTLs: A tale about mediation, collider, increase in power and how to distinguish them

Authors:

A. Campos Gonzalez, C. Gillies, Regeneron Genetics Center, G. Abecasis, M. Ferreira; Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

Protein quantitative trait loci (pQTLs) can be used as instrumental variables for protein levels to identify new therapeutic targets or disease biomarkers. Increasing the power of pQTL discovery is imperative. With this objective in mind, we ran *proteome-wide adjusted* genome wide association studies (PWGWAS) in the UKB plasma proteomics project. Briefly, for each focal protein of interest, we residualize the abundance of all other measured proteins prior to running a genome-wide association study. This approach has three expected outcomes compared to a standard GWAS: (i) increased signal for *direct* (cis) genetic associations; (ii) removal of pQTLs whose association is driven by the abundance of correlated proteins; and (iii) *indirect* (trans) pQTLs that gain significance upon conditioning. These last group requires additional consideration as some of these pQTLs may be spurious associations from the statistical adjustment, a phenomenon known as collider bias. Distinguishing relevant signal from collider is of interest as the former may provide useful biological insights whereas the latter is hard to interpret and may need to be QCed for downstream analyses. Both standard and PWGWAS GWAS identified around 41,700 sentinel pQTLs. Most cis pQTLs identified by standard GWAS were also identified by PWGWAS (90%). PWGWAS identified 184 (8.6%) additional cis pQTLs compared to standard GWAS. Despite both approaches identifying a similar number of trans pQTLs, the overlap between them was considerably lower. Of the 39,531 trans pQTLs identified by standard GWAS, only 33% (13,250) remained significantly associated in PWGWAS. PWGWAS removed signals driven by the abundance of other proteins. For example, a locus in *ASGR1* shows genome-wide significant associations with 520 proteins in standard GWAS, but

only with 61 proteins in PWGWAS. Trans pQTLs identified by both approaches (common trans pQTLs) are likely driven by a protein that was not measured here. On the other hand, trans pQTLs reaching significance only in PWGWAS (N=26,208) may be explained by two scenarios: (i) increased signal, akin to the scenario discussed above for common trans pQTLs; or (ii) a spurious association due to collider bias. It is worth noting that these two scenarios also occur when using other adjustment approaches such as proteomic PCs. We estimate that 14,676 (57%) of the PWGWGAS-only trans pQTLs may be driven by a collider protein (scenario ii). In this presentation, we will compare the results of standard, proteome PC-adjusted, and PWGWAS results and discuss the strategies we devise to distinguish between increased signal and collider, a necessary step to identify relevant trans pQTLs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4122F: PRS-Guided Sequencing and Large-Cohort Imputation: Unveiling Rare Variant Associations in Complex Diseases ⭐

Authors:

K-H. Lin, J. O'Connell, J. Shi, W. Wang, A. Kwong, K. Kalkus, K. Fletez-Brant, L. Noblin, M. Klucsor, M. Wise, J. Lubensky, P. Yee, 23andMe Research Team, A. Auton, X. Wang; 23andMe, Sunnyvale, CA

Abstract:

The genetic basis of complex diseases, such as obesity, has been widely studied via Genome-Wide Association Studies (GWAS). GWAS have identified many associations from common variants; however, they do not fully explain disease heritability. Rare variants are hypothesized to partly account for the missing heritability. However, detecting associations with rare variants typically faces challenges due to poor imputation quality and insufficient statistical power, hence likely requiring sequencing and large sample sizes in the hundreds of thousands or millions. Although modern population-based biobanks have reached large sample sizes, the high cost of sequencing at such scales remains a significant barrier. Polygenic Risk Scores (PRS) aggregate the cumulative effects of common genetic variants, and individuals whose disease status deviates from their PRS predictions may carry rare variants with large effects (risk or protective) that contribute to the total disease liability. Therefore, sequencing these individuals may provide an efficient method to identify carriers of disease-relevant rare variants. Subsequent imputation of these rare variants into a large cohort genotyped on arrays can be cost-effective and can significantly boost power for discovery of rare variant associations.
In this study, we leveraged common variant PRS in 4 million research participants in the

23andMe cohort to identify rare variant associations through a two-step approach. First, in order to enrich for rare variants that influence disease phenotypes, we selected individuals for exome sequencing from the extremes of disease PRS distributions who have discordant disease statuses (obesity, osteoarthritis, asthma, and type 2 diabetes) based on their common PRS prediction. Next, we constructed an imputation panel from the sequenced individuals and performed genotype imputation into the broader 23andMe cohort, followed by association testing. A total of 37,537 individuals were sequenced. We show that our PRS-based selection strategy is cost-effective, can enrich for rare variants, and lead to improved rare variant imputation quality that results in increased statistical power for detecting rare variant associations. We find novel gene-trait associations driven by rare variants that are typically not powered to be discovered without sequencing millions of individuals in large population cohorts. This work has significant implications for designing future work to deepen our understanding of complex diseases and to facilitate drug discovery efforts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4123F: Python-based pipeline for measuring imputation accuracy and assessing the impact of reference panel selection on genetic variant imputation in different study populations and across genomic regions

Authors:

M. Milojevic, A. Svane Frahm, D. D. Murray, J. Lundgren; RigsHosp.et - Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

Imputation is a widely used approach to fill in gaps in genetic data prior to performing genetic analyses and publicly available, free imputation platforms such as the TOPMed, Sanger or Michigan Imputation Server (MIS) have helped with standardization and making imputation more accessible for the average researcher. Imputation accuracy is known to be greatly affected by the reference panel used, and with multiple options being available on the imputation platforms, less experienced users may not know which are best suited for their study populations. Some users may also default to a large multiethnic panel as a “safe” choice without thorough consideration for the impact this will have on their genetics analyses. Given that the number of imputed genotypes can outnumber the assayed genotypes by several folds, it is important that the imputation is as accurate as possible or downstream analyses may produce incorrect results.
Here we present a bioinformatics pipeline to help users better understand the potential impact

of different reference panels on their imputation results and downstream analyses. Our python-based pipeline allows users to 1) perform QC on TOPMed or MIS raw output data, and 2) perform SNP genotype concordance analysis between multiple imputation data sets and measure imputation accuracy, if users have observational genotypes for comparison. Our pipeline generates visualizations of genotype concordance by chromosome and user-specified sample groupings, and outputs QC-ed data, along with QC summary files. We demonstrate the utility of the pipeline by applying it to a subset of our 12,000 individuals with HIV from international cohorts, who were genotyped at &gt770,000 loci on a custom Affymetrix SNP array that is enriched for immune dysfunction-associated genes. We performed imputation with the TOPMed and 1000 Genomes Project reference panels, increasing number of SNPs by &gt10x, and present how concordance levels can vary at a population and genetic locus/region level.
Our pipeline provides imputation users with a means to identify study participants or groups that are underrepresented in reference panels of interest, as well as genomic regions that are unreliably imputed across multiple reference panels and need to be directly assayed. Thus, the user will be able to make more informed decisions on reference panel selection and decide if imputation is a suitable approach given the study population composition and loci of interest.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4124F: Rapid Long-Range Linkage Disequilibrium Calculations at Biobank Scale using GPU Acceleration

Authors:

R. Kumar[1], P. Singhal[1], C. Carson[1], M. Conery[1], A. Rodriguez[2], T. Nandi[2], M. Ritchie[1], M. Thavappiragasm[2], R. Madduri[2], B. Voight[1], A. Verma[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Argonne Natl. Lab., Lemont, IL

Abstract:

Leveraging linkage disequilibrium (LD) information from one's own dataset (in-sample LD) is generally considered ideal for a variety of downstream analyses including statistical fine-mapping. However, due to computational challenges, most studies opt to use external reference panels such as those from 1000 Genomes. Current methods for computing in-sample LD, such as PLINK or LDStore2, are often impractical because they require [N^2]/2 computations for N variants (the upper triangle of a NxN matrix). This process increases with cohort size and thus can take weeks or months. To reduce the dimensionality of the search space, many of these methods assume that LD occurs between variants in close proximity by imposing distance

thresholds when selecting variant pairs for LD calculation. However, LD can be observed over very long distances and even across chromosomes, likely owing to non-recombination-based mechanisms of LD. To fully capture LD across all variant pairs in biobank-scale whole-genome sequencing datasets—which can contain up to hundreds of millions of variants across hundreds of thousands of individuals—new, more efficient computational strategies are essential to compute large-scale LD in a feasible time.
We present in this work a new approach (the first of its kind) that makes use of multiple GPUs and distributed computing to compute the R^2 for every pairwise set of variants in a dataset. When comparing calculations on chromosome 22 (1.8 million variants) of the 30x whole-genome-sequenced 1000 Genomes dataset only, we observe a ~6x speedup with our method when using a single consumer-level 12GB GPU (an NVIDIA RTX 2080TI) compared to the same calculation with PLINK on a high-end 64-threaded CPU (an Intel Xeon Gold 6338). Specifically, it takes <30 minutes on a single GPU compared to >3 hours on the 64-thread CPU. This speedup scales almost linearly with the number of GPUs - when using 4 such GPUs, we observe a ~24x speedup.
With this method, we successfully computed on the aforementioned 30x WGS 1000 Genomes dataset (comprising ~120 million variants and ~2500 samples) the entire LD matrix (over 1e16, or 10 quadrillion elements) in under 6 hours using 512 NVIDIA 40GB A100 GPUs on the Department of Energy (DOE) Argonne Leadership Computing Facility Polaris High-Performance Computing System. We make this tool, coded in Python using CuPy and custom C++ kernels, publicly available. Using this tool, researchers can design experiments that leverage the full extent of their genomic data without relying on external LD reference panels. This can lead to more accurate and population-specific findings, particularly for underrepresented groups in existing databases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4125F: RareEffect: Rare variant effect size estimation and polygenic risk prediction

Authors:

K. Nam[1], M. Kho[1], W. Zhou[2], B. Mukherjee[3], S. Lee[1]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Massachusetts Gen. Hosp., Chestnut Hill, MA, [3]Univ. of Michigan, Ann Arbor, MI

Abstract:

To elucidate the impact of rare variants on traits and enhance risk prediction, it is essential to estimate effect sizes in addition to calculating p-values. However, due to their low frequency,

estimating the effect sizes of rare variants is challenging. In this study, we introduce RareEffect, a novel method that addresses this challenge. RareEffect uses a variance component model, commonly employed rare variant tests like SKAT, and is designed to provide two levels of effect sizes: gene/region-level and variant-level, which offer improved interpretation. A primary challenge in using variance component models is computational demand, especially when estimating variance components in large samples. To address this, we implemented the FaST-LMM approach, which reduces computation complexity from $O(n^3)$ to $O(nk^2)$ where n is the number of samples (about 500,000 in the UK Biobank) and k denotes the number of variants in a gene (100-1000). Additionally, we developed a fast version of Firth bias correction, reducing computation time by nearly 200 times compared to conventional implementation. From simulation studies using UK Biobank whole exome sequencing (WES) data, RareEffect showed a lower RMSE and faster analysis completion compared to traditional methods such as linear and ridge regression. Using UK Biobank WES data of 393,247 White British samples, we further show that effect sizes derived from RareEffect improve polygenic risk prediction. For example, a combined model of RareEffect-based polygenic risk score (PRS) and common-variant PRS significantly improved R^2 values for HDL and LDL cholesterol prediction in high-risk individuals within the top/bottom 0.5%—achieving 0.6237 and 0.6645, respectively, compared to 0.5417 and 0.5823 with the common-variant PRS model alone. In summary, RareEffect enables the estimation of gene-level and variant-level effect sizes and the identification of high-risk individuals who might remain undetected by conventional PRS methods using common variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4126F: Regulatory context stratifies genotype-phenotype maps into biological pathways

Authors:

C. Boix, S. Sunyaev; Harvard Med. Sch., Boston, MA

Abstract:

The advent of biobank-scale datasets promises to help us understand how genetic associations converge onto molecular mechanisms in each trait. While the expectation is that genotype-phenotype maps contain underlying structure that will reveal both known and novel molecular pathways, top-down approaches to decompose genetic associations in biobank-scale datasets have struggled to recover recognizable structure or even to identify classical biological pathways. Here, starting from a bottom-up approach, we show that SNPs nearby known signaling and metabolic pathways exhibit significant phenotypic heterogeneity

due to the modular re-use of pathways across cellular contexts. However, stratifying these SNPs by their cell type and tissue-specific epigenomic profiles cleans up significant heterogeneity in their phenotype associations. Given this observation, and the fact that genotype associations commonly enrich for specific regulatory contexts, we hypothesize that we can recover biologically-relevant structure from genome-wide genotype-phenotype maps by conditioning on regulatory information across cellular contexts. Based on this intuition, we develop a statistical framework to model GWAS effect sizes through pathways but conditioned on known context-specific regulatory maps of the epigenome and transcriptome. We use this framework to score 50 known pathways in 833 cell and tissue contexts, identifying joint pathway and context-specific associations including: WNT and keratinocytes for balding; WNT and ovaries for female reproductive traits; insulin and pancreas for HbA1c; and cholesterol and small intestine for cholestasis. We extend our model to decompose the genome-trait-context tensor across 3,721 traits from UKBB along both known pathways and unknown pathways, enabling us to identify novel and known molecular building blocks of biomarkers, diseases, and phenotypic traits. Based on this work, we propose that identification of joint (pathway, trait, context) triples will improve our modeling and understanding of molecular mechanisms underlying complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4127F: Relatedness and rotator cuff tears: Leveraging biobank data to understand source of associations

Authors:

T. Basnet[1], R. Prakash[2], W. Liu[3], S. Herzberg[1], M. B. Breyer[1], N. B. Jain[2], A. Giri[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Univ. of Michigan, Ann Arbor, MI, [3]The Univ. of Texas Hlth.Sci. Ctr., Houston, TX

Abstract:

Prior studies evaluating familial predisposition to rotator cuff tears rely on surveys that inquire about family history, which may be prone to recall bias. We propose leveraging relatedness-metrics based on genetic data to compare odds of rotator cuff tears between genetically related individuals (1st-, 2nd- and 3rd-degree relatives) and genetically unrelated individuals (>3rd degree). We hypothesized that if this association is rooted in part due to genetics, evidence for familial disposition should be stronger in closely related individuals (1st degree vs. unrelated) than less closely related ones (3rd degree vs. unrelated). We used data from Vanderbilt University Medical Center biorepository (BioVU) to estimate relatedness among

individuals of predominantly European ancestry using KING. Pairs of individuals with kinship coefficient > 0.044 were considered related (0.177 to 0.354 for 1st degree; 0.088 to 0.177 for 2nd degree; and 0.044 to 0.088 for 3rd degree), while kinship coefficient ≤ 0.044 were considered as unrelated. We then used logistic regression to evaluate the association between familial-relatedness and rotator cuff tears adjusted for age, sex and 10 principal components (N-cases = 7,930; N-controls = 82,308). Compared to unrelated individuals, related individuals were more likely to have rotator cuff tears (odds ratio [OR] = 1.21; 95% confidence interval [95% CI] = 1.13, 1.30; p-value = $1.2x10^{-7}$). Compared to unrelated individuals, odds ratios were 1.28 (95% CI = 1.17, 1.40; p-value = $7.7x10^{-8}$) for 1st degree-related individuals, 1.16 (1.05, 1.27; p-value = $2.2x10^{-3)}$ for 2nd degree and 1.56 (1.10, 2.14; p-value = $9.0x10^{-3}$) for 3rd degree related individuals suggesting no significant trend between increasing degree of relatedness and increasing odds of rotator cuff tears. We evaluated the same associations between genetically-determined-relatedness and rotator cuff tears using the UKbiobank data. Compared to unrelated individuals, related individuals were more likely to have rotator cuff tears (OR = 1.09, 1.03, 1.16; p-value = 4.3x10-3), however, odds of rotator cuff tears was lower in 1st degree related individuals (OR = 1.04; 95% CI = 0.94, 1.14; p-value = $4x10^{-1}$), than in 2nd degree related individuals (OR = 1.33; 95% CI = 1.17, 1.52; p-value = $1.2x10^{-5}$), or 3rd degree related individuals (OR = 1.11; 95% CI = 1.03, 1.19; p-value = $6.1x10^{-3}$). Furthermore, SNP-based heritability with LDSC was 0.2% in BioVU and 0.62% in UKBiobank. Our assessment of two biobanks suggests rotator cuff tears do run in families, however this association may be due to lifestyle factors common to related individuals rather than genetics.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4128F: scTWAS: A powerful statistical framework for single-cell transcriptome-wide association studies

Authors:

Z. Lin[1], C. Su[2]; [1]Florida State Univ., Tallahassee, FL, [2]Emory Univ., Atlanta, GA

Abstract:

Transcriptome-wide association studies (TWAS) have successfully identified genes associated with complex traits and diseases. Yet most existing TWAS use bulk transcriptome data, failing to capture the cellular context of gene-trait associations. Recent advancements in single-cell RNA sequencing (scRNA-seq) enable the study of cell-type-specific genetic regulation of gene expression, which is of particular importance for the disease-related yet less abundant cell types (e.g., microglia for Alzheimer's disease (AD)). However, scRNA-seq data present substantial

challenges for traditional TWAS methods. Single-cell data are highly sparse, skewed, and heteroskedastic, while traditional methods were designed for bulk data, which can be modeled as normally distributed. This leads to heuristic data normalization in practice, which lacks theoretical justification and suffers from loss of information and consequently limited statistical power. Here, we propose scTWAS, a novel moment-based statistical framework to conduct cell-type-specific TWAS using scRNA-seq data with much improved statistical power. We adopt a novel expression-measurement model with a Poisson component for the sparse and skewed counts and with sequencing depths included to capture heteroscedastic variance. Based on the moment conditions, scTWAS then imputes the cell-type-specific gene expression with an iteratively reweighted penalized least squares procedure and infers cell-type-specific gene-trait associations using imputed expression. Simulation studies showed that scTWAS preserved type-I error control and greatly improved both stage 1 imputation accuracy and stage 2 TWAS power across cell types, with a greater improvement for less abundant cell types. We applied scTWAS to 13 immune cell types from the OneK1K study and found that it generated effective imputation models for an average of 27% (sd=0.18) more genes over FUSION elastic-net model based on the normalized pseudo-bulk data, with up to 63% more genes in less abundant cell types. In the subsequent Stage 2 analysis on 31 hematological traits, scTWAS identified an average of 11% (sd=0.11, max=40%) more gene-trait associations across cell types. We further applied scTWAS to microglia from the ROSMAP study and AD GWAS, which yielded 24% more accurately imputed genes and 17% more AD-associated genes in microglia than FUSION. Among significant scTWAS genes in microglia and its subtypes, 12 replicated existing literature on microglia AD genes (e.g. *BIN1*, *PICALM*) while 5 novel genes have not been previously implicated for microglia-specific AD pathogenesis (*FERMT2, ARL17B, LRRC37A, CHGA, NUP160*).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4129F: Sex-differences in deviation from Hardy-Weinberg Equilibrium

Authors:

E. Garg[1], J. Romain[2], L. Sun[2], A. D. Paterson[1,2]; [1]Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada

Abstract:

Quality control of variants in genome-wide association studies often includes testing for Hardy-Weinberg Equilibrium (HWE). However, this is usually implemented in a homogenous population without stratifying by sex. Deviation from HWE can be caused by violation of

numerous assumptions or technical errors. We have previously shown that sex-differences in minor allele frequency (sdMAF) are not only found on the X chromosome (predominantly near the boundaries between pseudoautosomal regions (PAR) and non-pseudoautosomal regions), but also on the autosomes. This work explores genome-wide sex-differences in deviation from HWE (sdHWD) across 2,490 unrelated individuals from five super-populations in the 1000 Genomes Project (1KGP) data. In order to reduce technical errors we used the high-coverage short-read whole-genome sequence 1KGP data aligned to the T2T (telomere-to-telomere v2) reference genome in a sex-specific manner. Our analysis is based on bi-allelic variants with non-missing genotypes and MAF>1% in both sexes of every super-population. We used a robust allele-based approach for HWE testing. To test for sdHWD we implemented a heterogeneity test using Cochran's Q. The heterogeneity test identified several regions with genome-wide significant sdHWD (P<5E-8). Variants in the Genome in a Bottle difficult-to-sequence regions were excluded. Multiple variants at the PAR boundaries on the X chromosome have significant deviation from HWE in males but not females, along with significant sdMAF, driven by higher MAF in males. On the autosomes 37 variants had significant sdHWD but this was not driven exclusively by one sex, and none of these autosomal variants had significant sdMAF. We have further examined these autosomal variants based on their genomic location and flanking sequence, as well as sequence variation identified by the Human Pangenome Reference Consortium. Our analysis shows that most sdHWD exists at the PAR boundaries and is rarely observed elsewhere in the genomes of 1KGP data. Limitations of our study include small sample size within each super-population stratified by sex.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4130F: Single-nucleus transcriptome-wide association study of human brain disorders

Authors:

S. Venkatesh[1], Z. Wu[1], M. Anyfantakis[1], G. Hoffman[1], D. Lee[1], D. Burstein[1], D. Mathur[1], R. Kosoy[1], F. Tsetsos[1], F. Prashant[1], C. Dillard[1], B. Zeng[1], A. Hong[1], C. Casey[1], M. Alvia[1], Z. Shao[1], S. Argiriou[1], K. Therrien[1], P. Auluck[2], D. Bennett[3], S. Marenco[2], V. Haroutunian[1], J. Bendl[1], J. Fullard[1], G. Voloudakis[1], P. Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Human Brain Collection Core, Natl. Inst. of Mental Hlth.-Intramural Res. Program, Bethesda, MD, [3]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL

Abstract:

Transcriptome-wide association studies (**TWAS**) leverage expression quantitative trait loci and genome-wide association studies to discover gene-trait associations (**GTAs**). Existing efforts have primarily utilized RNA sequencing from tissue homogenate which often obfuscates cell type-specific (**CTS**) effects. We utilized single-nucleus (sn) gene expression profiling from the PsychAD cohort (**EUR**opean n=920, **AFR**ican n=321, Admixed Americans n=118) to construct 94 sn transcriptomic imputation models (**snTIMs**) towards discovering CTS GTAs across a range of neuropsychiatric diseases (**NPDs**). Our models can impute 20,189 unique genes across 128,918 gene-cell-type combinations in EUR snTIMs. Across 12 NPDs, we report 6,132 unique GTAs in total including 567 unique GTAs in Alzheimer's Disease (**AD**). Notably, we show that GTAs found in "high resolution" snTIMs are enriched for novelty compared to GTAs found in tissue homogenate. We utilize a TWAS fine mapping tool, FOCUS, to identify putatively causal genes within genetic loci, and find 328 genes and 178 loci implicated in AD. We then perform heterogeneity and MASH analyses to identify genes that vary across NPDs, and perform LD-aware pathway analysis with JEPEGMIX2-P. Using this unique resource we are able to identify CTS GTAs that are not detected in similarly powered homogenate-based models trained in the same brain region. For example, *CTSH* in AD is exclusively fine-mapped in neurons and belongs to the significantly enriched pathway for lamellar bodies. The latter pathway is implicated in AD-related neuronal death after exposure to Aβ-42. We next utilized our power to investigate genes, pathways, and loci that are shared among disorders. We show a strong relationship between AD, Parkinson's Disease (**PD**), and Alcohol Use Disorder (**AUD**). We confirm and fine-map the association of *CELF1* with AD and AUD, and go on to show that *CELF1* is not associated with PD. Notably, due to *CELF1*'s role in RNA binding, this may implicate unique post-transcriptional biology in AD and AUD that is not present in PD. Next, we used our separately trained ancestry-specific snTIMs to perform ancestry-matched TWAS using individual level data at the Million Veterans Program. We show that ancestry-specific TWAS and subsequent pathway analysis has high concordance amongst their top ranked associations. Finally, we use individually imputed transcriptomics to perform phenome wide association studies, which reveal CTS phenotypic associations. For example, the aforementioned *CELF1* shows association with Anxiety Disorder, but only in inhibitory neurons. Overall, we show that snTIMs can be used to elucidate a wealth of novel biology.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4131F: snputils: A Python library for processing diverse genomes

Authors:

D. Bonet Sole[1,2], M. Comajoan Cara[3], M. Barrabés[4], A. Dominguez Mantes[5], R. Smeriglio[6], D. Agrawal[1], C. López[7], C. Thomassin[8], A. Calafell[8], A. Luis[3], J. Saurina[3], M. Franquesa[3], M. Perera[4], M. Geleta[9], A. Jaras[2], B. Oriol Sabat[10], J. Abante[11], S. Moreno[1], D. Mas Montserrat[1], A. G. Ioannidis[2,1]; [1]Stanford Univ., Stanford, CA, [2]Univ. of California Santa Cruz, Santa Cruz, CA, [3]Univ.t Politècnica de Catalunya, Barcelona, Spain, [4]Munster Technological Univ., Cork, Ireland, [5]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [6]Politecnico di Torino, Torino, Italy, [7]Univ.t Politècnica de Catalunya, Barcelona, Spain, Barcelona, Spain, [8]École Polytechnique, Palaiseau, France, [9]Univ. of California, Berkeley, Berkeley, CA, [10]Univ. of California, Los Angeles, Los Angeles, CA, [11]Univ.t de Barcelona, Barcelona, Spain

Abstract:

With the growth of diverse datasets in genomics and precision health, the handling and concurrent analysis of sequencing and ancestry data present significant computational challenges. Researchers often must develop their own custom software to manage these data types. We present **snputils**, a new Python library designed to streamline robust input/output operations, basic data manipulation, and advanced visualization for genome variation and ancestry data. The library supports a wide range of file formats, including VCF (.vcf, .vcf.gz), PLINK1 (.bed, .bim, .fam), PLINK2 (.pgen, .pvar, .psam), .msp local ancestry inference formats, and global ancestry proportion files. **snputils** integrates seamlessly with existing SNP manipulation tools while introducing new capabilities, enhancing data interoperability and optimizing analytical workflows. The library includes advanced dimensionality reduction methods such as ancestry-specific PCA and MDS, along with comprehensive admixture mapping functionalities. These tools are designed to optimize the extraction of meaningful genetic information from diverse data structures and large datasets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4132F: Statistical identification of cell type-specific spatially variable genes in spatial transcriptomics

Authors:

P. Wu[1], L. Shang[2], X. Zhou[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]MD Anderson Cancer Ctr., Houston, TX

Abstract:

An essential task in spatial transcriptomics involves identifying genes with spatial expression patterns, known as spatially variable genes (SVGs). However, a large proportion of SVGs are cell

type marker genes, offering limited biological insights beyond those attainable through standard omics technologies like single cell transcriptomics. Here, we propose a paradigm shift in SVG analysis to focus on the detection of a subset of SVGs that display diverse spatial expression patterns within a given cell type. These cell type-specific SVGs (ct-SVGs) represent crucial transcriptomic signatures underlying cellular heterogeneity, providing insights uniquely accessible through spatial transcriptomics. To systematically detect ct-SVGs, we present Celina, a statistical method that utilizes a spatially varying coefficient model to accurately capture each gene's spatial expression pattern in relation to the distribution of cell types across tissue locations, thus ensuring effective type I error control and high statistical power. We evaluated the performance of Celina through comprehensive simulations and applications to five real datasets, where we also adapted and examined existing methods originated from other analytic settings to detect ct-SVGs. Celina proves powerful compared to these ad hoc method adaptations in single cell resolution spatial transcriptomics and stands as the only effective solution for spot resolution spatial transcriptomics. In the real data applications, Celina uncovers ct-SVGs associated with tumor progression and patient survival in lung cancer, identifies metagenes with unique spatial patterns linked to cell proliferation and immune response in kidney cancer, and detects genes preferentially expressed near amyloid-β plaques in an Alzheimer's model. The ct-SVGs detected by Celina open doors for novel biologically informed downstream analyses, unveiling functional cellular heterogeneity at an unprecedented scale. Finally, Celina is generalizable to other spatial omics data, enabling the characterization of the spatial organization of complex tissue in a cell type-specific fashion.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4133F: Streamlined pipeline for Tractor GWAS Unveils Ancestry-Specific Variants in Admixed Populations

Authors:

N. Shah[1], A. X. Maihofer[2], T. Tan[1], J. Mauer[1], H. Lin[1], PGC-PTSD ancestry working group, C. Zai[3], M. Santoro[4], C. Nievergelt[5], E. Atkinson[1]; [1]Baylor Coll. of Med., Houston, TX, [2]UCSD, La Jolla, CA, [3]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [4]Federal Univ. of Sao Paulo, Sao Paulo, Brazil, [5]Univ California San Diego, La Jolla, CA

Abstract:

Admixed individuals have long been excluded from GWAS studies, leading to significant disparities and Eurocentric biases in subsequent efforts like constructing polygenic risk scores (PRS) and fine-mapping. While recent global sequencing of admixed individuals shows promise

in addressing this issue from a data perspective, limited statistical methods exist to analyze these populations at scale. Admixed populations have complex genetic makeups from multiple continents, complicating association analysis.
To address these challenges, our lab developed Tractor, a GWAS method that accounts for the local ancestral background of alleles in association analyses. Unlike traditional GWAS, Tractor involves multiple steps, including phasing, local ancestry inference, and subsequent regression analyses with local ancestries. We have streamlined these processes into a scalable, portable Nextflow workflow, enhancing user-friendliness and reducing the barrier to adoption. We deployed this workflow on hundreds of traits, including all 30 blood biomarkers for 4,500 AFR-EUR admixed individuals from the UK Biobank. Blood biomarkers were chosen for the pilot run due to their well-established role as risk factors for diseases, their strong heritability, and their effectiveness as therapeutic targets. We subsequently compared the ancestry-specific effect size estimates across component ancestries from Tractor to characterize trends in genetic effect size differences across phenotype classes, with control over the environment.
The Nextflow workflow has successfully replicated known hits and identified novel ancestry-specific hits, demonstrating Tractor's potential for new discoveries. For example, we identified the *APOE* variant for total cholesterol across both ancestries, aligning with its biological role in cholesterol metabolism. Additionally, we discovered a variant in the *ANGPLT8* gene specific to AFR ancestral haplotypes. A detailed frequency analysis revealed that this variant is three times more common in AFR ancestral haplotype blocks than in EUR, providing power for association uniquely in the AFR background. Through our work, we have identified numerous ancestry-specific variants that traditional GWAS might have missed, thus playing a vital role in identifying genetic variants and enhancing our understanding of complex phenotypes. This development represents a major stride toward greater inclusivity and precision in genetic research. The user-friendly Nextflow workflow enables users with limited computational experience to run it with ease, reducing barriers and promoting broader adoption.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4134F: Successful Implementation of a Method for the Reconstruction of Parent Genomes from Two or More Children

Authors:

C. Mason[1], Y. Wang[2]; [1]Ancestry.com DNA, LLC, Lehi, UT, [2]Ancestry.com DNA, LLC, San Francisco, CA

Abstract:

The proportion of genetic markers of a parent inherited by a set of children is reflected deterministically by the number of children and stochastically by the shared set of recombination breakpoints. While the identification of a large fraction of parental genomic markers has been shown to be possible from four or more children or when the genome of the other parent or close relative is available, we are not aware of any successful implementation for the genome reconstruction of parents from the genetic data of fewer children alone.
We developed a method that can reconstruct substantial portions of both parent genomes when as few as two long-range phased children genomes are available. Using a consistent set of 416,176 SNPs across the 22 autosomes, we reconstructed the genome of 2,000 parents - each from only two siblings. To assess accuracy we required the parent to have tested and that all these family members were research consented Ancestry DNA participants. The phased genomes of the siblings were first obtained by running IBDphase on their initially unphased data. Then, we used our method to reconstruct a median of 75.4% (IQR: 73.7-77.2%, range: 64.1-83.9%) of the total parental SNP alleles. Both parental alleles were determined for a median of 50.9% (IQR: 47.4-54.3%, range: 28.2-67.8%) of the SNP markers. Comparing with the parent tested data, we found the bi-allelic accuracy of our method to be >98% for 97.3% and >95% for 99.7% of the reconstructed parent genomes.
We further extended this approach to reconstruct *long-range* phased parental genomes from two children by including additional close or only distant relatives' genomes. To test this method, we utilized a set of 889,000 genomes of research consented Ancestry DNA participants to identify relatives sharing identity by descent regions totaling <340 cM with the children (likely to be second cousins or more distant). We then reconstructed the long-range phased genome of the same set of 2,000 parents using only the phased SNPs from the two children and these relatives. While in this analysis nearly half (48.3%) of the sibling pairs did not have sufficient relatives to long-range phase at least 5% of the parent reconstructed genome, the majority did - these 1,034 reconstructed genomes having a median of 42.3% (IQR: 32.3-48.1%, range: 5.0-64.5%) of the parent SNPs reconstructed and phased consistently across all autosomes for both allele copies.
The ability to reconstruct a long-range phased parental genome from ≥2 children and other relatives now allows for iterative reconstruction of genomes beyond the parent, showcasing the potential for obtaining genetic insights into distant ancestors and their DNA pedigrees.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4135F: The Bias of Using Cross-Validation in Genomic Predictions and Its Correction

Authors:

D. Wang, Y. Qian, D. Qixuan, M. Greenberg, Q. Long; Univ. of Calgary, Calgary, AB, Canada

Abstract:

Cross-validation (CV) is a widely used technique in statistical learning for model evaluation and selection. Various methods, such as Generalized Least Squares (GLS), Linear Mixed-Effects Models (LMM), and regularization techniques, are commonly applied in genomic predictions—a field that utilizes DNA polymorphisms to predict phenotypic traits. However, the bias in cross-validation can be caused by the correlation between samples. This issue is particularly relevant in genomic data, where the training and testing datasets may have different correlation structures, leading to an underestimation of the generalization error. In this work, we analyzed the bias of CV in eight methods: Ordinary Least Squares (OLS), GLS, LMM, Lasso, Ridge, elastic-net (ENET), and a hybrid method, the Bayesian Sparse Linear Mixed Model (BSLMM). Leveraging genomics data from the 1,000 Genomes Project and simulated phenotypes, our investigation revealed the presence of bias in all these methods. To address this bias, we implemented a method based on variance structures, known as Cross-Validation Correction (CVc). This approach aims to rectify the cross-validation error by providing a more accurate estimate of the generalization error. To quantify the performance of our adapted CVc for all these methods, we applied the trained models to an independently generated dataset, which served as a gold standard for validating the models and calculating the generalization error. The outcomes show that, by leveraging CVc, we corrected the CV bias for most of the frequentist methods mentioned above. Additionally, we utilized the equivalence property between the Bayesian method and Linear Mixed-Effects Models to transform BSLMM into an LMM, allowing us to apply CVc to BSLMM. Our work revealed substantial bias in the use of CV in genomics, a phenomenon under-appreciated by the field of statistical genomics and medicine. Furthermore, we have demonstrated that bias-corrected models can be successfully formed by adapting CVc, and demonstrated the improved transferability of CVc-based optimization across diverse populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4136F: The effects of rule-based phenotyping on cohort composition and genome-wide association study findings in the UK Biobank

Authors:

A. Newbury, G. Gursoy; Columbia Univ., New York, NY

Abstract:

Biobanks enable precision medicine by providing comprehensive and diverse genetic, clinical, and environmental data, which researchers and clinicians use to identify disease mechanisms, tailor treatments, and develop personalized healthcare strategies. With the release of comprehensive Electronic Health Records (EHRs) in biobanks that include information from various domains, such as conditions, medications, and procedures, more accurate and thorough disease cohort creation with rule-based phenotyping is now possible. Currently, it remains unclear how cohort composition varies between different rule-based phenotyping algorithms, how these different cohorts alter the results of downstream genome-wide association studies (GWAS), and whether they aid in bridging the missing heritability gap. There are many different approaches to rule-based phenotyping. The simplest approach involves identifying cases by 2 or more occurrences of a condition. Another approach common in phenotyping for GWAS is Phecode phenotyping, which identifies cases by an inclusion set of condition codes. There are also more complex rule-based phenotyping algorithms which leverage other domains in addition to participant conditions. These multi-domain phenotyping algorithms include phenotypes from the Observational Health Data Sciences and Informatics (OHDSI) Phenotype Library (PL), which contains definitions that have undergone peer review, and those created in collaboration with clinicians. Another source for multi-domain phenotyping algorithms is the UK BioBank algorithmically defined outcomes (ADO), which leverage conditions and self-report codes. In this work, we compare cohorts and their corresponding GWAS findings using these four different rule-based phenotyping algorithms across seven diseases. We find that phenotypes leveraging more than one domain (OHDSI PL and ADO) generally result in a greater number of cases and a greater number of significant GWAS hits than single-domain phenotypes. Multi-domain phenotyping algorithms additionally result in heritability estimates greater than or equal to single-domain definitions. Furthermore, we show that significant variants from OHDSI PL and ADO definitions remain biologically relevant through investigation of gene, eQTL, and regulatory element annotations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4137F: The genetic atlas of nurture and parent-of-origin effects in ~220,000 UK Biobank individuals ⭐

Authors:

R. Hofmeister[1,2], L. Knüsel[1], T. Cavinato[1], L. Darrous[1], O. Delaneau[3], Z. Kutalik[1]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Inst. of Genomics, Tartu, Estonia, [3]Regeneron, Tarrytown, NY

Abstract:

Introduction: Although genome-wide association studies (GWAS) have extensively analyzed overall and direct (transmitted allele) genetic effects on complex traits, genetic nurture and parent-of-origin effects remain less understood. This typically requires distinguishing transmitted from non-transmitted alleles and determining their parental origin. However, studies are often limited by small sample sizes due to the need for parental genomes or detailed genealogies.
Method: To address this limitation in large-scale biobanks without genealogical data, we introduce an approach that uses close relatives to differentiate transmitted from non-transmitted alleles and determine their parental origin. For each individual, we start by identifying close relatives and probabilistically determine their parental lineage using chromosome X, mtDNA, and sex-specific genetic maps. We then use alleles shared Identical-By-Descent (IBD) between the target and their maternal or paternal relatives for two goals. First, to determine the PofO of alleles: we leverage IBD within statistical phasing to perform inter-chromosomal phasing and deduce the PofO from the relative lineage. We used this to perform PofO effects association scans. Second, to infer non-transmitted parental alleles: we use the close relative's alleles not-shared IBD. We then employed non-transmitted parental alleles in Mendelian Randomisation (MR) to elucidate how parental characteristics impact offspring traits.
Results: Using UK Biobank data, we inferred PofO and non-transmitted alleles for ~220,000 and ~130,000 individuals, respectively (vs ~5'000 parent-offspring pairs). Using the resulting sample sizes in PofO-specific GWAS confirmed known and revealed new PofO effects. In particular, we found PofO effects for standing height and type-2-diabetes at the 11p15.5 imprinted region, and for BMI and telomere length at non-imprinted regions. On the other hand, our intergenerational MR analysis with non-transmitted parental allele effects identified a widespread combined impact of parental obesogenic- and socio-economic environments. Notably it highlighted their influence on offspring BMI and education attainment, that we replicated using 178,000 sibling pairs from 19 cohorts.
Conclusion: Overcoming the traditional requirement for parental information by leveraging inter-individual relatedness in biobanks allows us to evaluate parent-of-origin and nurturing effects in hundreds of thousands of samples. This approach also holds potential for expanding our understanding of genetic nurture and parent-of-origin effects at an unprecedented scale.
Grant SNSF 310030-189147

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4139F: The genetic epidemiology of Usher syndrome in a South Florida population

Authors:

Z. Cromar, R. Chen, B. Lam, X. Liu, S. Blanton; Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract:

Usher syndrome (USH) is the leading genetic cause of congenital deaf-blindness, affecting ~400,000 individuals worldwide. USH, an autosomal recessive disorder, is clinically characterized by partial or complete congenital sensorineural hearing loss followed by progressive vision loss due to retinitis pigmentosa. There are three main subtypes (USH1, USH2, USH3) categorized by age of symptom onset and severity. With ten known causative genes, understanding the genetic epidemiology of USH can help identify novel mutations and facilitate definitive diagnosis and treatment. This retrospective study characterizes the mutation spectrum of USH in an ethnically diverse South Florida population. We retrospectively analyzed medical records of 154 patients diagnosed with USH at Bascom Palmer Eye Institute and/or the University of Miami Department of Otolaryngology. Collected genetic data included gene affected, cDNA, zygosity, variant classification, etc. Clinical and demographic data included USH subtype and self-reported race and ethnicity. Corroboration of molecular diagnoses from genetic testing and clinical diagnoses was confirmed using NIH ClinVar, dbSNP, and gnomAD databases. Novel mutations were assessed for pathogenicity using PolyPhen2. Of the 154 patients, 8 identified as African-American (AA), 6 as Asian (A), 43 as Hispanic (H), and 92 as White, non-Hispanic (WNH); 5 patients did not answer. The subtype distribution among those reporting race/ethnicity was as follows: AA: 4 USH1, 4 USH2; A: 2 USH1, 2 USH2, 1 USH3, 1 not known (NK); H: 13 USH1, 24 USH2, 2 USH3, 4 NK; and WNH: 13 USH1, 40 USH2, 3 USH3, 7 NK. The distribution of subtypes was similar between the H and WNH groups, with USH2 accounting for 45% and 43% of cases, respectively. Only 61(40%) of these patients had a molecular genetic diagnosis in their chart:1 AA (25%), 2 A (33%), 16 H (37%), and 27 WNH (29%). Of the 13 patients with USH1, 77% had pathogenic variants in *MYO7A* (1 AA, 1 A, 3 H, 2 WNH, 3 NK). Of the 46 with USH2, *USH2A* was the most common gene (86.8%, 1 A, 10 H, 20 WNH, 9 NK). The 2 patients with USH3 (both WNH) had a variant in *CLRN1* in a homozygous state. The African American patient had heterozygous mutations in *MYO7A* and two Asian patients had heterozygous mutations in *USH2A* and homozygous mutations in *MYO7A*. This is the first genetic characterization of an ethnically diverse South Florida population with USH, which can help direct patient diagnosis and medical care. Genetic testing resolved 2 individuals with unknown subtype. As clinical trials for treatment commence, it is imperative that molecular testing is done for all individuals.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4140F: The pivotal role of the X-chromosome in the genetic architecture of the human brain

Authors:

Z. Jiang[1], P. Sullivan[2], T. Li[1], B. Zhao[3], X. Wang[1], T. Luo[1], S. Huang[1], P. Guan[1], J. Chen[1], Y. Yang[4], J. Stein[5], Y. Li[2], D. Liu[6], S. Lei[7], H. Zhu[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ North Carolina, Chapel Hill, NC, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]UNC-Chapel Hill, Durham, NC, [5]UNC Chapel Hill, Chapel Hill, NC, [6]Penn State Univ., Hershey, PA, [7]Univ. of Toronto, Toronto, ON, Canada

Abstract:

Genes on the X-chromosome are extensively expressed in the human brain. However, little is known for the X-chromosome's impact on the brain anatomy, microstructure, and functional network. We examine 1,045 complex brain imaging traits from 38,529 participants in the UK Biobank. We unveil potential autosome-X-chromosome interactions, while proposing an atlas outlining dosage compensation (DC) for brain imaging traits. The rfMRI and tfMRI traits that favor no DC are predominantly associated with the primary and secondary visual networks. Through association studies, we identify 72 genome-wide significant trait-locus pairs (including 29 new associations) that share genetic architectures with brain-related disorders, notably schizophrenia. We prioritize five genes, *FAM3A*, *PJA1*, *TMLHE*, *PLXNA3*, and *ZNF275* through eQTL mapping that also have causal impacts on the brain imaging traits. These genes are linked to schizophrenia and autism. Among all identified protein-coding genes, 40% are concentrated in Xq28. Furthermore, we discover unique sex-specific associations and assess variations in genetic effects between sexes. We find males have five times as many significant associations as females (26 vs 5). And brain phenotypic variations in males can be reflected by diverse X-chromosome genetic profiles. Our research offers critical insights into the X-chromosome's role in the human brain, underscoring its contribution to the differences observed in brain structure and functionality between sexes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4141F: The role of sleep traits in diabetes onset: Evidence from Mendelian Randomization using self-reported data

Authors:

N. Nisamudheen, D. Velayutham, P. Jithesh; Hamad Bin Khalifa Univ., Doha, Qatar

Abstract:

Self-reported data, often collected in cohort studies, is a valuable resource for understanding health outcomes, behaviors, disease prevalence, and risk factors, yet underutilized in epidemiological research. While observational causal studies have linked sleep traits with diabetes, evidence using self-reported diabetes data for causal connection is lacking. Here, we performed a two-sample Mendelian randomization analysis using Inverse Variance Weighting (IVW), IVW with multiplicative random effects (IVW-MRE), Maximum Likelihood, MR-Egger regression, and Weighted Median models. We analyzed genetic variants linked to five sleep traits (sleep duration, insomnia, snoring, daytime dozing, and chronotype) from the IEU Open GWAS project database on self-reported diabetes from the 'Non-cancer illness code: Self-report', UK Biobank. The study utilized MR-Egger and MR-PRESSO regression to evaluate pleiotropy and outliers, IVW Q statistics to detect heterogeneity, the MR-Steiger test to assess directionality, and leave-one-out sensitivity analysis to ensure the reliability of MR analysis. Our results with Maximum Likelihood analysis demonstrated positive causal associations between genetically predicted insomnia ($p = 0.002$, OR = 1.021, 95% CI: 1.008-1.035) and daytime dozing ($p = 0.014$, OR = 1.029, 95% CI: 1.006-1.052) with diabetes, while IVW and IVW-MRE analysis showed a trend towards significance. We also observed that snoring showed mixed evidence, suggesting no significant causal effect on diabetes risk. Furthermore, genetically predicted sleep duration was marginally associated with diabetes ($p = 0.053$, OR = 0.992, 95% CI: 0.984-1.000) with the weighted median method, indicating a potential small protective effect. However, no causal association was found between chronotype and diabetes. In conclusion, our study underscores the significant causal evidence between insomnia and daytime dozing with diabetes risk using self-reported data, which align closely with established non-self-reported causal relationships. These findings highlight the critical role of self-reported health outcomes in advancing lifestyle intervention strategies and clinical therapeutic approaches.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4142F: Tissue-specific epigenomic profiles inform pleiotropic partitioning of disease loci

Authors:

G. Kerner[1], A. Price[2]; [1]Harvard Sch. of Publ. Hlth., Boston, MA, [2]Harvard Sch Pub Hlth, Boston, MA

Abstract:

Common diseases are highly pleiotropic, motivating pleiotropic partitioning of disease loci (Suzuki et al. 2024 Nature, Smith et al. 2024 Nat Med). Tissue-specific disease mechanisms have been used to validate these partitions but not to identify or formally evaluate them. Here, we introduce a method for pleiotropic partitioning of disease loci informed by tissue-specific epigenomic profiles, and a metric to evaluate its performance. First, we probabilistically assign fine-mapped SNPs to causal tissues using the EM algorithm, based on posterior inclusion probabilities and tissue-specific epigenomic profiles. Second, we cluster fine-mapped SNPs based on both pleiotropic associations and probabilistic tissue assignments using a Bayesian non-negative matrix factorization approach, iterating to convergence. We evaluate performance based on the accuracy (r) of predicting observed pleiotropic associations using tissue-specific profiles trained using off-chromosome data; we also evaluate the accuracy of predicting SNP cluster membership. We applied our method to GWAS of 45 highly heritable diseases/traits (average N=272K), using bulk (Boix et al. 2021 Nature) and single-cell (Zhang et al. 2021 Cell) chromatin data. Tissue-specific profiles predicted pleiotropic associations for many traits, including type 2 diabetes (T2D) (r=0.34), neutrophil proportion (NP) (r=0.17) and hematocrit (HT) (r=0.17) (avg r=0.16 across 13 traits with significant r). Notably, avg r across all traits was 40% higher compared to using pleiotropic associations only (without tissue profiles), as in current methods. For T2D, we identified 2 clusters for which tissue-specific profiles strongly predicted cluster membership, including a cluster positively associated to BMI and systolic blood pressure that was enriched for neural and kidney chromatin marks, and a cluster negatively associated to BMI that was enriched for marks in digestive and pancreatic gamma delta cells (which regulate insulin levels). For NP, we identified 3 strongly predicted clusters, including one negatively associated to eosinophil count that was enriched for metabolic (muscle and adipose) chromatin marks and another negatively associated to monocyte proportion that was enriched for immune (B-cell and monocyte) marks. For HT, we identified 2 strongly predicted clusters, including a cluster negatively associated to triglycerides that was enriched for liver marks and a cluster positively associated to red blood cell count that was enriched for hematopoietic stem cell marks. Our method and metric show that tissue-specific epigenomic profiles inform pleiotropic partitioning of disease loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4143F: Trans-ancestry meta-regression analysis MR-MEGA replicates European IBD loci using multiple ancestries

Authors:

E. Dunbar[1], C. Simpson[1], S. Brant[2]; [1]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Rutgers Robert Wood Johnson Med. Sch., Somerset, NJ

Abstract:

Introduction: Inflammatory bowel disease (IBD) is a disease affecting the intestinal tract consisting of two subtypes: Crohn's disease and ulcerative colitis. Most known IBD risk loci have been identified in people of European ancestry. Some have been replicated and identified in East Asian and African American populations; however, much of the heritability remains unexplained. The purpose of this study was to use the trans-ancestry method MR-MEGA to leverage the power of larger sample sizes by meta-analyzing multiple populations to replicate known loci in IBD. Methods: MR-MEGA, a trans-ancestry meta-regression analysis method was used to combine summary statistics of inflammatory bowel disease GWAS. Five studies including four populations were included: African American, East Asian, Finnish, and Non-Finnish European. MR-MEGA's genomic control options were used to adjust for inflation of the summary statistics. Only SNPs that were available in all five studies were included in the analysis. Our target genome-wide significance threshold was 5e-08. FUMA was used to annotate the results. Resulting loci were compared to known loci from prior IBD studies to identify replicated loci. This list of 241 known loci was pulled from and compiled by de Lange, et al. in a 2017 study using European ancestry individuals with IBD. Results: FUMA annotation of the MR-MEGA results identified 122 genomic risk loci. Of these, 95 overlapped/replicated known loci from the previous study in European ancestry individuals with IBD. There were 146 known loci that did not replicate. Discussion/Conclusion: We have used the multi-ancestry method MR-MEGA to replicate loci previously identified only in one population. Replication of single ancestry loci across ancestries increases confidence in the validity of those loci and allows for future generalization across ancestries. Loci not replicated may be specific to Europeans or may not have had significant SNPs included in MR-MEGA due to not being present in all population datasets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4144F: Transcriptome wide association study identifies novel candidate susceptibility genes for inguinal hernia

Authors:

D. Chaar[1], C. Jiang[1], B. Cowan[2], S. Patel[2], M. Kvale[3,1], J. Yin[1], R. Mostaedi[4], N. Ahituv[3,5], E. Jorgenson[6], T. Hoffman[3,7], H. Choquet[1]; [1]Div. of Res., Kaiser Permanente Northern California, Pleasanton, CA, [2]UCSF-East Bay Surgery, Oakland, CA, [3]Inst. for Human Genetics, Univ. of

California, San Francisco (UCSF), San Francisco, CA, [4]Kaiser Permanente Northern California, Richmond, CA, [5]Dept. of Bioengineering and Therapeutic Sci., UCSF, San Francisco, CA, [6]Regeneron Genetics Ctr., Tarrytown, NY, [7]Dept. of Epidemiology and Biostatistics, UCSF, San Francisco, CA

Abstract:

Purpose: Genome-wide association studies (GWAS) have identified 92 genetic susceptibility loci across 66 genes for inguinal hernia, however, how the majority of these loci impact inguinal hernia etiology is unknown. To identify novel genes associated with inguinal hernia and interpret the transcriptional effect of those genes, we conducted a transcriptome-wide association study (TWAS).

Methods: We performed tissue-specific TWAS analysis to assess associations between imputed gene expression from 54 tissues (including 49 from the Genotype Tissue Expression (GTEx) Project v8) with inguinal hernia susceptibility using FUSION software. Meta-analyzed GWAS summary statistics from 33,379 inguinal hernia cases and 436,717 controls, all of European ancestry and from two cohorts (the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) and the UK Biobank) were used. A total of 325,513 gene-tissue pairs (representing 37,920 unique genes across 54 tissue reference panels) were tested for associations between imputed gene expression with inguinal hernia susceptibility. Associations with a Bonferroni p-value $< 1.54 \times 10^{-7}$ (=0.05/325,513) were considered significant. TWAS genes were considered novel if they were located over 1Mb apart from previously reported inguinal hernia GWAS loci.

Results: In tissue-specific TWAS analyses, 261 gene-tissue pairs were associated with inguinal hernia after correcting for multiple testing. These gene-tissue pairs were represented by 96 unique genes across the 54 tissue reference panels. Importantly, 9 of these genes (*AL592148.3, MON1A, GNL3, STIMATE, CFL1, LINC00674, AC005332.6, WIPI1,* and *UBE2CP2*) were novel, as they did not overlap with any previously-identified inguinal hernia-associated loci. Moreover, subcutaneous adipose tissues represented one of the highest proportions of the 261 Bonferroni-significant gene-tissue pairs.

Conclusions: Our results provide evidence of the utility of imputation-based TWAS approaches to characterize known GWAS risk loci and identify novel genes that may increase our understanding of inguinal hernia etiology. Our findings also highlight the fact that expression of genes associated with inguinal hernia susceptibility is not necessarily restricted to connective tissue.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4145F: Transferability of single- and cross-tissue transcriptome imputation models across ancestry groups

Authors:

I. Pagnuco, S. Eyre, M. Rattray, A. Morris; Univ. of Manchester, Manchester, United Kingdom

Abstract:

Transcriptome-wide association studies (TWAS) explore the associations between genetically regulated gene expression and complex traits and diseases. The methodologies developed for TWAS involve first imputing gene expression using expression quantitative trait loci (eQTL) as predictors, and then testing the association of the imputed gene expression with the disease or trait. The power of TWAS to detect association depends, in part, on the predictive power of the gene expression imputation models. Training these imputation models requires genotype and gene expression data from the same set of samples and can be performed using single-tissue or cross-tissue approaches. However, publicly accessible transcriptomics resources, such as the Genotype Tissue Expression (GTEx) Project, are biased towards individuals of European ancestry, and may therefore lead to less accurate gene expression prediction models for individuals from other ancestry groups. This study explored the transferability of eQTL into different ancestry test datasets by comparing the performance of two gene expression imputation model: PrediXcan, which utilises a tissue-specific approach; and UTMOST, which employs a cross-tissue approach. Both models were trained using a dataset comprising 49 tissues from the GTEx Project, exclusively composed of European ancestry individuals but were subsequently tested on two distinct datasets representing European ancestry and African American individuals. The findings indicate that, for most tissues, both approaches demonstrated improved performance when the ancestry of the training dataset matches that of the testing dataset. Moreover, the cross-tissue approach generally outperformed the single tissue approach. The study highlights that the eQTL detected by gene expression imputation models can be influenced by both ancestry and tissue context. The continuous development of population-specific reference panels across multiple tissues has the potential to increase gene expression prediction accuracy, thereby improving the power of TWAS analysis, and enhancing understanding of the biological processes underpinning complex traits and diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4147F: Two novel, unbiased, sibling-based genetic methods greatly increase precision of heritability estimates and quantify gene-environment interactions

Authors:

D. Steinecke[1], P. Turley[2], D. J. Benjamin[1]; [1]Univ. of California Los Angeles, Los Angeles, CA, [2]Univ. of Southern California, Los Angeles, CA

Abstract:

Understanding the causal mechanisms between genotype and phenotype is critical for biological science, medical application, and public policy. A key parameter in this line of research is heritability. However, the practical utility of heritability estimates is impeded by tradeoffs between controlling for confounds and maximizing statistical power, as well as by the ambiguous definition of heritability in the presence of gene-environment interactions (GxE). Furthermore, accounting for confounds requires family data, which is scarce, leading to imprecise heritability estimates.

Here, we present two novel, unconfounded heritability estimators—Within-Family Sibling Regression (WFSR) and Between-Family Sibling Regression (BFSR)—that are substantially more precise than existing sibling-based methods. Both control for confounds by only considering genetic and phenotypic variation between siblings. WFSR estimates heritability based on sibling differences within a family, similar to other sibling-based heritability estimators. However, WFSR obtains greater precision by using a new, IBS-based relatedness measure that captures more within-family phenotypic variation. BFSR extends this method further by taking these sibling differences and comparing them between families, obtaining greater precision because BFSR uses more data (family pairs), than WFSR does (individual families).

In theory and in a broad range of simulations, we find that WFSR and BFSR obtain unbiased estimates of heritability. WFSR and BFSR respectively obtain two and ten times smaller standard errors than state-of-the-art sibling-based heritability estimators do. Furthermore, we show that GxE affects WFSR and BFSR estimates differently when environmental variation is greater between families than within families. Comparing these estimators produces a natural measure of the role of GxE, enabling us to quantify how much GxE may contribute to the “missing heritability” problem. In this study, we are additionally applying our methods to the UK Biobank, estimating heritability and GxE for a variety of anthropometric, cardiovascular, psychiatric, and behavioral phenotypes, bringing us closer to understanding how genetics and the environment jointly give rise to an array of phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4148F: Uncovering causal gene-tissue pairs and variants: A multivariable TWAS method that controls for infinitesimal effects ⭐

Authors:

Y. Yang[1], X. Zhu[1], N. Lorincz-Comi,[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Cleveland Clinic Fndn., Cleveland, OH

Abstract:

Multivariable TWAS (MV-TWAS) methods, such as causal TWAS (cTWAS) and Tissue-Gene Fine-Mapping (TGFM), are increasingly popular for identifying causal genes in a cis region. These methods consider a causal diagram where eQTLs regulate the expressions of tissue-specific causal genes, which subsequently affect a phenotype. They perform fine-mapping to address challenges caused by eQTLs of tissue-specific causal genes being in linkage disequilibrium (LD) with variants in: (1) nearby non-causal genes, (2) causal genes of non-causal tissues, and (3) nearby direct causal variants, by using the sum of single effects (SuSiE). However, Cui et al. (2024) reported high replication failure rates with current Bayesian fine-mapping methods, including SuSiE, potentially due to not considering widespread infinitesimal effects. We introduce a new MV-TWAS method called the Tissue-Gene pairs, direct causal Variants, and Infinitesimal effects (TGVI) selector. This method identifies causal gene-tissue pairs and direct causal variants using SuSiE while simultaneously addressing infinitesimal effects through Restricted Maximum Likelihood (REML). The Pratt index, a metric that measures the ratio of local genetic variance of the outcome explained by a gene-tissue pair or direct causal variant, is applied to quantify the importance. We applied the TGVI selector to 40 metabolic and cardiovascular traits using their largest GWAS datasets, along with eQTL and sQTL summary data from 28 GTEx tissues, using an LD reference panel with over 9 million variants. We reported candidate causal gene-tissue pairs and direct causal variants, identified major relevant tissues, and conducted enrichment analysis to uncover underlying biological processes. On average, the TGVI selector and TGFM identified a total of 135.6 and 175.2 causal gene-tissue pairs, and 88.2 and 148.2 direct causal variants with Pratt indices > 0.2, respectively. By accounting for infinitesimal effects, the TGVI selector removes likely falsely discovered gene-tissue pairs and variants whose contributions (measured by the Pratt index) are relatively smaller. For example, the TGVI selector localized PCSK9-Blood (Pratt = 0.19) and rs11591147 (Pratt = 0.50) as the only causal gene-tissue pair and direct causal variant for LDL-C. In contrast, TGFM identified PCSK9-Blood (Pratt = 0.18) and lnc-PCSK9-4 (Pratt = -0.02), as well as rs11591147 (Pratt = 0.49) and seven other direct variants (maximum Pratt = 0.041) as potential causals at the PCSK9 locus. Our

results suggest that the TGVI selector can identify causal elements of GWAS loci with better accuracy and resolution than TGFM.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4150F: Unveiling the power of allelic series: enhancements and applications of COAST

Authors:

Z. McCaw[1], J. Gao[2], R. Dey[1], S. Tucker[1], C. O'Dushlaine[1], T. Soare[1]; [1]Insitro, South San Francisco, CA, [2]Univ. of Toronto, Toronto, ON, Canada

Abstract:

An allelic series is a collection of variants in a gene or pathway where increasingly deleterious mutations have increasingly large phenotypic effects. Such genes constitute “experiments of nature”, wherein the phenotypic consequences of variation in gene function are directly observable. These genes are of particular interest for therapeutic development because the existence of a dose-response relationship between gene function and phenotypic impact suggests how to perturb the gene for clinical benefit. The coding-variant allelic series test (COAST) was the first rare variant association test specifically tailored for the identification of allelic series. Here, we describe several extensions of COAST, and its application to the full collection of unrelated exomes from the UK Biobank (UKB). First, we experiment with replacing the default weights used by COAST with weights derived from recent large language models, such as AlphaMissense. Second, we implement a summary statistic-based version of COAST, enabling identification of allelic series starting from per-variant association statistics rather than individual-level data. Third, we present strategies for meta-analyzing across ancestries, allowing for the identification of cross-ancestry allelic series. Finally, we compare results from our preliminary analysis of lipid traits among 150K UKB participants with results from the expanded cohort of 350K UKB participants. We publicly release a workflow for running COAST on the UKB’s Research Analysis Platform.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4151F: Valid inference for machine learning-assisted GWAS

Authors:

J. Miao[1], Y. Wu[1], Z. Sun[1], X. Miao[1], T. Lu[2], J. Zhao[1], Q. Lu[1]; [1]Univ. of Wisconsin–Madison, Madison, WI, [2]Univ. of Toronto, Toronto, ON, Canada

Abstract:

Machine learning (ML) has revolutionized analytical strategies in almost all scientific disciplines including human genetics and genomics. A rising trend in complex trait genetics research is the ML-assisted genome-wide association study (GWAS), which applies advanced ML techniques to predict phenotypes that are difficult and expensive to measure (e.g., undiagnosed diseases, imaging-derived outcomes, molecular traits in rare tissues), and then conducts GWAS on these ML-imputed outcomes. However, all existing approaches for ML-assisted GWAS treat imputed phenotypes as observed and completely neglect the inherent uncertainty associated with 'black-box' ML algorithms. In this work, we first demonstrate the risk of pervasive false positive associations in existing ML-assisted GWAS. Using real data benchmarking, we find that a shocking 81% of the associations in GWAS of ML-imputed type 2 diabetes fail to replicate in the ground truth GWAS of type 2 diabetes. We then introduce POP-GWAS, a principled statistical framework that reimagines ML-assisted GWAS, ensuring valid and powerful results irrespective of the accuracy of imputation, the choice of ML algorithm, and variables used for imputation. It is a versatile tool that can account for binary phenotypes, sample relatedness, and selection bias, making it suitable for broad applications. It also only requires GWAS summary statistics as input, and is computationally fast. We provide a theoretical guarantee that POP-GWAS is the statistically optimal solution to ML-assisted GWAS. Using POP-GWAS, we conducted the largest-to-date GWAS and rare-variant association analysis on bone mineral density (BMD) derived from dual-energy X-ray absorptiometry imaging (DXA) at 14 skeletal sites, achieving a 9.7%-50.7% gain in effective sample size compared to conventional approaches. We identified 89 novel genome-wide significant loci not previously implicated in BMD GWAS and revealed the skeletal site-specific genetic architecture of BMD. It also nominated novel head-specific BMD genes including *LGR5* supported by evidence of *Lgr5*-deleted mice exhibiting a range of craniofacial abnormalities. We further identified 47 novel genes associated with BMD through ML-assisted rare-variant association analysis, pinpointing potential therapeutic targets for osteoporosis. This work introduces key methodological advances that will fundamentally reshape the analytical strategies in future ML-assisted genetic association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4152F: Viewing the role of aging through the prism of changing genetic effects

Authors:

X. Miao, S. Smith, E. Tucker-Drob, A. Harpak; Univ. of Texas at Austin, Austin, TX

Abstract:

The role of genetic variation in shaping complex trait variation is dynamic within a population. This dynamism may result from differences in exposures, lifestyle, or changes in the importance of biological pathways as individuals within the population age. To better understand the changing role of genetic variation as populations age, we develop a method to estimate changes in genetic effects as a function of age. In particular, we assume there exists a set of archetypal functional relationships of variance due to genetic effects with age. Therefore, we model variants genome-wide as sampling from a single distribution-a mixture of possible such relationships, and apply it across traits and diseases in a sample of individuals aged 40-69 in the UK Biobank. For example, in weight and body proportion-related traits we find that genetic effect sizes are constant until around age 50 and then begin to decrease rapidly over the next 20 years. Furthermore, our analysis of weight-related traits reveals that the FTO gene, a known contributor to anthropometric trait variance, shows a substantial change in its effect with age. In contrast, we observe a monotonic decline of effect sizes with age in several traits, including calcium, red blood cell count, urea, and other biomarkers. While previous research has shown a decline in heritability with age, our results suggest that in this suite of traits the decrease in heritability with cohort age is in large part due to diminished magnitude of genetic effects rather than changes in lifestyle and environmental exposures alone.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Friday Session

Board 4153F: Vitamin D receptor-binding variants confer risk of Sjögren's disease in Europeans

Authors:

D. Yilmaz[1,2], K. Taylor[3], L. Criswell[4], C. Shiboski[5], L. F. Barcellos[1,2]; [1]Univ. of California, Berkeley, Berkeley, CA, [2]Ctr. for Computational Biology, Div. of Computing, Data Sci. and Society, Univ. of California, Berkeley, Berkeley, CA, [3]Div. of Rheumatology, Univ. of California, San Francisco, San Francisco, CA, [4]NIH, NIAMS, Potomac, MD, [5]Dept. of Orofacial Sci., Univ. of California, San Francisco, San Francisco, CA

Abstract:

Sjögren's Disease (SjD) is a chronic, multisystem autoimmune disease that causes significant morbidity. Patients typically present with ocular dryness and dry mouth due to exocrine gland infiltration; other systemic features may include inflammatory arthritis, subcutaneous lupus

erythematosus, neuropathies, interstitial lung disease, and primary biliary cirrhosis. Previous studies of autoimmune diseases, including SjD, have investigated the role of low vitamin D as a risk factor using Mendelian randomization (MR) analysis. Vitamin D is crucial for many biological processes. After ingestion or absorption, serum vitamin D is converted to its more stable form, 25-hydroxyvitamin D [25(OH)D], and then to 1,25-dihydroxyvitamin D [1,25$(OH)_2$D]. 25(OH)D signals through the nuclear vitamin D receptor (VDR), a ligand-regulated transcription factor modulating vitamin D-regulated gene expression. Directly testing for associations between VDR binding and disease phenotypes in large human studies is challenging. SNPs associated with genetic variation in VDR binding affinity (VDR binding variants or VDR-BVs) have been identified using ChIP-exo data from calcitriol-stimulated lymphoblastoid cell lines followed by Allele-seq. We recently studied these VDR-BVs as genetic instrumental variables using MR analysis in multiple sclerosis (MS), identifying evidence for causal associations with several VDR-BVs. The present study aimed to examine these VDR-BV candidates in SjD. We utilized genetic data for 556 SjD cases from the Sjögren's International Collaborative Clinical Alliance (SICCA) cohort and 3,517 unrelated controls; all participants were of European ancestry. Cases were those with symptoms suggestive of SjD, confirmed using the 2016 ACR/EULAR criteria based on a total score of ≥4 for the weighted sum of 5 items: anti-SSA(Ro) antibody positivity and FLS and FS ≥1 foci/mm^2 scoring 3; OSS ≥5, Schirmer test ≤5 mm/5 min, and UWS flow rate ≤0.1 mL/min each scoring 1. Genotyping was performed using the Illumina Omni2.5M platform and data were imputed against reference haplotypes from the Haplotype Reference Consortium using IMPUTE4. After quality control measures, SNPs were excluded if minor allele frequency (MAF) was <5% resulting in 17 candidate VDR-BVs for analysis. Associations between VDR-BVs and SjD were tested using logistic regression in PLINK, adjusting for the first five genetic principal components. Two VDR-BVs were associated with SjD at $p<0.05$: rs2531804 (OR: 1.24, 95% CI: 1.09-1.41) and rs2881514 (OR: 1.14, 95% CI: 1.00-1.29). These findings provide further evidence that at least one known VDR-BV associated with MS is also implicated in SjD.